

Comparative Analysis of Antithrombotic Strategies in Atrial Fibrillation Patients with Ischemic Stroke: A Propensity Score and Inverse Probability Weighting Approach with Three Groups

Ambrogi Arianna⁽¹⁾, Panetta Valentina⁽¹⁾, Folcarelli Silvia⁽¹⁾, Piras Marina⁽²⁾, Diomedi Marina⁽²⁾, Maestrini Ilaria⁽²⁾

(1) *L'altrastatistica Srl*

(2) *Stroke Center, Department of Systems Medicine, University Hospital of Rome Tor Vergata, Rome, Italy*

CORRESPONDING AUTHOR: Ambrogi Arianna, ariambrogi@gmail.com

INTRODUCTION

Acute ischemic stroke (AIS) is a severe complication in patients with atrial fibrillation (AF), even in those already treated with direct oral anticoagulants (DOACs). Growing evidence indicates that systemic inflammation plays a key role in promoting the prothrombotic state and increasing the risk of thromboembolic events in AF. In observational studies, treatment allocation is not assigned randomly but rather reflects real-world clinical decision-making. DOACs treatment is prescribed based on the patient's estimated thromboembolic risk, typically assessed using established scoring systems such as CHA2DS2-VASc score. However, despite appropriate anticoagulation, a residual thromboembolic risk persists in a subset of patients, as documented in real-world studies. The characterisation of this subgroup of patients is essential to elucidate pathophysiological mechanisms and guide potential therapeutic modifications. Accordingly, within a cohort of AIS patients, different antithrombotic strategies can be encountered among patients with AF. This variability, combined with the lack of randomization, can introduce confounding and selection bias, making it difficult to accurately compare clinical outcomes [1]. In this context, methods such as propensity score analysis are essential to adjust for these imbalances and improve the validity of comparisons. In this retrospective cohort of patients aged ≥ 70 with acute ischemic stroke (AIS) due to large vessel occlusion (LVO), three groups were identified: (1) AF patients treated with direct oral anticoagulants (AFwDOACs), (2) AF patients without anticoagulant therapy (AFwoDOACs), and (3) patients with large-artery atherosclerosis (LAA). The primary aim of the study was to compare the inflammatory profiles between the two AF groups -those

receiving DOACs and those not- within the context of cardioembolic AIS related to LVO. The LAA group served as a reference cohort, allowing for contextualization of inflammatory markers in a non-cardioembolic stroke etiology.

OBJECTIVES

To compare the impact of different statistical methods -Propensity Score Matching (PSM) and Inverse Probability Weighting (IPW)- in identifying clinical and laboratory differences between treatment groups. Specifically, we aimed to: (1) evaluate the ability of these methods to balance covariates across groups; (2) compare findings across methods in both pairwise and three-group analyses; (3) assess how statistical approach choice impacts the identification of clinically meaningful biomarkers.

METHODS

In this study, we retrospectively analyzed patients admitted between 2015 and 2024 with acute ischemic stroke due to large vessel occlusion (LVO). Patients were categorized into three groups based on prior clinical characteristics: (1) AF patients treated with direct oral anticoagulants (AFwDOAC), (2) AF patients not receiving oral anticoagulants (AFwoDOAC), and (3) patients with large-artery atherosclerosis (LAA). The third group served as a reference cohort, functioning as a control group to contextualize inflammatory profiles in a non-cardioembolic stroke population.

A major challenge of this observational design was achieving covariate balance across three distinct groups. While

propensity score (PS) methods such as matching and inverse probability weighting (IPW) are widely used, they are most commonly applied in two-group comparisons [2]. To address this, we implemented two complementary strategies: (1) pairwise balancing (comparing Group 1 vs Group 2 and Group 1 vs Group 3 separately), and (2) simultaneous balancing across all three groups using a multinomial PS model.

Covariates included in all PS models were: age, sex, baseline NIHSS score, pre-stroke statin use, and vascular territory. To improve covariate balance and ensure model stability, the analysis was restricted to patients aged 70 years or older, as the inclusion of younger patients introduced instability in PS and IPW estimation.

PAIRWISE COMPARISONS:

For each pairwise comparison, propensity score matching was performed using nearest-neighbour matching with a 1:2 matching ratio and a caliper of 0.2 standard deviations. The caliper ensured that only treated and control individuals with sufficiently similar propensity scores were matched, reducing potential bias from poor-quality matches [3]. The standardized mean difference was used to evaluate balancing between covariates.

In parallel, IPW was applied using stabilized weights to reduce the influence of extreme weights [4]. The Average Treatment Effect (ATE) was used as the estimand. To limit the impact of outliers, weights were truncated at the 1st and 99th percentiles.

THREE-GROUP COMPARISON:

To balance covariates across all three groups simultaneously, propensity scores were estimated using a multinomial logistic regression model using stabilized and truncated weights, again with the ATE as the estimand. This model extends binary logistic regression to multiple categories, estimating the conditional probability of belonging to each group given the covariates. These probabilities were used to generate stabilized inverse probability weights for multi-group IPW, allowing for covariate adjustment across all groups within a single model. This approach allowed for the inclusion of all patients in a single weighted model, improving comparability across the three clinical profiles.

Covariate balance was assessed graphically using Love plots (see Figure 1), which display standardized mean differences before and after weighting or matching. An STD value ≤ 0.1 was considered a good balancing. For statistical testing, weighted t-tests were conducted for both pairwise and three-group comparisons.

A broad panel of biochemical and haematological markers was analyzed across groups, including lipid profiles, inflammatory indices (e.g., neutrophil-to-lymphocyte ratio, monocyte-to-lymphocyte ratio), and coagulation parameters (e.g., D-dimer). When global differences were identified in the three-group analysis using weighted ANOVA, post hoc pairwise comparisons were performed using model-based t-tests.

For additional comparison, an unweighted ANOVA was also conducted on the raw data, to evaluate differences in the absence of statistical adjustment.

Residual inflammatory risk (RIR) -defined as hsCRP > 2 mg/dL- and residual cholesterol risk (RCR) -defined as LDL >

70 mg/dL-were analyzed using chi-square tests.

All analyses were performed using RStudio (v2025.05.496).

RESULTS

Both PSM and IPW effectively improved covariate balance (see Figure 1), although the number of units used varied between methods:

- Three-group IPW: 48/59 in group 1, 287/294 in group 2 and 125/151 in group 3;
- IPW pairwise (Group 1 vs 2) 52/59 in group 1 and 293/294 in group 2;
- IPW pairwise (Group 1 vs 3): 43/59 in group 1 and 138/151 in group 3;
- PSM (Group 1 vs 2): 56/59 in group 1 and 111/294 in group 2;
- PSM (Group 1 vs 3): 54/59 in group 1 and 89/151 in group 3.

In pairwise comparisons:

- PSM identified significantly higher MLR and lower triglycerides in Group 3 vs Group 1 ($p = 0.041$ and $p=0.003$ respectively).

- IPW identified significantly lower triglycerides ($p=0.008$) and lower D-dimer ($p=0.012$) in Group 3 vs Group 1 and detected significantly higher D-dimer in Group 1 vs Group 2 ($p = 0.024$).

- In the three-group IPW-weighted ANOVA, MLR, triglycerides, and D-dimer were all significantly different across groups. Interestingly, Total-Cholesterol also reached significance under IPW but not in unadjusted models, potentially due to improved covariate balance revealing subtler metabolic effects.

- In the unweighted (raw data) ANOVA only triglycerides were statistically significant.

The post-hoc analyses confirmed significant differences in D-dimer (Group 1 vs 2 and Group 1 vs 3) and triglycerides (Group 1 vs 3). Other significant results from the weighted ANOVA were primarily attributable to differences between Group 2 and Group 3.

These findings demonstrate that the choice of adjustment method can shift which biomarkers appear relevant, reinforcing the importance of using multiple statistical approaches to capture meaningful clinical signals.

CONCLUSIONS

The results showed suggest that Propensity Score Matching (PSM) identified a significant difference in MLR between groups; however, this result was not confirmed by post hoc analyses following the three-group IPW-weighted ANOVA. This discrepancy may be attributed to the matching ratio (1:2), which necessarily excludes a substantial portion of observations from Groups 2 and 3, potentially limiting representativeness and power. In fact, MLR was only significantly different between Groups 2 and 3 in post hoc tests, but this comparison is outside the primary scope of this study. Moreover, PSM failed to detect significant differences in D-dimer between Groups 1 and 2, which were instead clearly identified

through IPW, both in pairwise and three-group comparisons.

The limitations of PSM in studies involving more than two groups are underscored by the need to select separate matched datasets for each pairwise comparison. This results in inconsistent representation of Group 1 across comparisons and makes integration of findings more difficult. In contrast, IPW applied across all three groups, in combination with post hoc weighted analyses, successfully identified clinically relevant differences in D-dimer and lipid markers. IPW appears to be a robust and flexible method for covariate adjustment in in three-group observational.

Unweighted analyses were not able to detect differences beyond triglycerides, this highlights the importance of using covariate balancing techniques, particularly when dealing with groups of highly unequal size. These results emphasize the critical role of appropriate statistical methods in observational research and support the application of multinomial IPW as a valuable strategy when multiple treatment groups are involved.

The variation in findings across PSM, IPW, and unadjusted models underscores the importance of triangulating results from multiple analytical strategies, especially in unbalanced observational datasets.

REFERENCES

1. Halpern EF., Behind the Numbers: Inverse Probability Weighting. *Radiology*, 2014 Jun; 271 (3):631-2
2. Austin PC., A Tutorial and Case Study in Propensity Score Analysis: An Application to Estimating the Effect of In-Hospital Smoking Cessation Counseling on Mortality. *Multivariate Behav Res.*, 2011; 46(1):119-51
3. Courboin E., Comparing Robotic-Assisted to Open Radical Cystectomy in the Management of Non-Muscle-Invasive Bladder Cancer: A Propensity Score Matched-Pair Analysis, *Cancers* 2023, 15, 4732.
4. Young J., Inverse probability weighted estimation of risk under representative interventions in observational studies. *J Am Stat Assoc.* 2019;114(526):938-947.

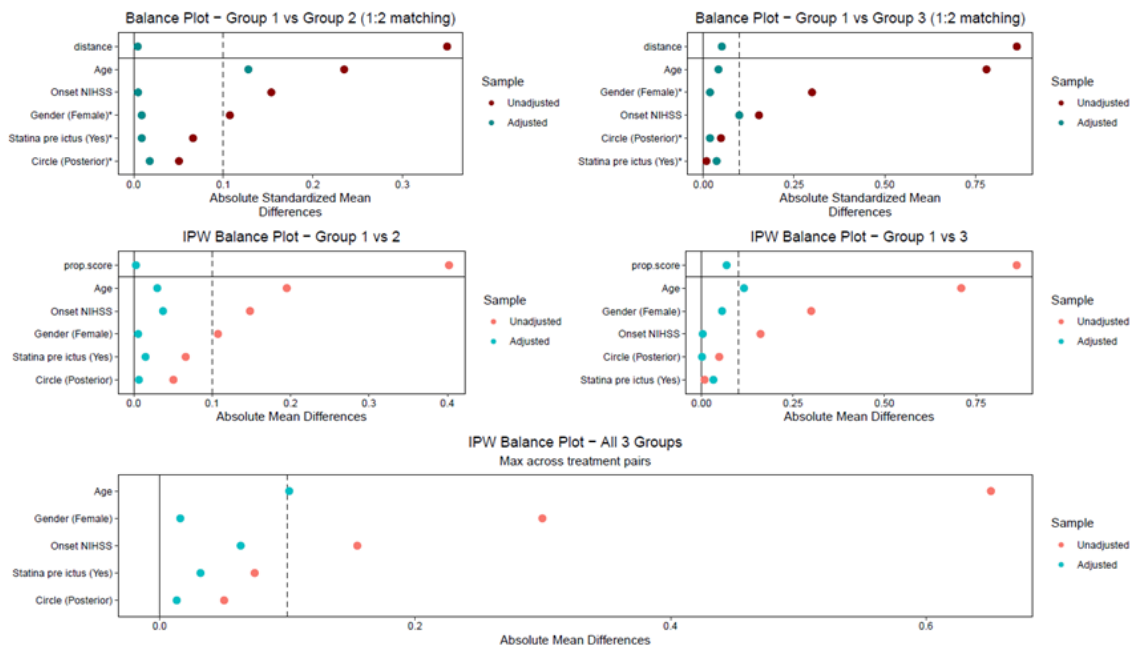


Figure 1. Covariate balance across groups before and after adjustment using PSM and IPW. Plots show standardized mean differences across age, sex, NIHSS, statin use, and vascular territory. Adjustment substantially reduced imbalance in all two-by-two comparisons and the overall three-group analysis