Leveraging Machine Learning for Causal Inference: A Case Study in Interventional MRI for Atrial Flutter

Andrei Alexandru Mircea[1], Jorge Solana Munoz[1], Cosima Jahnke[2], Kerstin Bode[2], Etienne Pruvot[1], Juerg Schwitter[1], Ingo Paetsch[2], Adrian Luca[1]

[1] Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland
[2] Heart Center, University Hospital Leipzig, Leipzig, Germany

Abstract

Background: *Normalized elongation (NE) is a novel biomechanical marker of the linear stress–strain regime of the right atrium. Previous work demonstrated a significant negative association between external cardioversion (EC) and NE in patients undergoing atrial flutter ablation in an interventional MRI (iMRI) setting. This study aimed to evaluate whether this relationship is causal using a combination of causal inference and machine learning methods.*

Methods: *A cohort of 32 patients undergoing iMRI-guided atrial flutter ablation was analyzed, including 15 who received EC prior to ablation. Smoking status and dyslipidaemia were considered potential confounders. Causal inference was performed using random forest partial linear regression (DoubleML), linear regression (DoWhy), covariate adjustment, propensity score matching, and propensity score inverse probability weighting. Robustness was assessed using placebo refutation and Rosenbaum sensitivity analysis.*

Results: *Both the linear regression and the partial linear regression models estimated a significant causal effect of EC on NE ($\beta = -0.36$, $p < 0.05$). Inverse probability weighting produced a comparable average treatment effect (ATE = -0.38, $p = 0.025$), whereas propensity score matching was not significant, likely due to the small sample size. Placebo testing eliminated the observed effect ($\beta = -0.015$, $p = 0.86$), and Rosenbaum sensitivity analysis ($\Gamma = 3.1$) indicated that a relatively strong unmeasured confounder would be required to negate the findings.*

Conclusions: *Multiple causal inference approaches consistently support a causal negative effect of external cardioversion on normalized elongation. These findings demonstrate the feasibility of combining statistical and machine learning causal inference methods in small observational clinical cohorts.*

1. Introduction

Normalized elongation (NE) is a novel biomechanical marker for the linear stress-strain regime of the right atrium [1]. Our prior work on atrial flutter ablation in an interventional MRI (iMRI) setting showed a significant negative association between external cardioversion (EC) and NE [1]. The medical field has historically used the Bradford-Hill criteria to demonstrate causality [2]. Machine learning techniques can nowadays provide a new framework to test for causality [3,4]. This framework represents a great alternative for small, observational cohorts used in clinical studies.

Our project is aimed at providing an *in silico* framework to assess whether the EC-NE relationship is causal.

2. Methods

2.1. Demographics and interventional data

The cohort comprised 32 patients who underwent atrial flutter ablation in an iMRI environment (IMRICOR). Fifteen patients underwent cardioversion prior to ablation. Previously, we showed that EC, dyslipidemia (Dys) and smoking status (SS) were associated with NE values [1]. Dyslipidemia was defined as having an LDL cholesterol greater than 3.0

mmol/L. Smoker status was defined as having consumed more than 100 cigarettes in his/her lifetime and having smoked in the last 12 months. The cohort traits are presented in Table 1.

Table 1. Cohort demographics and interventional data

Patients (n = 32)	EC (n=15)	No EC (n=17)
Smoker, n(%)	9 (60%)	8 (47%)
Non-smoker, n(%)	6 (40%)	9 (53%)
Dyslipidaemia, n(%)	9 (60%)	8 (47%)
No dyslipidemia, n(%)	6 (40%)	9 (53%)
Normalized elongation (mean ± std)	0.52±0.3	0.94±0.75

EC- electrical cardioversion before ablation

2.2. Causality methodology

Causality was assessed using a mix of statistical and machine learning methods. The machine learning methods involved training a random forest regressor (Double machine learning partial linear-regressor – DoubleML library) and a linear regressor model (DoWhy library) [3,4]. The directed acyclic graph showing the relationships is depicted in Figure 1. NE was considered the outcome. EC was considered the treatment. Both the smoker status and dyslipidaemia were confounders.

Covariate adjustment (DoWhy library) as well as propensity score matching and inverse weighting were used to assess the EC-NE causal relationship [5]. The covariate adjustment models were ordinary least square models, where the log(NE) was used to account for non-normal distribution. In the case of covariate adjustment, two models were developed for both (1) NE ~ EC and (2) NE ~ EC + SS + Dys for comparison purpose. The propensity score matching and inverse weighting used only the NE ~ EC + SS + Dys relationship.

2.3. Testing for robustness

The EC-NE causal relationship was tested for robustness using both a placebo refuter (DoWhy library) and Rosenbaum bounds. The former method replaces the treatment with random variables and re-estimates the causal effect on the new treatment - outcome relationship. The latter tested the sensitivity to unobserved confounding i.e. how strong a confounder needs to be to alter the causal relationship. This was quantified with an analogous to odds-ratio called the gamma parameter [6].

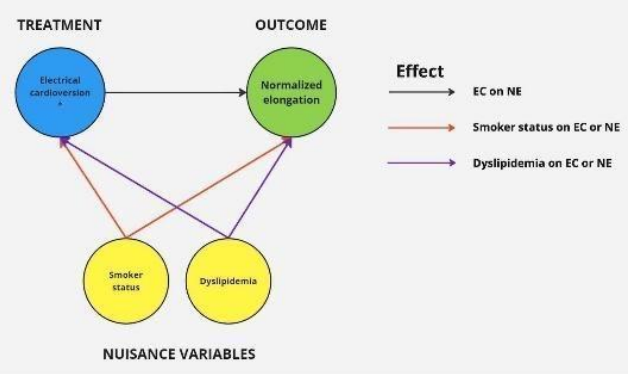

Figure 1. Causal directed acyclic graph for the normalized elongation-electrical cardioversion relationship

3. Results

3.1 Machine learning results

The Double*MLPLR* model estimated a significant causal coefficient of -0.36 (p=0.005) for the effect of EC on NE. The *DoWhy* model yielded similar results, with a causal coefficient of -0.36 ($p < 0.05$).

3.2 Propensity score and covariate adjustment results

The two models considered for covariate adjustment were in stark contrast. The model accounting for Dys and SS had an R squared of

0.25, with a p value of 0.09. Among the three variables included in the model (EC, SS, and Dys), the closest to significance was EC with a p value of 0.09. In contrast, the model accounting for EC only (eq. 1) had an R squared of 0.11 and a p value of 0.05. Skewness was computed to verify the normal distribution of NE, required for these ordinary least square models. Both models had a skew near 0.

The propensity score methods had a similar pattern. Using the matching method, EC was estimated to an average treatment effect (ATE) of -0.24, but with a p value of 0.16. On the other hand, inverse weighting had a calculated ATE of -0.38 with a p value of 0.025.

Table 2 shows a comprehensive review of the results.

Table 2. Methods and results of assessing NE-EC causality

Method for assessing causality	Variable testing the NE-EC relationship	Value of the variable	P value
Random forest with partial linear regressor (DoubleML)	Causal coefficient	-0.36	0.005
Linear regressor model (DoWhy)		-0.36	0.05
Covariate adjustment: log(NE)* ~ EC + SS + Dys (least square error)	R^2	0.25	0.09
Covariate adjustment: log(NE)* ~ EC (least square error)	R^2	0.11	0.05
Propensity score: matching	Average treatment effect (ATE)	-0.24	0.16
Propensity score: Inverse weighting		-0.38	0.025

3.3 Robustness results

The placebo refuter reduced the effect to -0.015 (p=0.86), indicating that the observed effect is unlikely to arise from random variation.

The gamma sensitivity parameter was 3.1, indicating that an unobserved confounder would need an odds ratio of at least 3.1 to negate the observed effect.

4. Discussion and Limitations

Our work showcases how multiple machine learning and statistical models can be used jointly to assess for causal relationships in observational studies. Thus, we made use of a random forest machine learning partial linear regression, a machine learning linear regressor, propensity scores and covariate adjustment models as well as statistical tests of robustness (Rosenbaum bounds and a placebo refuter). Consistency among these tests is a framework to select which variables from an observational study could be causal. Therefore, resources for future randomized control trials could be directed towards testing variables which are highly associated on statistical tests and consistently returning as causal within the statistical and machine learning causal framework.

A key limitation is our small cohort (n=32). Covariate adjustment and propensity score modelling are constrained by the limited sample size. The covariate adjustment requires at least 10 patients per feature, 30 being the lower limit of required for our cohort to calculating an R^2 and its p value. The propensity score model has the same shortcoming as one feature requires at least 10 patients, with the matching process providing 15 patients per class (cardioverted/non-cardioverted). All in all, a larger sample could provide more accurate covariate adjustment and propensity score models.

5. Conclusion

The EC–NE relationship is consistently supported by causal inference machine learning and statistical methods in a clinical data framework.

References

[1] Andrei Alexandru Mircea, Agathe Pusturle de Cidrac, Adrian Luca, Cheryl Teres, Patrizio Pascale, Mathieu Le Bloa, Ciro Ascione, Giulia Domenichini, Mattia Pagnoni, Alessandra Pia Porretta, Cosima Jahnke, Kerstin Bode, Sabrina Oebel, Frank Lindemann, Panagiotis Antiochos, Ambra Masi, Jorge Solana-Muñoz, Juerg Schwitter, Etienne Pruvot, Ingo Paetsch, Correlation of cavotricuspid isthmus dynamics with clinical parameters: insights from interventional cardiac magnetic resonance imaging, *European Heart Journal - Imaging Methods and Practice*, Volume 3, Issue 4, October 2025, qyaf106, https://doi.org/10.1093/ehjimp/qyaf106

[2] Shimonovich, M., Pearce, A., Thomson, H., Keyes, K., & Katikireddi, S. V. (2021). Assessing causality in epidemiology: revisiting Bradford Hill to incorporate developments in causal thinking. European journal of epidemiology, 36(9), 873–887. https://doi.org/10.1007/s10654-020-00703-7

[3] Sharma, A., Kiciman, E., 2020. DoWhy: An End-to-End Library for Causal Inference. https://doi.org/10.48550/arXiv.2011.04216

[4] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., Robins, J., 2024. Double/Debiased Machine Learning for Treatment and Causal Parameters. https://doi.org/10.48550/arXiv.1608.00060

[5] https://ocw.mit.edu/courses/6-s897-machine-learning-for-healthcare-spring-2019/resources/mit6_s897s19_lec15/

[6] PAUL R. ROSENBAUM, Sensitivity analysis for certain permutation inferences in matched observational studies, Biometrika, Volume 74, Issue 1, March 1987, Pages 13–26, https://doi.org/10.1093/biomet/74.1.13

Address for correspondence:
Andrei Alexandru Mircea,
25 Rue Ernest Barble,
LU-1210, Luxembourg ville
Luxembourg
andrei-alexandru.mircea@chuv.ch