



**HAL**  
open science

## **Modelling for Principal Stratification: Analyzing Stroke Events in CADASIL Patients Across NOTCH3 Variant**

Léa Aguilhon, Sophie Tezenas Du Montcel, Juliette Ortholand

### ► **To cite this version:**

Léa Aguilhon, Sophie Tezenas Du Montcel, Juliette Ortholand. Modelling for Principal Stratification: Analyzing Stroke Events in CADASIL Patients Across NOTCH3 Variant. ISCB 2025 - 46th Annual Conference of the International Society for Clinical Biostatistics, Aug 2025, Basel, Switzerland. <hal-05285233>

**HAL Id: hal-05285233**

**<https://hal.science/hal-05285233v1>**

Submitted on 26 Sep 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

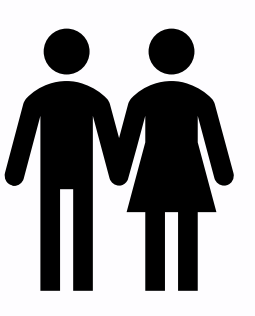
# Modelling for Principal Stratification: Analyzing Stroke Events in CADASIL Patients Across NOTCH3 Variant

## Background & Objectives

Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CADASIL), a cerebral small vessel disease caused by NOTCH3 mutations, shows heterogeneity in age of onset, symptom severity, and stroke recurrence. Variants in EGFr domains 1-6 may worsen disease progression, but post-stroke trajectories remain unclear. Traditional studies face bias from competing risks (death truncating stroke observation). To address this, we propose a causal inference approach: first comparing mutation types and first stroke incidence, then analyzing second strokes using Principal Stratification, a method that accounts for immortal time bias by modeling latent survival outcomes. Finally, we assess mortality differences between variant groups. This framework enables unbiased estimation of variant-specific stroke risks, improving disease trajectory understanding for clinical trials.

## Methods

### Population

- 326 CERVCO's CADASIL patients
  - First visit between 2005 and 2021
  - Have done at least one stroke before or after inclusion (165 had a 2nd stroke)
-  Reference time = Age first stroke

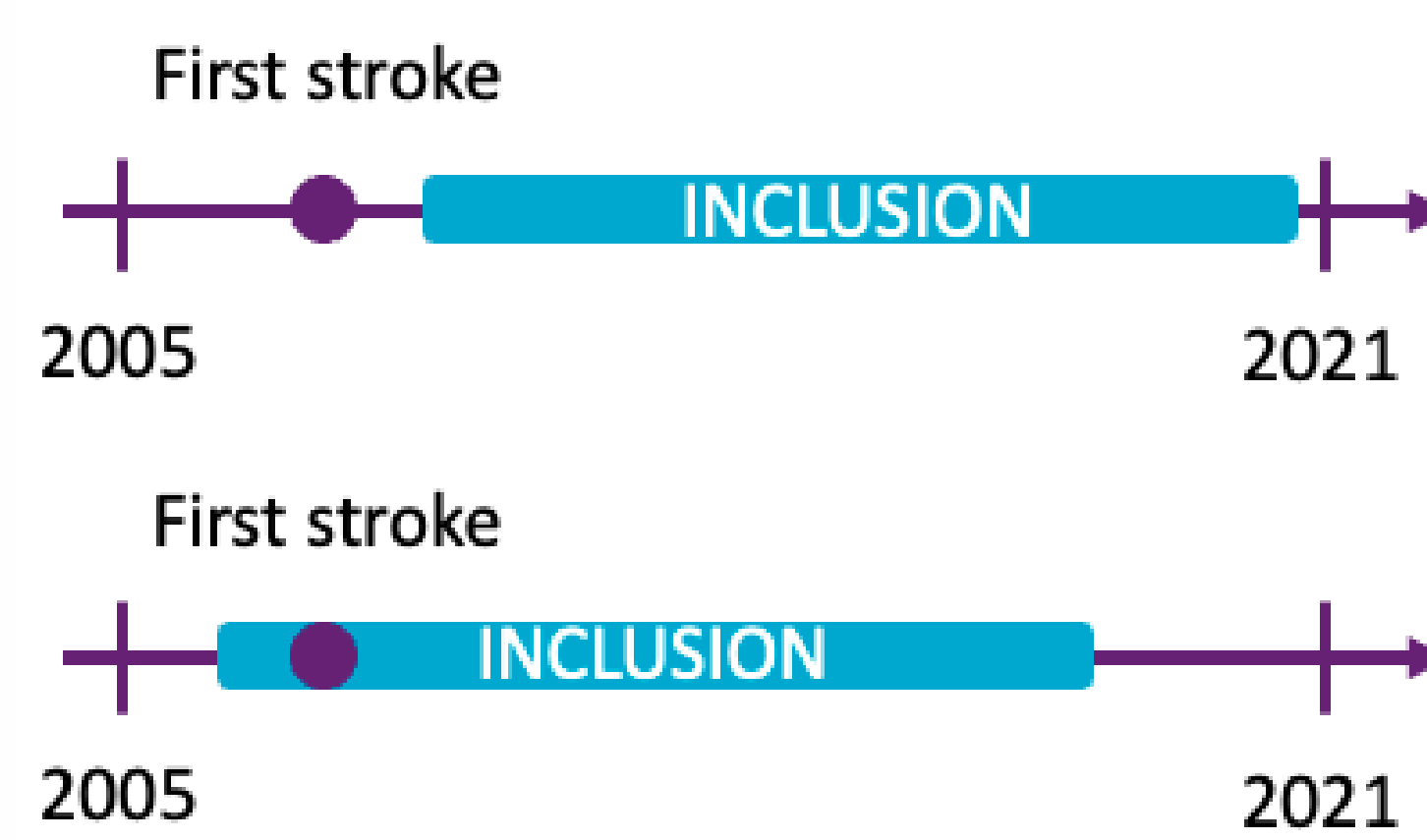


Figure 1: Schema inclusion criteria with first stroke

### Covariates

- Sex
- Education level
- EGF domain mutation
- Age first stroke
- Cardiovascular risk factors
  - Alcohol
  - Hypertension
  - Hypercholesterolemia
  - Smoking

### Events outcomes

- Death
- Second stroke

Stratas	Y(0) 7-34 mutation	Y(1) 1-6 mutation
Always-survivors	Alive	Alive
Protected	Dead	Alive
Harmed	Alive	Dead
Never-survivors	Dead	Dead

Table 1: Classification of strata for principal stratification based on predicted survival status from a Cox model, based on mutation status

### Statistical Method

- Propensity score matching** for covariates balancing to create a pseudo-randomized and population groups (1-6 vs 7-34 mutations).  
⇒ Avoid observational bias
- Survival model adjusted (Cox model)** to predict principal stratification  
⇒ Neutralizes competing influences on survival, so the principal stratification reflects mutation-specific effects
- Selection of Always-survivors**  
⇒ Avoids immortal time bias and competing risk from death, isolating the pure effect on stroke recurrence
- RMST (Restricted Mean Survival Time)** comparison between EGFr 1-6 and 7-34 groups  
⇒ Provides an interpretable estimand, robust to censoring and model assumptions

## Results

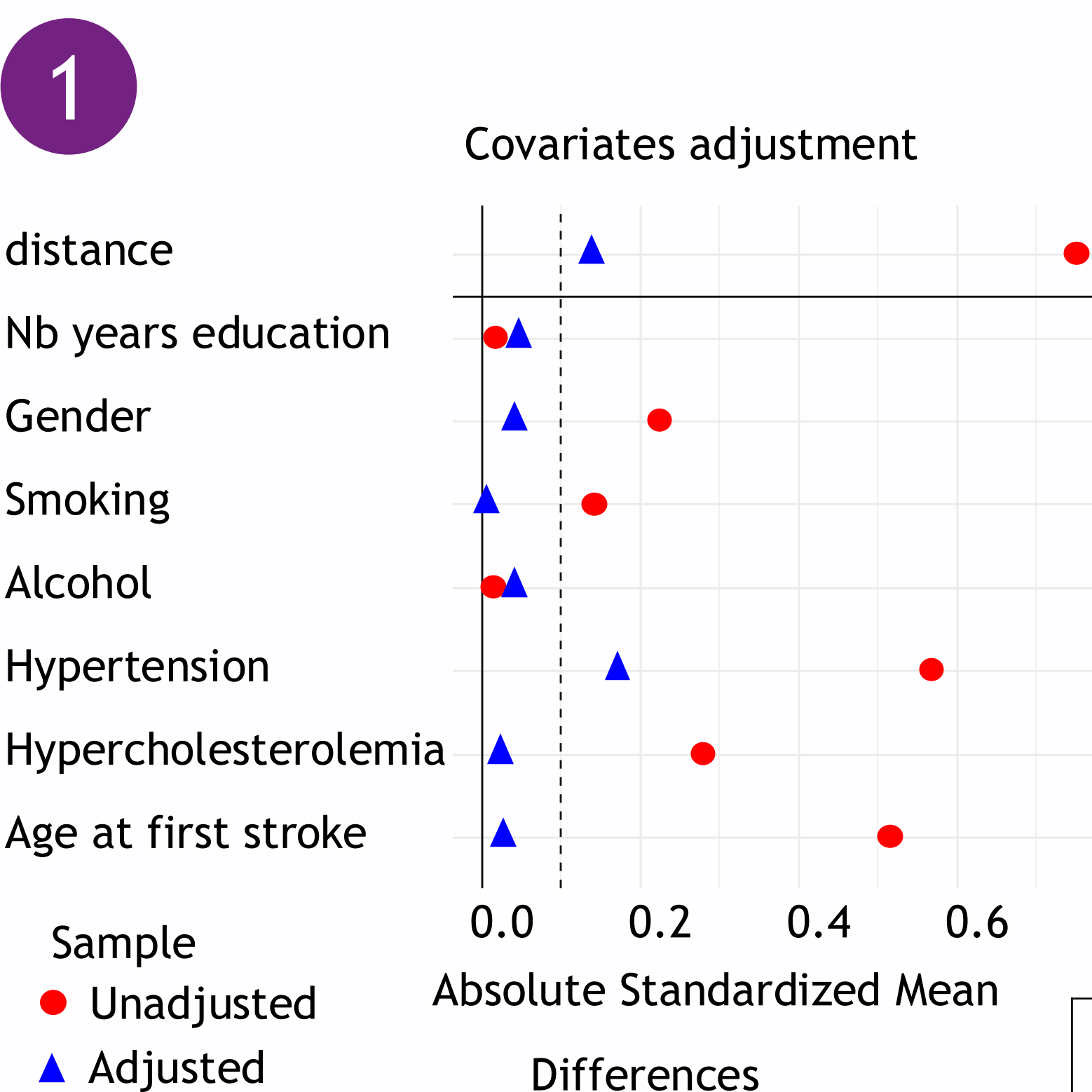


Figure 2: Love plot assessing covariate balance before and after matching

Propensity score permits to have more equilibrated groups (1-6 vs 7-34 EGFr mutation)

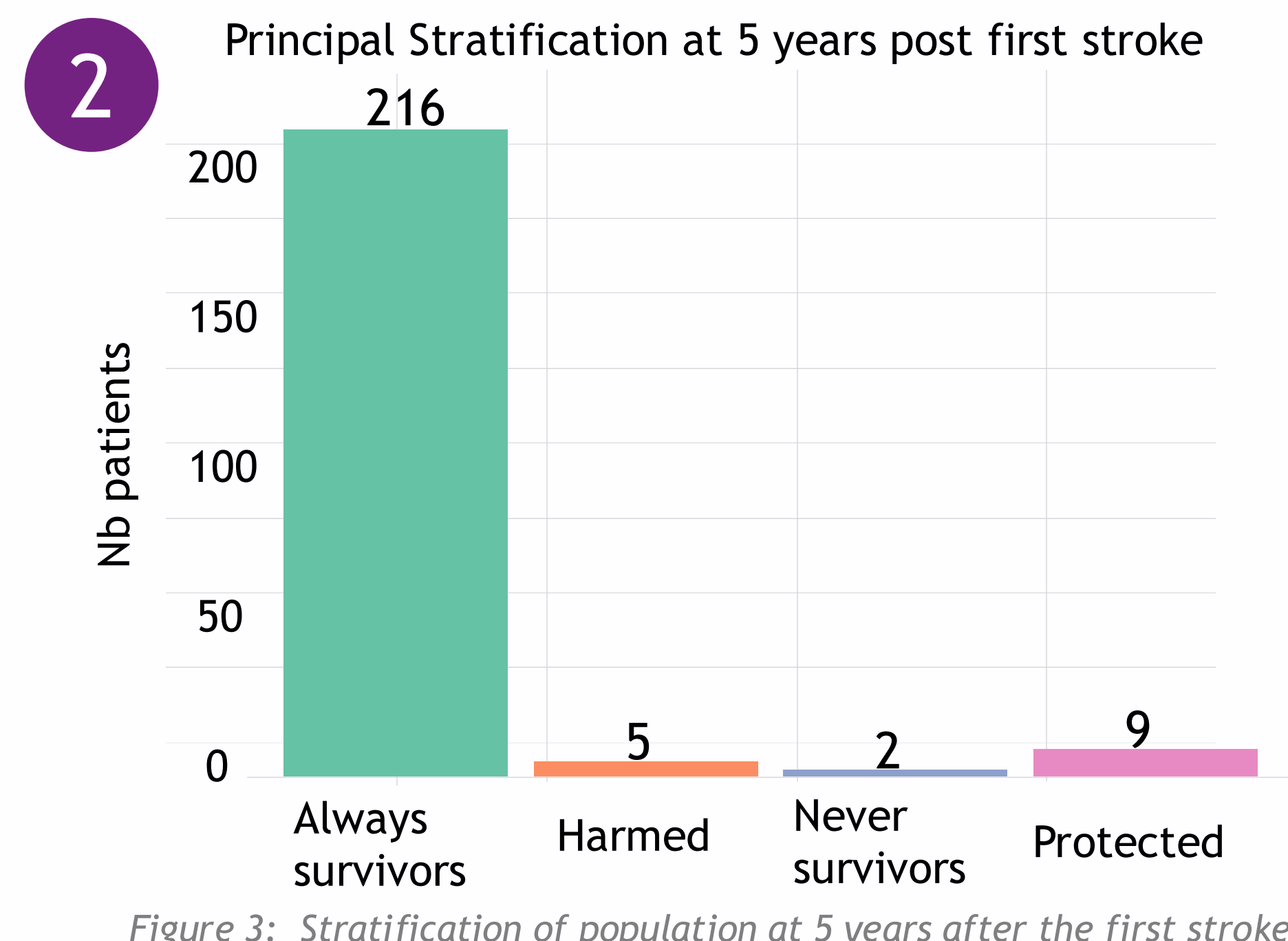


Figure 3: Stratification of population at 5 years after the first stroke

Years to predict	Nb always survivors	RMST 1-6 mutation	RMST 7-34 mutation	Difference	p-value
2 years	232	1.70 [1.59-1.82]	1.74 [1.62-1.83]	0.020	0.797
5 years	216	3.83 [1.48-2.17]	3.88 [2.74-4.22]	0.057	0.814
10 years	195	6.43 [5.74-7.22]	6.52 [5.67-7.38]	0.127	0.830
15 years	154	8.64 [7.30-9.98]	8.57 [7.03-10.11]	0.110	0.916

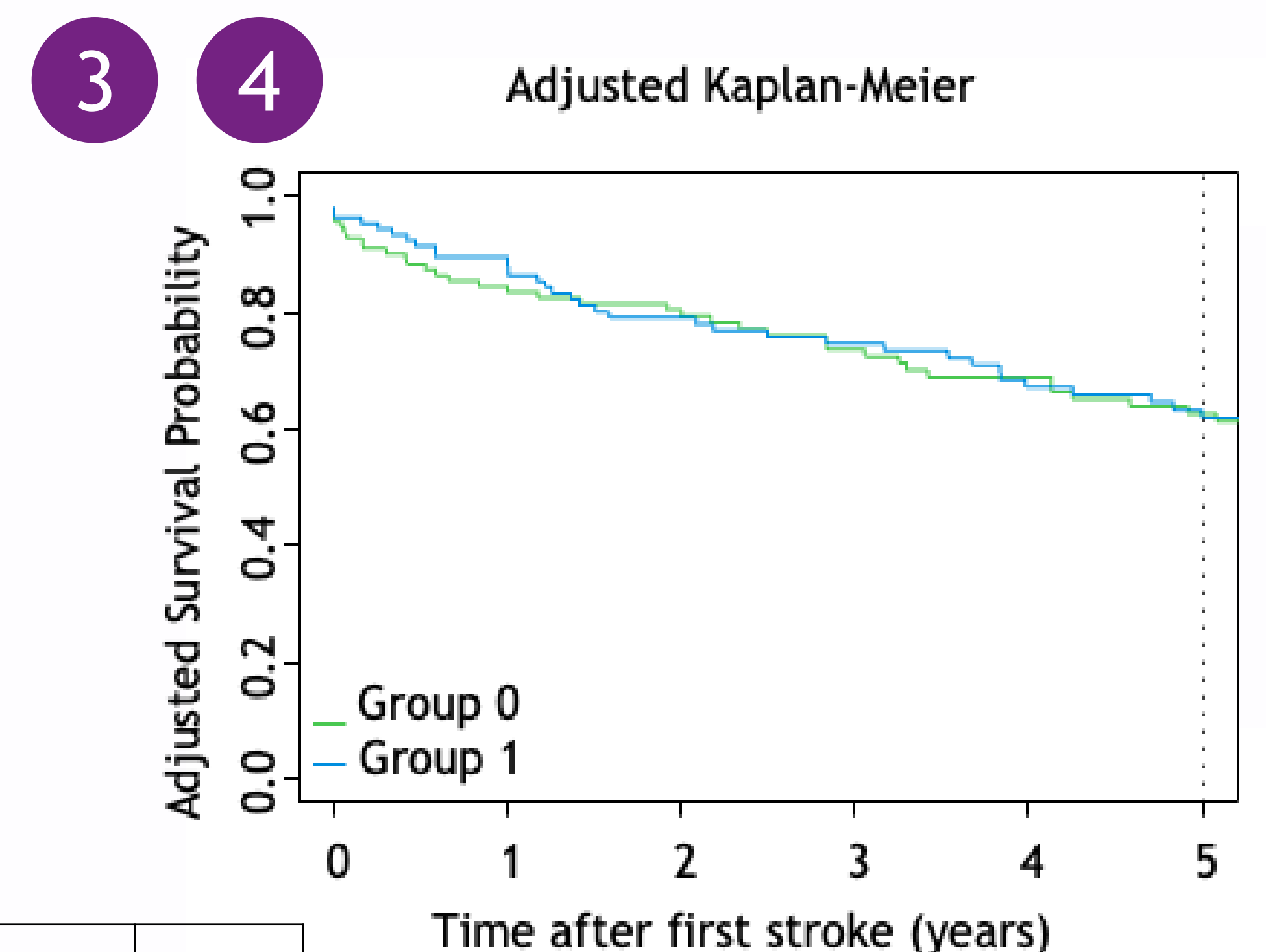


Figure 4: Adjusted Kaplan-Meier curve estimating second stroke risk within 5 years following initial stroke

No significant difference are observed between 1-6 and 7-34 mutation groups when studying their second stroke.

Table 2: RMST estimand for 1-6 and 7-34 EGFr mutation type at 2,5,10 and 15 years after the first stroke

## Conclusion

There were no significant differences in second stroke study between groups (1-6 and 7-34 EGFr mutation). Further confirmation analysis could lead to a reference time point for this heterogeneous disease, to develop future trials.

- References:**
- Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics*. 2002 Mar;58(1):21-9. doi: 10.1111/j.0006-341x.2002.00021.x. PMID: 11890317; PMCID: PMC4137767
  - Han K, Jung I. Restricted Mean Survival Time for Survival Analysis: A Quick Guide for Clinical Researchers. *Korean J Radiol*. 2022 May;23(5):495-499. doi: 10.3348/kjr.2022.0061. PMID: 35506526; PMCID: PMC9081686.