

Original Article

Open Access



Modeling treatment effect heterogeneity in prophylactic lumbar drainage: a Double Machine Learning reanalysis of EARLYDRAIN

Shrinit Babel^{1,2} , Syed RH Peeran³, Gandham E. Jonathan³

¹Morsani College of Medicine, University of South Florida, Tampa, FL, 33612, USA.

²Whiting School of Engineering, Johns Hopkins University, Baltimore, MD 21218, USA.

³Department of Neurosurgery, Christian Medical College, Vellore 632004, India.

Correspondence to: Shrinit Babel, University of South Florida Morsani College of Medicine, Tampa, FL 33612, USA. E-mail: shrinitbabel@usf.edu

How to cite this article: Babel S, Peeran SRH, Jonathan GE. Modeling treatment effect heterogeneity in prophylactic lumbar drainage: a Double Machine Learning reanalysis of EARLYDRAIN. *Art Int Surg*. 2025;5:545-56. <https://dx.doi.org/10.20517/ais.2025.75>

Received: 6 Aug 2025 **First Decision:** 15 Sep 2025 **Revised:** 30 Sep 2025 **Accepted:** 30 Oct 2025 **Published:** 15 Dec 2025

Academic Editor: Andrew Gumbs **Copy Editor:** Xing-Yue Zhang **Production Editor:** Xing-Yue Zhang

Abstract

Aim: The EARLYDRAIN trial illustrated that prophylactic lumbar drainage (LD) could reduce poor outcomes in patients with aneurysmal subarachnoid hemorrhage, although not uniformly. We aim to reanalyze trial data using Double Machine Learning (DML) to estimate individualized treatment effects and identify patients or patient subgroups most likely to benefit.

Methods: We applied a DML framework with causal Random Forests to data from 287 patients randomized in the EARLYDRAIN trial to prophylactic LD or standard care. Six binary outcomes were analyzed: vasospasm, cerebral infarction, infection, favorable six-month outcome [modified Rankin Scale (mRS) ≤ 2], functional independence [Glasgow Outcomes Scale-Extended (GOS-E) ≥ 5], and shunt dependency. Average treatment effects (ATEs) and conditional ATEs (CATEs) were estimated. Uniform Manifold Approximation and Projection of the CATE values identified treatment-response clusters.

Results: Across the full cohort, prophylactic LD showed no consistent ATE across outcomes [e.g., mRS ATE: 0.02; 95% confidence interval (CI): -0.13-0.17]; CATE distributions revealed significant heterogeneity, with four treatment-response phenotypes. Younger patients with elevated intracranial pressure and lower drainage volumes derived greater benefit, and potentially reduced vasospasm risk. Patients older than 60 with higher systolic blood



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



pressure, greater hemorrhage burden, and positive fluid balance experienced limited benefit and increased shunt dependency. A web-based application (<https://earlydrain.streamlit.app>) was developed to translate these findings into clinical decision support.

Conclusion: DML analysis of the EARLYDRAIN trial was able to substantiate the heterogeneity in effects of prophylactic LD from the initial trial. DML offers a scalable framework to reveal treatment heterogeneity masked in trial averages and support precision medicine in neurosurgery.

Keywords: Lumbar drainage, Double Machine Learning, early drain

INTRODUCTION

Aneurysmal subarachnoid hemorrhage (aSAH) can result in devastating neurological sequelae and is associated with high morbidity and mortality. Even among patients who survive the initial aneurysm rupture and receive definitive treatment, either surgical clipping or endovascular coiling, outcomes vary: some recover well, whereas others develop complications such as cerebral vasospasm during their intensive care unit (ICU) stay or experience poor long-term neurological outcomes. Adjunctive clinical strategies are therefore critical for preventing short-term complications and improving long-term functional recovery.

Prophylactic lumbar drainage (LD) has been increasingly considered in the post-treatment phase of aSAH to reduce intracranial pressure (ICP), promote blood clearance from the subarachnoid space, and ultimately decrease the incidence of secondary insults by providing increased cerebral oxygenation and blood flow. The EARLYDRAIN trial was a multicenter randomized controlled trial (RCT) that enrolled adult patients with aSAH across centers in Germany, Switzerland, and Canada to evaluate the role of prophylactic LD^[1]. Particularly, patients were randomized to receive either prophylactic LD initiated within 48 hours after aneurysm treatment or standard of care alone.

The EARLYDRAIN analysis demonstrated that prophylactic LD was associated with a significant reduction in poor neurological outcomes across the entire cohort at six months, as measured by the modified Rankin Scale (mRS 3-6 at six months with LD: 32.6% vs. 44.8%, $P = 0.04$).

A meta-analysis of 14 studies involving more than 2,400 patients demonstrated that prophylactic LD was associated with significantly lower rates of symptomatic vasospasm and other early complications^[2]. However, its impact on long-term functional outcomes was inconsistent. Another prospective trial of 210 patients (the LUMAS trial; Lumbar Drainage in Subarachnoid Haemorrhage; ClinicalTrials.gov identifier NCT00842049) similarly demonstrated reduced delayed ischemic neurological deficits and improved early clinical outcomes with prophylactic LD placement. However, there were no significant improvements in long-term outcomes at six months^[3].

Furthermore, LD itself can lead to complications. A retrospective study comparing serial lumbar punctures to prophylactic LD found that the LD group had a significantly higher risk of meningitis^[4]. There is also the risk of lumbar overdrainage leading to low-pressure headaches, and in rare cases, intracranial (ICH) or spinal hemorrhage. For example, a case study described syringomyelia following lumbar cerebrospinal fluid (CSF) overdrainage in the acute phase of subarachnoid hemorrhage (SAH)^[5].

Rather than universally placing LDs prophylactically in all patients with aSAH, it may be more effective to identify which individuals are most likely to benefit and which may be at greater risk for complications.

Quantifying the precise magnitude of benefit for each patient can also support decision-making. The treatment effect of prophylactic LD is heterogeneous, with some subgroups deriving greater benefit than others depending on individual clinical characteristics such as baseline ICP, CSF dynamics, and clot burden^[6].

To identify and quantify these individualized treatment effects, it is important to account for confounding factors in observational or complex clinical trial data. There are limited RCT-based subgroup analyses to date, and traditional subgroup analyses are often limited in granularity and may lack the power to detect heterogeneous effects. Conversely, conventional predictive analytics and machine learning techniques can capture flexible patterns in data, but they cannot distinguish between causation and correlation.

Double Machine Learning (DML) is an emerging technique that integrates the strengths of both machine learning and causal insights to predict treatment effects^[7]. Originally developed for econometrics, DML was introduced by Chernozhukov et al. as a technique using orthogonalization techniques to control for confounding regardless of the data collection technique. This can ultimately be used to estimate both the average treatment effect (ATE) over the entire cohort and the conditional ATE (CATE), which reflects treatment impact for individual patients based on specific characteristics.

Capable of modeling heterogeneous treatment effects within a cohort, DML has numerous applications within clinical research, including intensive care outcomes, treatment selection for cardiovascular disease, and optimizing interventions in oncology^[8]. In neurosurgery, where treatment effects vary greatly across patient-level factors in neurophysiology and hemodynamics, DML can provide a scalable approach to harness past data towards decision-making.

In this study, we aim to apply a DML framework and reanalyze the EARLYDRAIN trial to estimate the individualized effects of prophylactic LD on various clinically relevant outcomes. We hypothesize that the role of prophylactic LD is not uniform, and DML will help model the patient-specific treatment effects to better inform individualized care.

METHODS

The methodology used in this study incorporates a Double Machine Learning (DML) framework. All code was written using Python 3.12.0 (Python Software Foundation), and the code is available. The hypothesized causal relationship we modeled is illustrated in a directed acyclic graph (DAG) shown in [Figure 1](#).

Data collection and preprocessing

The EARLYDRAIN dataset used in this study was sourced from Mendeley Data, an open-access dataset repository. The dataset contains information for 287 adult patients collected from 2011 to 2016. We used both the baseline clinical dataset and the dynamic ICU physiological data, which incorporated a comprehensive set of demographic, clinical, radiologic, treatment, and outcome variables. These included demographics, clinical severity scores on admission, imaging characteristics, treatment information such as coiling or clipping, use of nimodipine, magnesium, or statins, time to treatment or imaging, and admission neurological findings related to paresis, aphasia, and sedation status. Daily-level mean ICP at 7 a.m. during the first days post-SAH, mean arterial pressure (MAP), systolic and diastolic blood pressures, hemoglobin level, and net fluid balance, as well as daily and cumulative CSF drainage volumes via initial lumbar drain, external ventricular drain (EVD), or both, were also included. Specific feature descriptions are provided in the supplementary README file on GitHub, which can be accessed via https://github.com/shrinitbabel/Precision_Medicine_EARLYDRAIN_DML.

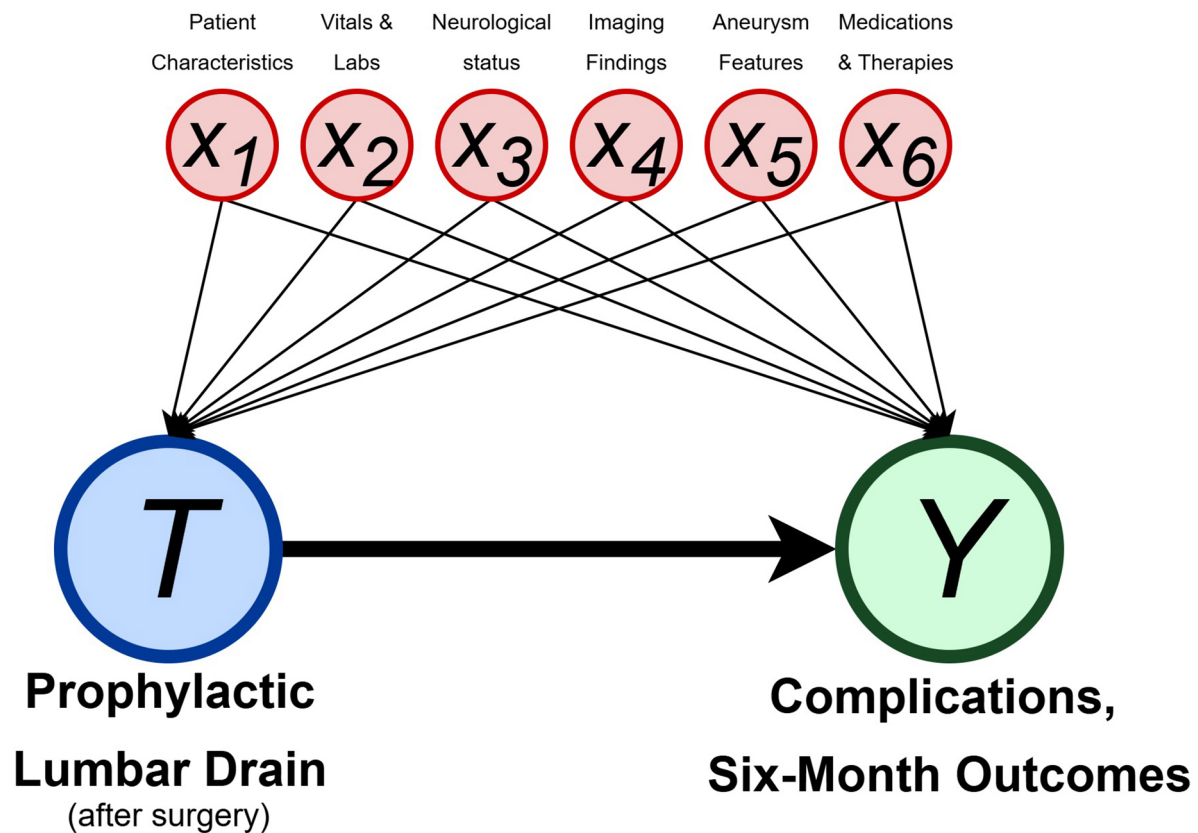


Figure 1. Directed acyclic graph (DAG) illustrating the causal model used in this study. The treatment variable T represents whether a patient received a prophylactic lumbar drain after aneurysm repair. The outcome Y represents outcome variables, including those related to the ICU course and six-month functional outcomes. Covariates are represented from x_1 to x_6 , and include demographics, vitals and labs, neurological status, imaging findings, aneurysm characteristics, and concomitant medications or therapies. These covariates influence both treatment assignment and outcomes, and are adjusted for using Double Machine Learning (DML) to estimate unbiased treatment effects. ICU: Intensive care unit.

We defined six binary clinical outcomes, including short-term ICU-related complications and six-month outcomes: (1) clinical vasospasm, (2) infection at discharge, (3) cerebral infarction, (4) Modified Rankin Scale (MRS) ≤ 2 at six months, (5) Glasgow Outcome Scale-Extended (GOS-E) ≥ 5 at six months (indicating functional independence), and (6) shunt dependency at six months (indicating increased morbidity or insufficient intervention). MRS was preferably dichotomized due to the small sample size, although an additional experimental CATE model was also generated in the pipeline modeling MRS as an ordinal scale.

Missing data were handled using a multivariate Bayesian ridge regression model with iterative imputation from scikit-learn, allowing patients with partially missing physiologic measurements to be retained. Continuous variables were standardized, and categorical features were encoded as one-hot or binary vectors. There were a total of 26 covariates. Intention-to-treat assignments were used for the treatment variable. Data from repeated daily measurements were standardized into patient-level features by computing the moving mean value across each patient's ICU stay.

Double Machine Learning analysis

To estimate individualized treatment effects of prophylactic LD on each outcome, we used the CausalForestDML method from the EconML Python library to orthogonalize and cross-fit the nuisance models [Figure 2]. The nuisance functions were estimated using nonparametric random forests, which were

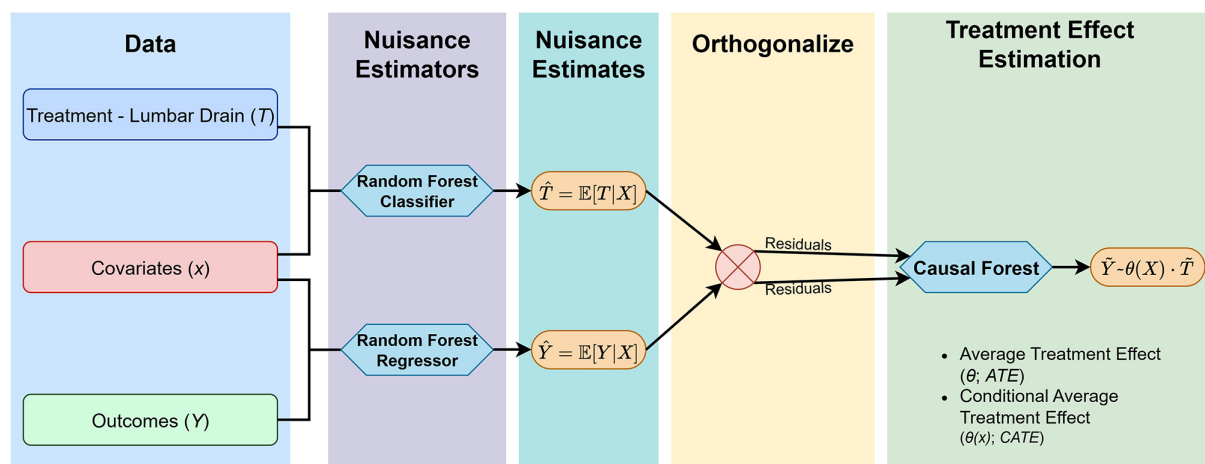


Figure 2. Double Machine Learning (DML) workflow. The pipeline begins with input data comprising treatment assignment, covariates, and outcomes. Nuisance functions for treatment and outcome models are estimated using random forests. These predictions are then orthogonalized by computing residuals, which are then passed to a causal forest algorithm that estimates both the average treatment effect (ATE), and the conditional average treatment effect (CATE).

especially preferred for their flexibility and frequent use with clinical data. All models were fit using 1,000 estimators with a maximum depth of 10 and a minimum of 10 samples per leaf. For each binary outcome, we estimated the ATE and CATE across individuals. The ATE and its 95% confidence interval (CI) were obtained, and we visualized the CATE distribution for each outcome. Feature importances for the causal forest model were also recorded. To assess the stability of the ATE estimates, we used an 80:20 train-test split to perform internal validation and compare ATEs obtained from the test set to those derived from the full dataset. A Spearman rank correlation was used to quantify the consistency in outcomes, with a significance threshold of 0.05 used to signify stability.

To explore heterogeneity in treatment response, we clustered patients based on their CATE vectors across multiple outcomes. For this step only, outcomes representing adverse outcomes - vasospasm, secondary infarction, infection, and six-month shunt dependency - were inverted so that a higher CATE consistently reflected net patient benefit. These were incorporated into a patient-by-outcome matrix, which was normalized, followed by dimensionality reduction using Uniform Manifold Approximation and Projection (UMAP). K-means clustering was applied to the embedding, and the optimal number of clusters was determined by comparing silhouette scores; the configuration with the highest score, indicating the best cluster separation, was selected.

RESULTS

In the 287-patient cohort, the mean age was 55.5 ± 12.4 years, with 69% being female. Aneurysms were predominantly located in the anterior circulation (85%), and 51% of cases were treated with endovascular coiling, while the remainder underwent surgical clipping. In addition to the randomized administration of prophylactic LD, nearly all patients received nimodipine (99%), while 16% were on statins, and 54% received magnesium during their ICU stay. The total percent missing across the dataset was 7.3%, with height being the most affected covariate (16.4%).

Stability of the treatment effect estimates was validated through a post-hoc comparison between ATEs from a held-out test set to those from the full dataset. The Spearman correlation between these two estimates across outcomes was $\rho = 0.82$ ($P = 0.044$), suggesting strong consistency in ranking and reliability of the

model. A scatterplot comparison of this is available in the GitHub repository.

The CATE distributions are illustrated in [Figure 3](#). Across all outcomes, prophylactic LD did not demonstrate a consistent or universal benefit; its effect varied across the cohort. The estimated ATEs of prophylactic LD were -0.08 for vasospasm (95%CI: -0.24 to 0.08), -0.049 for secondary infarction (95%CI: -0.21 to 0.12), and 0.095 for infection (95%CI: -0.066 to 0.26). For six-month outcomes, the ATE was 0.02 for mRS ≤ 2 (95%CI: -0.13 to 0.17), 0.01 for functional independence (95%CI: -0.12 to 0.15), and -0.049 for shunt dependency (95%CI: -0.2 to 0.12). Across all outcomes, the ranking of covariate feature importances varied slightly, although several variables consistently emerged as the most influential. These included age, net fluid balance, 7 a.m. and peak ICP recordings, systolic blood pressure, mean CSF drainage, and Body Mass Index (BMI). These features contributed substantially to the heterogeneity in treatment response and were important to the learned CATE models.

The clustering analysis revealed four distinct treatment response phenotypes with a silhouette score of 0.42, suggesting adequate cluster separation [[Figure 4A](#)]. Each cluster was characterized by distinct clinical and physiological profiles and ultimately different patterns of benefit or harm from prophylactic LD [[Table 1](#)]. Cluster 0 consisted of relatively younger patients (median age, 52 years old) with moderate Hunt and Hess scores, lower ICP values. Patients in this cluster had the most consistent positive treatment effects across outcomes. CSF drainage was also highest overall. Cluster 1 comprised the oldest group of patients (median age, 59 years old) with the highest systolic blood pressures, increased radiographic Fisher severity, and a higher prevalence of intraventricular (IVH) and ICH hemorrhage. These patients exhibited less benefit, with a negative mean CATE for functional independence (-0.02) and an elevated risk of shunt dependency (+0.1). Cluster 2 had the highest fluid balance, although it had a similar radiographic Fisher severity score as Cluster 1. While they showed some benefits regarding infarction prevention, they also reduced the potential benefits and harm of vasospasm and infection. Cluster 3 was the youngest group with the lowest Hunt and Hess scores and the highest ICP values, yet it had the lowest CSF drainage volumes. This group had favorable effects across long-term outcomes, although a less pronounced short-term benefit. Notably, this group had the most significant reduction in shunt dependency.

The CATE framework also enables visualization of outcome trade-offs and CATE values across different covariates. [Figure 4B](#) illustrates the trade-off between reducing vasospasm risk and achieving a gain in functional independence (GOS-E ≥ 5) at six months. Patients in Cluster 1 tended to experience consistent benefits in terms of reduced vasospasm risk but showed minimal improvement in long-term functional independence. In contrast, Cluster 3 demonstrated more favorable trade-offs, with most patients exhibiting both reduced vasospasm and an increased likelihood of functional independence. Similarly, [Figure 5](#) illustrates how the treatment effect of prophylactic LD on vasospasm risk varies as a function of mean CSF drainage and peak ICP. Patients with low CSF drainage and higher ICP experienced the most significant reduction in vasospasm risk with prophylactic LD. Conversely, those with high CSF output and lower peak ICP experienced minimal benefit or even increased risk for vasospasms, potentially due to over-drainage.

Finally, we developed a web-based application using Streamlit, designed to translate the causal findings (available at: <https://earlydrain.streamlit.app>). This tool allows users to input patient-level characteristics and obtain individualized estimates of the treatment effect of prophylactic LD across multiple outcomes [[Figure 6](#)]. The interface displays the predicted benefit or harm of prophylactic LD based on the learned CATEs, allowing users to visualize how the causal effect changes with various parameters. The risks and benefits are displayed as a percentage, which can be interpreted as the absolute change in the probability of an outcome due to the treatment, or an absolute risk difference.

Table 1. Cluster characteristics

Feature	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Size	61	78	81	67
Age [*]	51 (46, 60)	59 (51, 66.75)	60 (53, 70)	50 (45, 53)
BMI	25.71 (22.9, 27.8)	25.6 (23.0, 27.6)	24.7 (23.4, 27.9)	24.9 (22.5, 27.0)
IVH on CT ^{**} , %	28 (45.9%)	59 (75.6%)	59 (72.8%)	29 (43.2%)
ICH on CT ^{**} , %	12 (19.6%)	42 (53.8%)	34 (42.0%)	18 (26.9%)
Paresis on admission, %	17 (27.9%)	20 (25.6%)	40 (49.4%)	10 (14.9%)
Aphasia on admission ^{**} , %	19 (31.1%)	23 (29.5%)	43 (53.1%)	10 (14.9%)
On nimodipine, %	61 (100%)	77 (98.7%)	79 (97.5%)	67 (100%)
On statin, %	9 (14.8%)	10 (12.8%)	15 (18.5%)	12 (17.9%)
On magnesium ^{***} , %	32 (52.5%)	34 (43.6%)	55 (67.9%)	35 (52.2%)
Mean ICP mm Hg [*] (peak)	13.6 (11, 15.6)	19.75 (16.6, 24.2)	13.2 (11.5, 15.5)	21.3 (17.7, 24.6)
Hunt and Hess breakdown	Grade 1: 15 Grade 2: 20 Grade 3: 13 Grade 4: 8 Grade 5: 5	Grade 1: 11 Grade 2: 12 Grade 3: 16 Grade 4: 18 Grade 5: 21	Grade 1: 11 Grade 2: 16 Grade 3: 14 Grade 4: 11 Grade 5: 29	Grade 1: 17 Grade 2: 21 Grade 3: 16 Grade 4: 7 Grade 5: 6

Age, BMI, and mean ICP are reported as median (interquartile range); categorical variables are reported as *n* (%). ^{*}Significant difference across clusters for continuous variables by Kruskal-Wallis test (all $P < 0.001$); ^{**}Significant difference across clusters for categorical variables by chi-square test (IVH on CT, ICH on CT, paresis on admission, aphasia on admission; all $P < 0.01$); ^{***}Significant by chi-square test by $P < 0.05$. Nimodipine and statin use did not differ significantly between clusters ($P = 0.39$ and 0.75 , respectively). BMI: Body mass index; ICP: intracranial pressure; IVH: intraventricular hemorrhage; CT: computed tomography; ICH: intracranial hemorrhage.

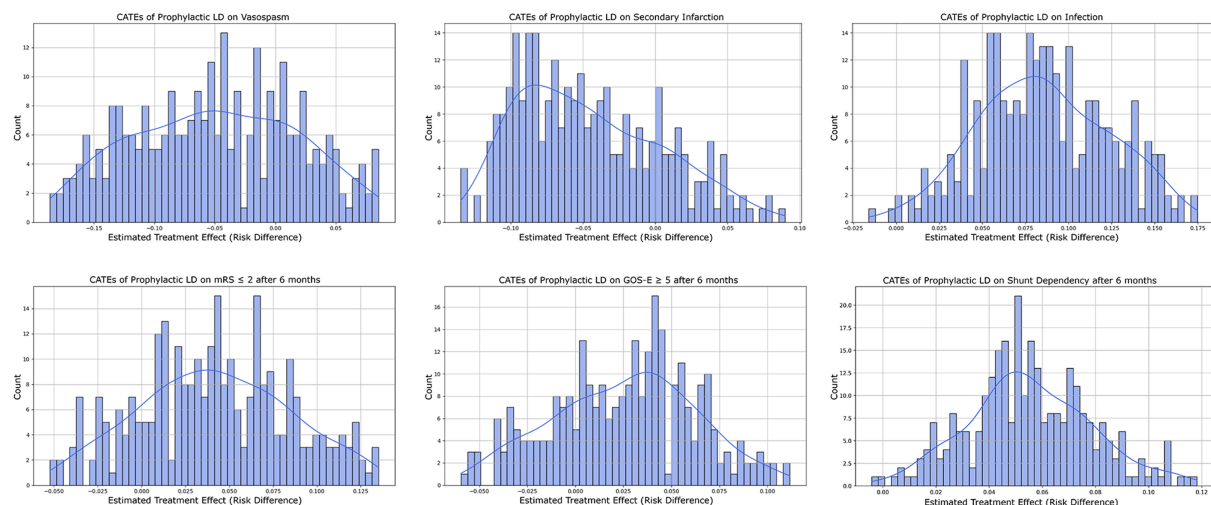


Figure 3. Conditional average treatment effect (CATE) distributions for prophylactic lumbar drainage (LD) across six clinical outcomes. Each panel displays the estimated treatment effect (risk difference) from the Causal Forest model. Outcomes include: vasospasm, secondary infarction, and infection during ICU course (top row); and modified Rankin Scale ($mRS \leq 2$), Glasgow Outcome Scale-Extended ($GOS-E \geq 5$), and shunt dependency at six months (bottom row). These distributions suggest heterogeneity in treatment effects across patients. ICU: intensive care unit.

DISCUSSION

The goal of our study was to provide a detailed understanding of the effect of prophylactic LD in the setting of aSAH using data from the EARLYDRAIN trial. Traditional RCTs and standard machine learning approaches often fail to reveal patient-specific treatment effects. However, DML bridges the gap by simulating counterfactual scenarios: it estimates how an individual patient's outcome would change if they received a different treatment. This allows for precise estimation of individualized treatment effects.

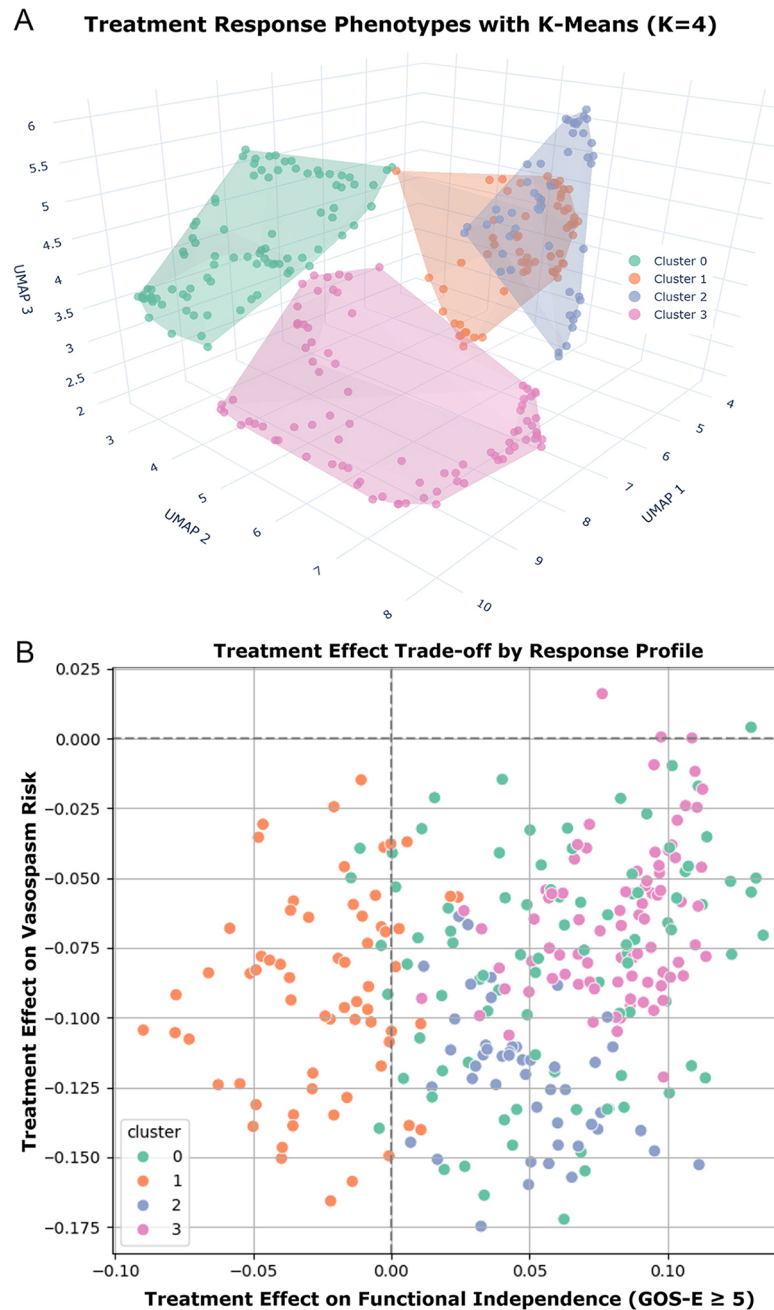


Figure 4. Clustering and trade-off analysis of treatment response phenotypes using CATEs across outcomes. (A) Clustering of patients based on their CATE vectors for all six outcomes, visualized in a UMAP-reduced space, revealed four distinct treatment response phenotypes (K-Means, K = 4); (B) Trade-off plot showing the relationship between the estimated treatment effect of prophylactic LD on vasospasm risk (y-axis) and six-month functional independence (GOS-E $\geq$ 5, x-axis). Each point represents an individual patient, colored by cluster membership. Cluster 1 demonstrated consistent vasospasm reduction without improvements in functional outcome, while Cluster 3 exhibited favorable trade-offs across both dimensions. Clusters 0 and 2 demonstrated a similar trade-off profile as Cluster 3, albeit with greater heterogeneity. CATEs: Conditional average treatment effects; CATE: conditional average treatment effect; UMAP: Uniform Manifold Approximation and Projection; LD: lumbar drainage; GOS-E: Glasgow Outcome Scale-Extended.

Therefore, we leveraged this strategy to model causal effects on six outcomes: three short-term outcomes (vasospasm, secondary infarction, and infection) and three six-month outcomes (mRS, GOS-E, and shunt dependency).

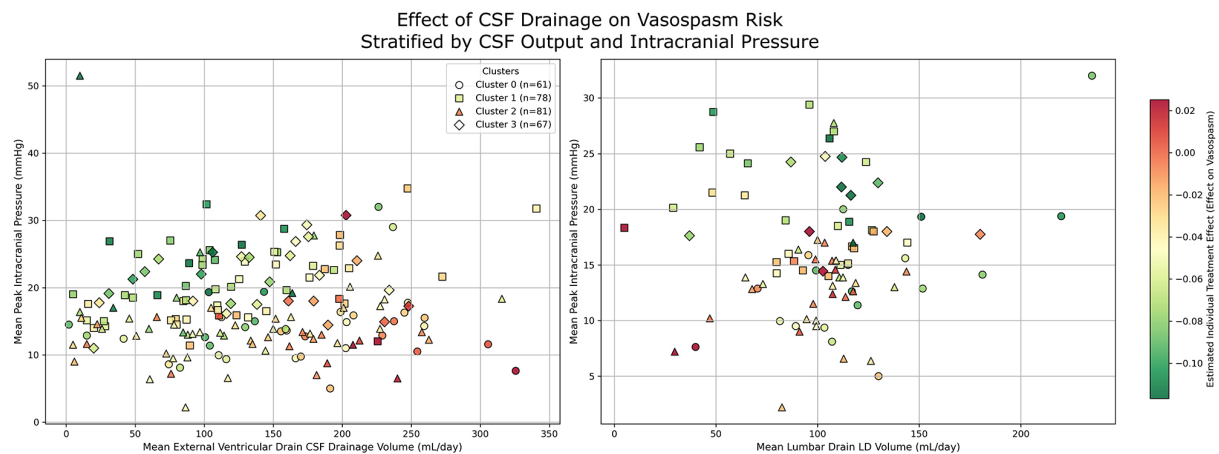


Figure 5. Individualized treatment effect of CSF drainage on vasospasm risk, stratified by mean CSF drainage volume and peak intracranial pressure (ICP) across external ventricular drain (EVD) (left) and prophylactic lumbar drain (LD) (right). Each point represents a patient, with color indicating the estimated treatment effect (CATE) on the risk of vasospasm. Green (negative CATE) represents a reduced risk of vasospasm, while red (positive CATE) represents an increased risk. Cluster relationships are also illustrated (circle = cluster 0; square = cluster 1; triangle = cluster 2; rhombus = cluster 3). CSF: Cerebrospinal fluid.

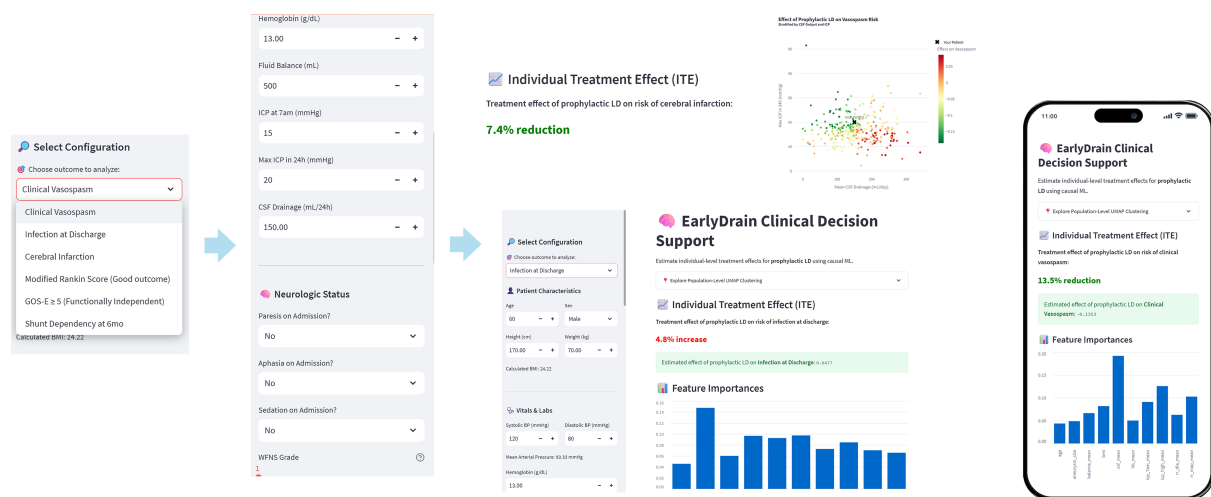


Figure 6. EARLYDRAIN Clinical Decision Support web application workflow (<https://earlydrain.streamlit.app>). The tool enables users to input patient-level characteristics and calculate individualized treatment effects (ITEs) for prophylactic lumbar drainage across multiple outcomes. Outputs include the estimated effect (expressed as absolute risk reduction or increase), ranked feature importances, and interactive visualizations of treatment effect stratified by key physiological parameters.

While the ATEs across the six outcomes demonstrated no consistent benefit of prophylactic LD across the entire cohort, the CATE distributions suggested substantial heterogeneity in treatment response. The clustering analysis identified distinct subgroups that experienced unique treatment responses, either beneficial or detrimental, depending on their baseline characteristics.

A particular interest was our analysis of the vasospasm outcome. The primary purpose of prophylactic LD in the context of aSAH is to promote the clearance of blood in the subarachnoid space, a significant trigger for cerebral vasospasm and, subsequently, delayed cerebral ischemia (DCI). Furthermore, as it drains CSF, LD reduces ICP, which can help mitigate secondary brain injury by maintaining the cerebral perfusion pressure in a desired range. Expectedly, the treatment effect of prophylactic LD was primarily influenced by

the mean CSF drainage, peak daily ICP, and MAP at 7 a.m. Our analysis complements the original EARLYDRAIN trial results in that prophylactic LD has no meaningful improvement in vasospasm risk overall; with our causal analysis, we hypothesize that benefits may rather be observed in specific patients. The benefit of prophylactic LD in reducing the risk of vasospasm was more pronounced in patients with higher ICP and lower drainage output. Conversely, our analysis replicated the concept of overdrainage-induced cerebral vasospasm described in past studies, particularly in patients with high CSF drainage volumes and lower peak ICP^[9].

aSAH-induced vasospasm, even if clinically silent, is independently associated with increased long-term morbidity and mortality, especially cognitive decline, emotional limitations, and functional limitations^[10]. The CATE framework enabled us to quantify the causal relationship between treatment, vasospasm risk during the ICU course, and long-term functional independence (GOS-E ≥ 5), as shown in Figure 4B. Although nearly all patients appeared to benefit from prophylactic LD in terms of reduced vasospasm risk, for a subset of patients (Cluster 1), this did not translate into improvements in functional outcomes. The limited functional recovery in Cluster 1 may be attributable to their distinct clinical profile - patients aged around 60-63 years old, elevated systolic blood pressure, and higher incidence of IVH and ICH - all factors previously associated with impaired long-term recovery following aSAH, independent of vasospasm severity^[11-13]. The web-based application helps quantify the magnitude of benefit or harm, providing a scalable way to bridge the causal inference with personalized, data-driven decision-making at the bedside.

The primary endpoint of the EARLYDRAIN trial was a favorable six-month neurological outcome, defined as the proportion of patients with an mRS score of 0-2. The DML analysis complements and extends these findings, which were previously limited to the ATE across the entire cohort. While features such as the number or size of aneurysms, aneurysm circulation territory, and WFNS grade were less influential in modifying treatment response, physiological features - namely, net fluid balance, age, ICP burden, and blood pressure metrics - were key drivers of treatment effect heterogeneity. For example, patients older than 60 and those with a net positive fluid balance showed attenuated effects of prophylactic LD on a favorable six-month mRS; advanced age and volume overload might reduce the potential advantage of CSF diversion. This finding is corroborated by a past retrospective study, in which a higher net positive fluid balance during the ICU stay in the setting of aSAH, regardless of management choices, was independently associated with poor outcomes at follow-up (mRS 3-6)^[14].

There are some limitations with DML. The two-stage design of DML means that initial errors in estimating nuisance parameters (e.g., treatment or outcome models) can percolate in subsequent steps^[7]. Even with the orthogonalization step designed to mitigate this risk, the potential for propagating early-stage errors still exists. Another inherent limitation is that DML assumes overlap in covariate distributions between treated and untreated groups, thereby resulting in unstable estimates, especially when treatment assignment is imbalanced^[15]. Although this limitation is more relevant in observational datasets than in randomized datasets such as ours, the small sample size in our analysis may still contribute to variability in effect estimation. With the limited cohort size ($n = 287$) and number of covariates ($n = 26$), the risk of overfitting and unstable CATE estimates should be considered. Although the orthogonalization step can attempt to mitigate this, the treatment effect estimates, especially in smaller subgroups, should be interpreted with caution and as exploratory. Furthermore, our analysis treats LD as an isolated intervention, although in reality, it interacts with other components of standard care, including nimodipine, statins, and fluid management. Future work should extend our framework to a larger dataset and include a causal synergy component to identify the effect of many treatments.

In conclusion, our research utilized DML to analyze treatment effect heterogeneity for preventive LD in aSAH, finally revealing patient-level and subgroup-level variations in both short- and long-term responses. We converted these causal insights into an accessible, Streamlit-based web application that predicts personalized treatment impacts across various outcomes. This technology enables neurosurgeons to quantify the potential benefits and risks for each patient, facilitating visualization of outcome trade-offs and supporting data-driven decision-making. In addition to expanding the insights from the EARLYDRAIN study, we illustrate how DML might enhance trial or observational data to improve neurosurgical care.

DECLARATIONS

Acknowledgments

The authors would like to thank the EARLYDRAIN investigators for publishing their data in open access, which made this work possible.

Authors' contributions

Conceptualization, methodology, data acquisition, data analysis, data curation, drafting of the manuscript, figures and table preparation: Babel S

Supervision, manuscript review and editing, guidance on clinical interpretation: Peeran SRH

Supervision, manuscript review and editing, project oversight and support: Jonathan GE

Availability of data and materials

All analysis code used in this study is available on GitHub at: https://github.com/shrinitbabel/Precision_Medicine_EARLYDRAIN_DML.

The patient-level dataset used in this study is publicly available from the original EARLYDRAIN trial authors and can be accessed via Mendeley Data: <https://data.mendeley.com/datasets/c4jc7k5ptx/1>. DOI:10.17632/c4jc7k5ptx.1.

Financial support and sponsorship

None.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

This study analyzed only publicly available, de-identified data from the EARLYDRAIN randomized trial hosted on Mendeley Data (DOI:10.17632/c4jc7k5ptx.1). The dataset is released under a CC BY 4.0 license, which permits unrestricted reuse with attribution. No new data were collected, and no interaction with human participants occurred. Because the data are publicly available under an open license and contain no identifiable information, additional institutional ethics approval and individual informed consent were not required for this secondary analysis. We complied with all license terms and handled data in line with relevant guidelines.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2025.

REFERENCES

1. Wolf S, Mielke D, Barner C, et al; EARLYDRAIN Study Group. Effectiveness of lumbar cerebrospinal fluid drain among patients with aneurysmal subarachnoid hemorrhage: a randomized clinical trial. *JAMA Neurol*. 2023;80:833-42. DOI PubMed PMC
2. Lee KS, Chari A, Motiwala M, Khan NR, Arthur AS, Lawton MT. Effectiveness of cerebrospinal fluid lumbar drainage among patients with aneurysmal subarachnoid hemorrhage: an updated systematic review and meta-analysis. *World Neurosurg*. 2024;183:246-253.e12. DOI PubMed
3. Al-Tamimi YZ, Bhargava D, Feltbower RG, et al. Lumbar drainage of cerebrospinal fluid after aneurysmal subarachnoid hemorrhage: a prospective, randomized, controlled trial (LUMAS). *Stroke*. 2012;43:677-82. DOI PubMed
4. Liang C, Yang L, Guo S. Serial lumbar puncture reduces cerebrospinal fluid (CSF) infection during removal of hemorrhagic CSF in aneurysmal subarachnoid hemorrhage after endovascular coiling. *J Biomed Res*. 2018;32:305-10. DOI PubMed PMC
5. Machida A, Fujii M, Ishihara T, et al. Syringomyelia due to lumbar spinal fluid drainage in the acute phase of subarachnoid hemorrhage: a case report. *J Stroke Cerebrovasc Dis*. 2018;27:e11-4. DOI PubMed
6. Chen Q, Zhang S, Pei J. Effects of lumbar cerebrospinal fluid drainage on infection, rebleeding, clinical vasospasm, cerebral infarction, hydrocephalus and mortality among patients with aneurysmal subarachnoid hemorrhage. *Mol Cell Biomech*. 2025;22:1056. DOI
7. Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W. Double/Debiased/Neyman Machine Learning of treatment effects. *American Econ Rev*. 2017;107:261-5. DOI
8. Feuerriegel S, Frauen D, Melnychuk V, et al. Causal machine learning for predicting treatment outcomes. *Nat Med*. 2024;30:958-68. DOI PubMed
9. Kasuya H, Shimizu T, Kagawa M. The effect of continuous drainage of cerebrospinal fluid in patients with subarachnoid hemorrhage: a retrospective analysis of 108 patients. *Neurosurgery*. 1991;28:56-9. DOI PubMed
10. Bracard S, Schmitt E. Vasospasm and delayed consequences. *Interv Neuroradiol*. 2008;14 Suppl 1:17-22. DOI PubMed PMC
11. Pavelka M, Necarsulmer J, Ho J, Sasaki-Adams D. Vasospasm risk following aneurysmal subarachnoid hemorrhage in older adults. *J Neurosurg*. 2023;139:1302-10. DOI PubMed PMC
12. Macdonald RL, Rosengart A, Huo D, Karrison T. Factors associated with the development of vasospasm after planned surgical treatment of aneurysmal subarachnoid hemorrhage. *J Neurosurg*. 2003;99:644-52. DOI PubMed
13. Regula JU, Schill J, Ringleb PA, Sykora M. Cerebral vasospasm and delayed cerebral ischemia in intraventricular hemorrhage. *Neurocrit Care*. 2014;20:460-5. DOI PubMed
14. Kissoon NR, Mandrekar JN, Fugate JE, Lanzino G, Wijedicks EF, Rabinstein AA. Positive fluid balance is associated with poor outcomes in subarachnoid hemorrhage. *J Stroke Cerebrovasc Dis*. 2015;24:2245-51. DOI PubMed
15. Xu S, Zheng B, Su B, et al. Efficient estimation of weighted cumulative treatment effects by double/debiased machine learning. *arXiv* 2023;arXiv:2305.02373v1. Available from <https://www.semanticscholar.org/paper/Efficient-estimation-of-weighted-cumulative-effects-Xu-Zheng/753b15ce4fcb3a1806d9982e77aab4b63e0b7f> [accessed 4 December 2025].