



Machine Learning in Healthcare

Survival Analysis Group
Causal Inference Group

Big Data Summer Institute | University of Michigan
24 July 2025

Introduction

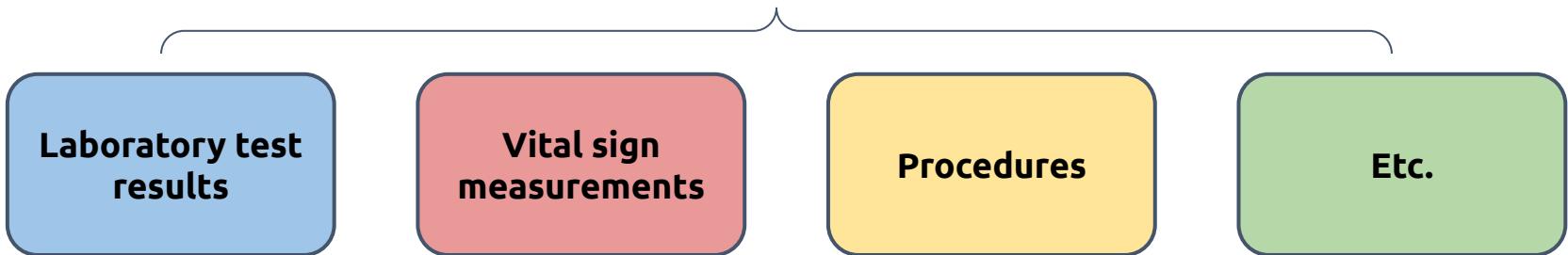
Background: Brain Bleeds

- Brain bleeds are severe medical conditions which can result in long-lasting complications or death
- Intracerebral Hemorrhages:
 - 18.9 million cases of intracerebral hemorrhages (ICH) occur annually [1]
 - Present a mortality rate of 40-50% in the first days following the event
- Patients discharged face a 22% mortality rate for 90-day survival [2]

Data Background

MIMIC-III

- “Medical Information Mart for Intensive Care” (3rd version)
- Health-related data from ~60,000 critical care patients from the Beth Israel Deaconess Medical Center in Boston, MA



Subset of interest: 1,330 brain bleed patients



Predicting Post-Discharge Survival Probability of Brain Bleed Patients

A Machine Learning Approach



Eliana Dietrich¹, Hanrui Yu², Joshua Brown³, Quinn Murphy⁴

Advisors: Dr. Rahul Ladhania⁵, Dr. Katherine Brumberg⁵, Abby Loe⁵

University of Washington¹, Peking University², Johns Hopkins University³, Hofstra University⁴, University of Michigan⁵



Our Goal

Predict the probability of survival at different times for patients post-discharge

Why?

- Improve patients' care & long-term outcomes
- Better inform providers in post-discharge intervention decisions

Examining Our Sample: Balances & Imbalances

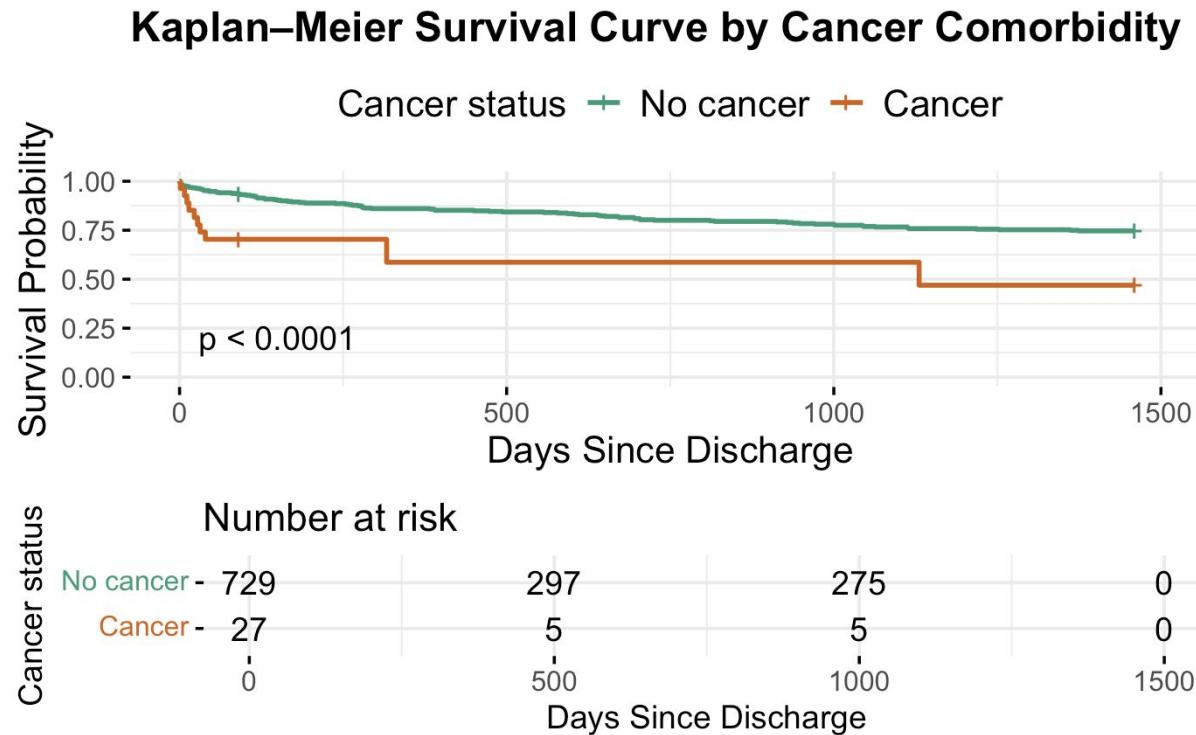
Table 1: Descriptive Statistics

Variables	N (%)
Gender	
Male	413 (54.6%)
Female	343 (45.4%)
Race/Ethnicity	
White	476 (63.0%)
Unknown	161 (21.3%)
Black or African American	57 (7.5%)
Hispanic/Latino	31 (4.1%)
Asian	26 (3.4%)
American Indian or Alaska Native	1 (0.1%)
More than one race	4 (0.5%)
Database Source	
Metavision	387 (51.1%)
Carevue	369 (48.9%)

Survival Analysis

What is a Kaplan-Meier Plot?

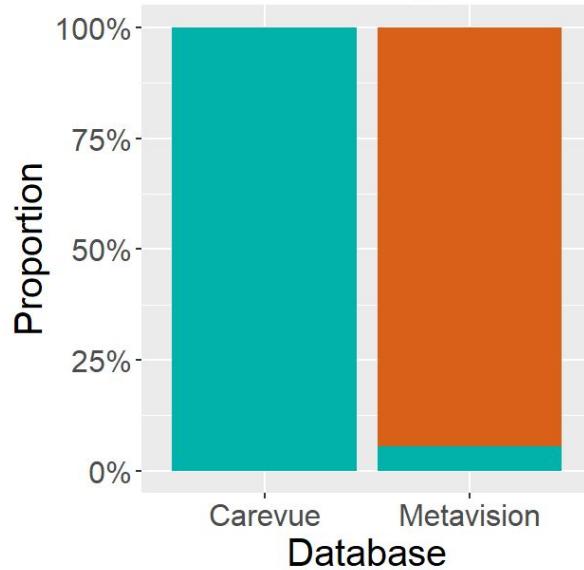
- **Kaplan-Meier Plot:** Calculates the probability of survival beyond a certain threshold
 - Survival trends & comparisons between groups
- Our threshold: t (days after discharge)



What is Censored Data?

- **Censored Data:** Observations where the time of the event is unknown
 - Patient lost to follow-up
 - Event occurs after study ends
- Censoring can lead to biased estimates of survival probabilities
 - Must account for censoring

Status of Patients by Record Type



Censor Status Before 4 Years

█ Censored
█ Non-censored

Formalizing Survival Data: Bivariate Survival Analysis

For each patient i we have : (X_i, δ_i)

- X_i - Length of observation
- δ_i - If an event was observed
- T_i - True time to event
- C_i - Censoring time

$$X_i = \min(T_i, C_i)$$

$$\delta_i = \mathbb{1}\{T_i \leq C_i\}$$

Our Scientific Question

$$P(T > t | \underset{\sim}{Z}) = g(\underset{\sim}{Z})$$

$$t \in \{7, 30, 60, 180, 365, 730\}$$

$\underset{\sim}{Z}$ = Predictors/Covariates

Methods

The Approach

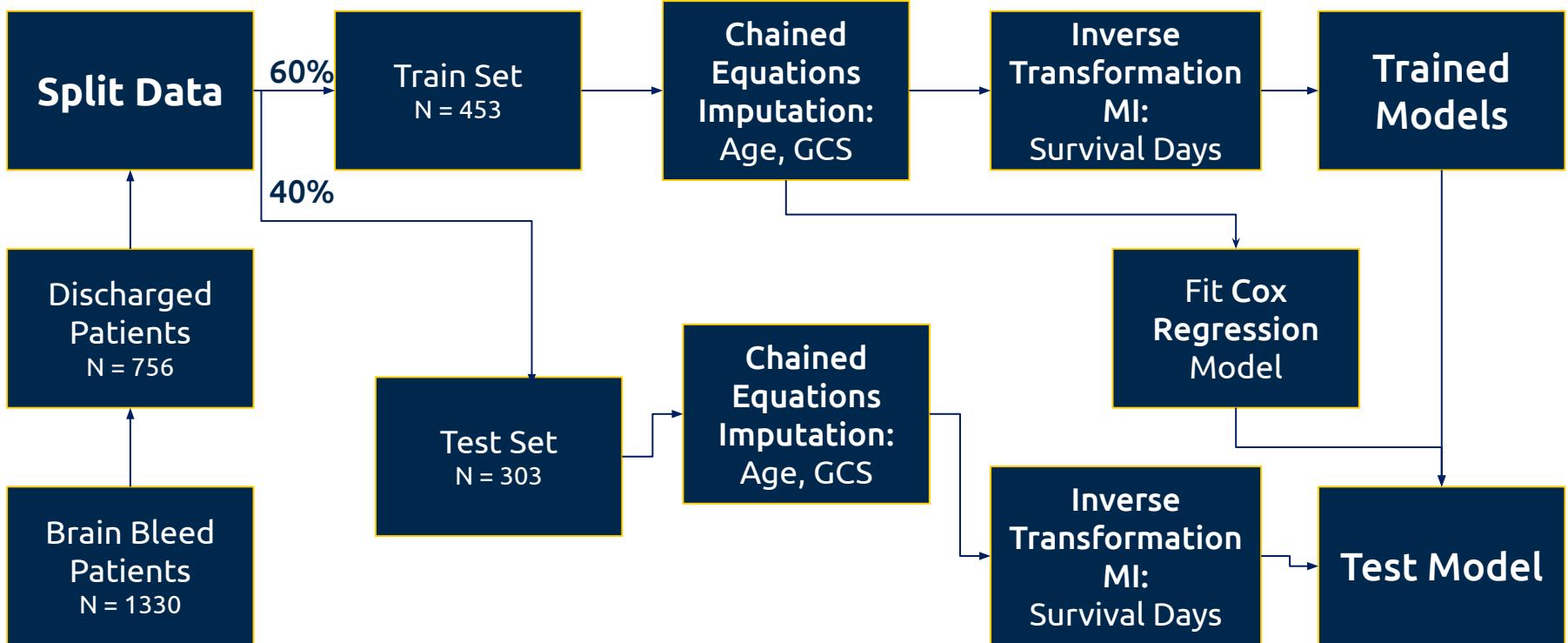
Multiple Imputation

Instead of approaching survival as a bivariate distribution, we will **impute** missing survival days using the Inverse Transformation Multiple Imputation algorithm

Who has missing survival days again..?

Metavision: censored **survival days** for patients after 90 days

The Approach



Machine Learning Methods

Baseline
Model

Cox Regression (All Predictors)

Random Forest (RF)

Cox Regression (Selected Predictors)

Machine Learning Methods

Baseline
Model

Cox Regression (All Predictors)

Random Forest (RF)

Cox Regression (Selected Predictors)

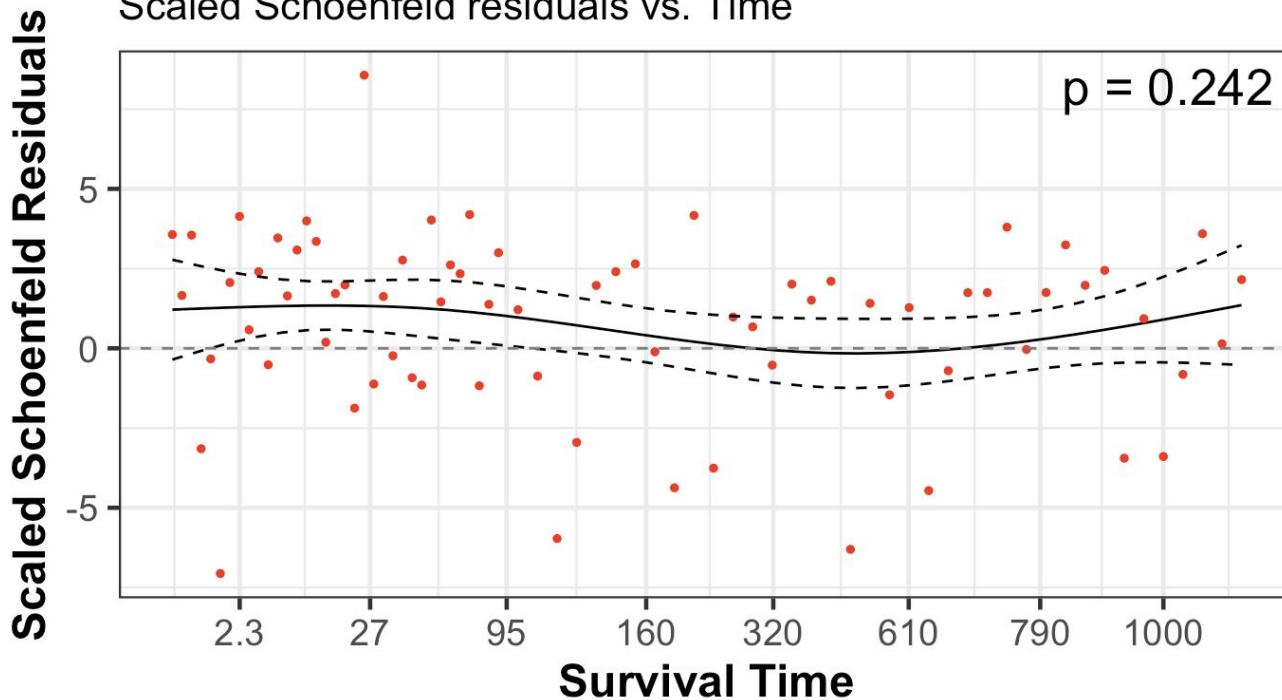
What is Cox Regression?

- Uses predictor variables to model hazard rate function
 - **Hazard rate:** risk of event occurring at specific time-point [\[3\]](#)
- Assumes proportional hazards for all predictors
 - **Proportional Hazards:** ratio of hazards between groups is proportional over time [\[4\]](#)

Proportional Hazards

Proportional-Hazards Check: Age At Admission

Scaled Schoenfeld residuals vs. Time



Cox Regression (All Predictors)

- Includes all predictors (**83 total**)
 - Baseline model
 - Commonly used in survival analysis
 - **Uses censored data**
-

$$\log\left(\frac{\lambda_i}{\lambda_0}\right) = \beta_1(\text{Age})_{i1} + \beta_2 \mathbb{1}(\text{Gender} = \text{Female})_{i2} + \cdots + \beta_n \mathbb{1}(\text{Cancer})_{i83}$$

Machine Learning Methods

Cox Regression (All Predictors)

Random Forest (RF)

Cox Regression (Selected Predictors)

Machine Learning Methods

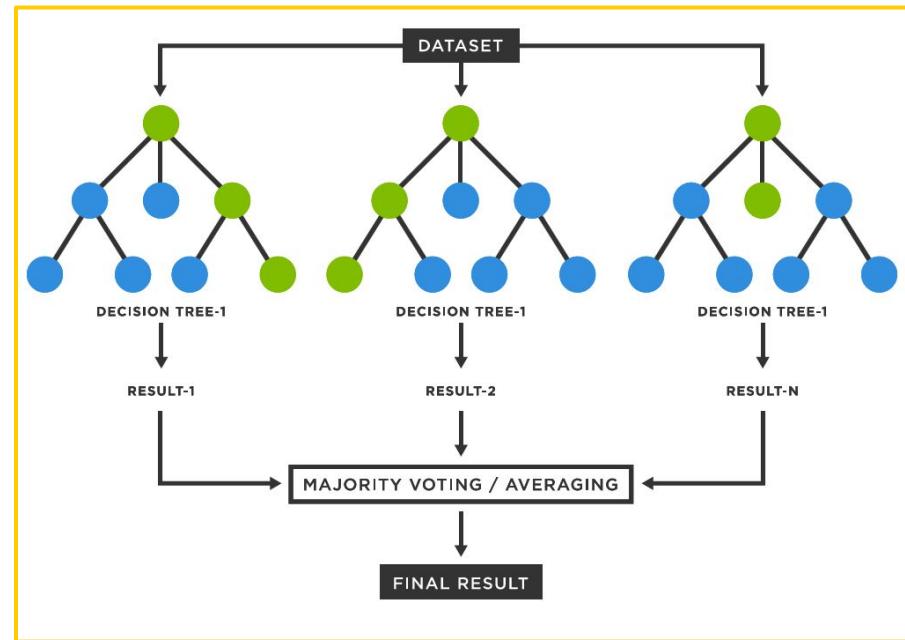
Cox Regression (All Predictors)

Random Forest (RF)

Cox Regression (Selected Predictors)

Random Forest (RF)

- Random samples of features used to train trees
- Can perform classification and show estimated event probability
- Uses Brier Score to determine best splits for each tree
- Is able to give variable importance
- **Uses imputed data**



Machine Learning Methods

Cox Regression (All Predictors)

Random Forest (RF)

Cox Regression (Selected Predictors)

Machine Learning Methods

Cox Regression (All Predictors)

Random Forest (RF)

Cox Regression (Selected Predictors)

Cox Regression (selected predictors)

- Includes selected predictors (**29 total**)
- **Uses censored data**
- Predictors selected by importance as determined by
Random Forest model
 - Prevents overfitting

Cox Regression (selected predictors)

Continuous Predictors

Age
Glasgow Coma Scale (GCS)

Glasgow Coma Score:

3-8: severe injury
9-12: moderate injury
13-15: mild injury



Religion

Catholic (ref.)
Other
Not Specified

ADMISSION LOCATION

Emergency Room (ref.)
Other
Clinic Referral/Premature

Insurance

Medicare (ref.)
Private
Other

Comorbidities

Cancer
Hepatitis C
Liver Cirrhosis
Chronic Kidney Disease
Chronic Pulmonary Disease

Discharge Location

Rehab/Distinct Part Hospital (ref.)
Other
Home

Language

English (ref.)
Spanish
Other

Procedures

Endocrine
Urinary
Cardiovascular
Nervous
Other

Prescriptions

Ophthalmic
Beta Blockers
Diuretics
Antiemetics
Antiepileptics

Metrics to Compare

C-index:

Measures ordering of predictions against ordering of actual survival time^[6]

- Good for categorization
- Does not reflect how well predicted probabilities are calibrated
- Very intuitive “who’s first” metric

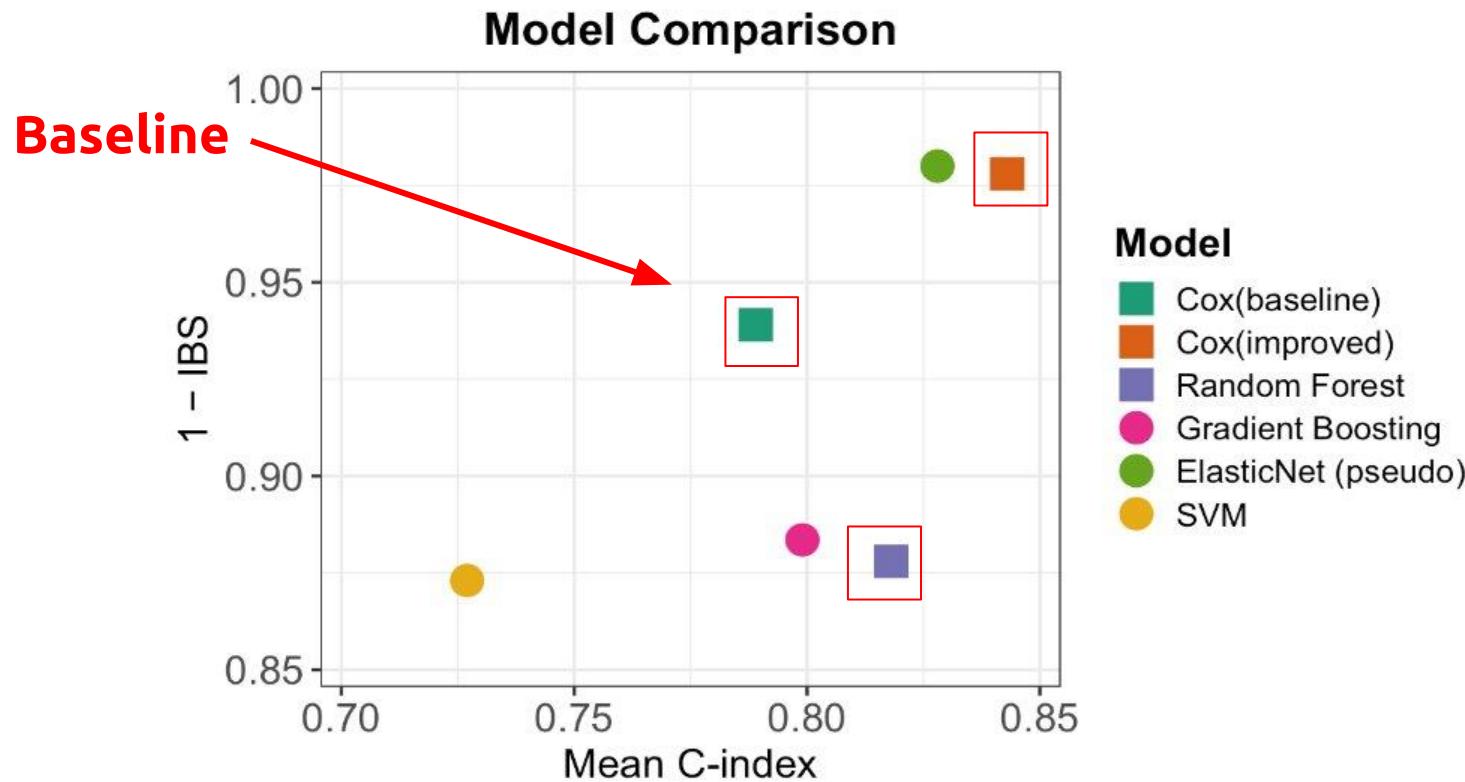
Integrated Brier Score (IBS):

Measures accuracy of survival probability predictions overtime^[7]

- Sensitive to systematic over or under-prediction
- No universal “good” threshold – must compare models
- Not that intuitive

Results

Model Metrics: A Visual Comparison



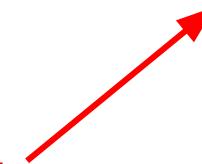
Model Metrics: C-Index & IBS

= 1 is best



Models	Mean C-Index	Integrated Brier Score
Cox Regression (All)	0.789	0.061
Cox Regression (Selected Predictors)	0.843	0.022
Random Forest	0.820	0.070

= 0 is best



Cox Regression (All Predictors): Significant Variables

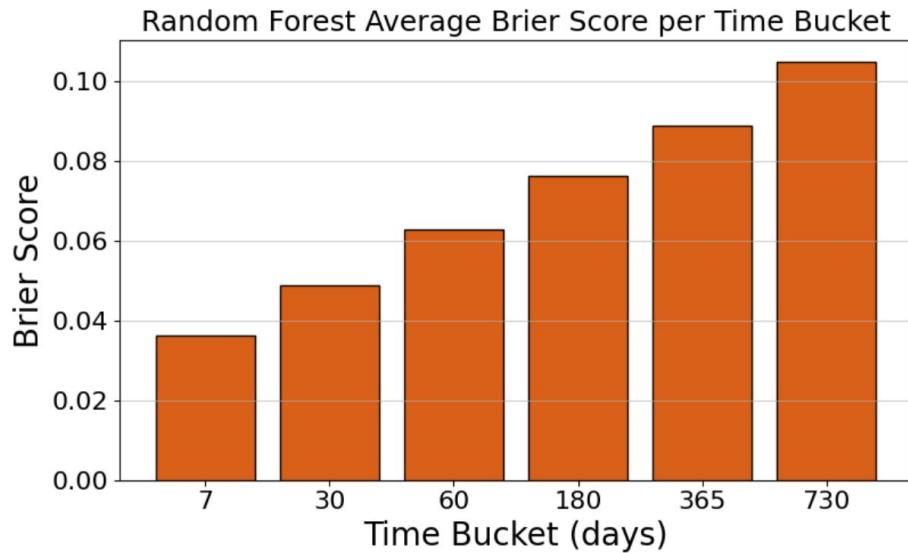
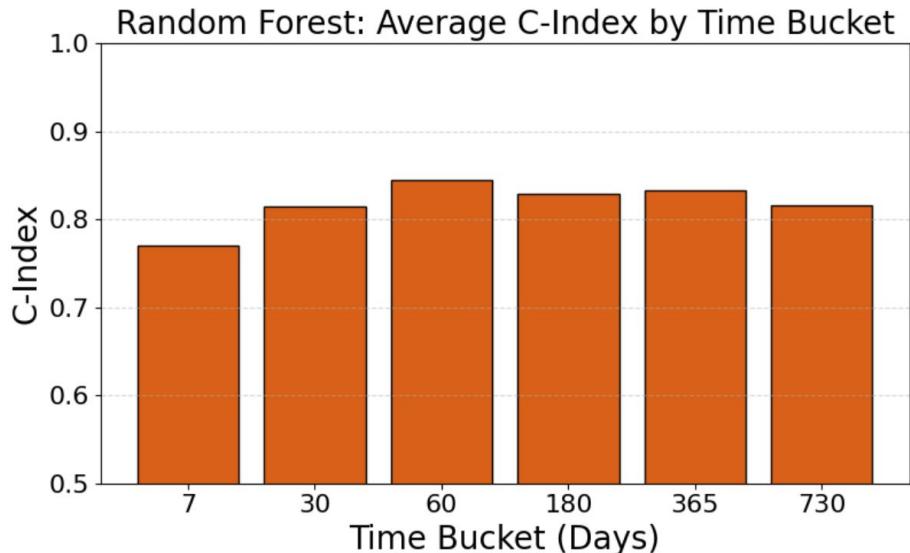
Variables (term)	Full Model $\exp(\hat{\beta})$ 95% CI
Glasgow Coma Scale Total***	0.32 (0.22, 0.48)
Comorbidity - Cancer**	1.88 (1.23, 2.86)
Age**	3.01 (1.16, 7.80)
Drug - Beta Blockers**	1.74 (1.09, 2.78)
Procedure - Endocrine**	1.32 (1.06, 1.64)
Comorbidity - Hepatitis C***	1.95 (1.35, 2.83)

$\exp(\hat{\beta}) < 1$ is associated with a decreased risk of death

p < 0.001 ***

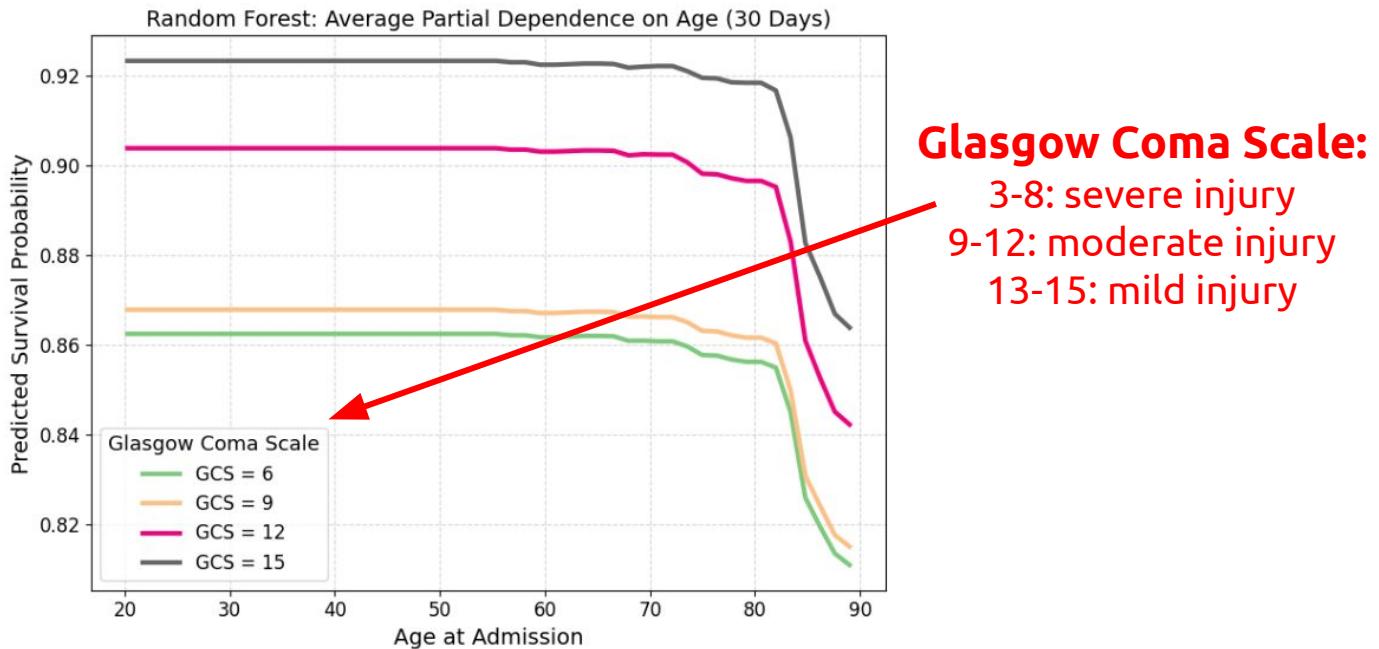
p < 0.05 **

Random Forest: C-Index vs. Brier Score



The model's ability to distinguish between patients with similar survival days (C-Index) is strong, and BS shows that predicted survival probabilities are less precise for longer predictions, but still strong.

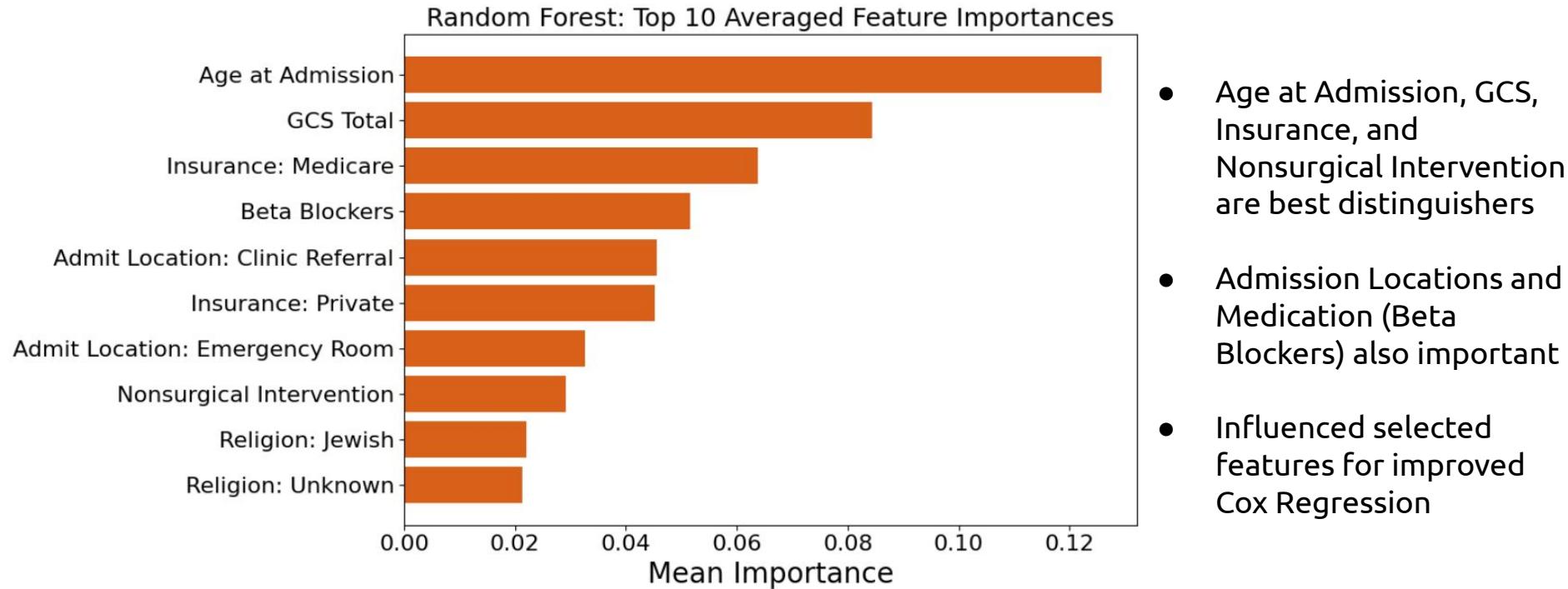
RF: Dependence Plots for Certain Features



Random Forests still leave some room for model interpretation, e.g., examination of the dependence of predicted survival probabilities on certain features for each patient

RF: Feature Importance

M



Cox Regression (Selected Predictors): Significant Variables

Variables (term)	Full Model	Selected Model
	$\exp(\hat{\beta})$ 95% CI	$\exp(\hat{\beta})$ 95% CI
Glasgow Coma Scale Total***	0.32 (0.22, 0.48)	0.50 (0.39, 0.65)
Comorbidity - Cancer***	1.88 (1.23, 2.86)	1.43 (1.17, 1.73)
Age**	3.01 (1.16, 7.80)	2.11 (1.21, 3.68)
Drug - Beta Blockers**	1.74 (1.09, 2.78)	1.39 (1.07, 1.81)
Procedure - Endocrine**	1.32 (1.06, 1.64)	1.21 (1.03, 1.42)
Comorbidity - Hepatitis C**	1.95 (1.35, 2.83)	1.26 (1.02, 1.55)

$\exp(\hat{\beta}) < 1$ is associated with a decreased risk of death

p < 0.001 ***

p < 0.05 **

Conclusions

Main Takeaways: Ideal Models

- **Random Forest**

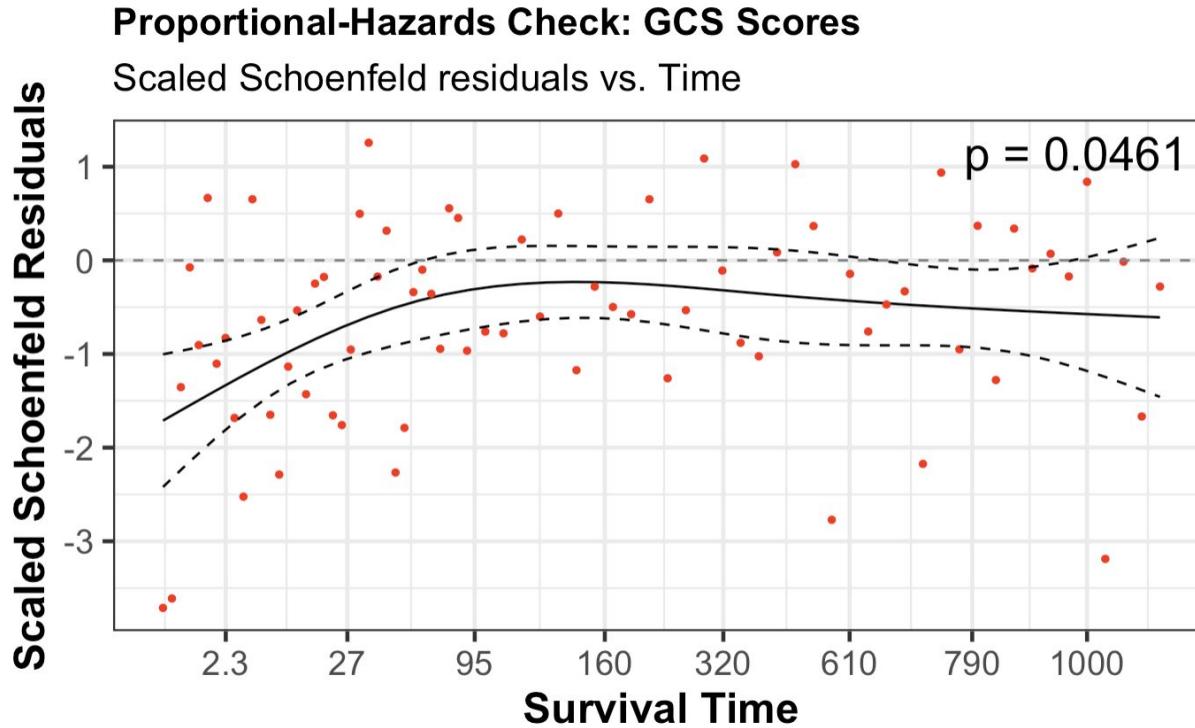
- Clinicians know what factors are influential but not exactly “how” they are associated

- **Cox Proportional Hazards**

- Clinicians know what factors are influential and “how” they are associated, **but** the model requires an assumption of proportional predictor-hazard relationships

	Models	
Features	Random Forest	Cox PH
Flexible Predictor-hazard Relations		
Variable Importance		
Interpretable		 (Limited)

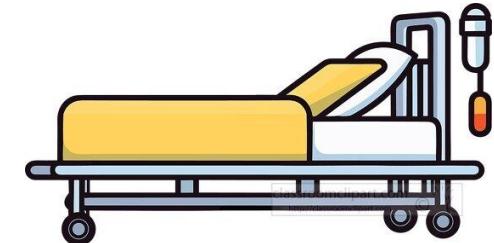
Main Takeaways: Failure of Proportional Hazard Assumption



Main Takeaways: Overall Variable Importance



- Age
- Discharge Location
- Comorbidities
 - Cancer
 - Hepatitis C
- In-Hospital
 - GCS Total
 - Endocrine Procedures
- Medication (Beta Blockers)



Limitations of Our Study

- **Overall Sample Size**
 - Only 756 patients out of 1330 were discharged
- **Censoring**
 - Censoring **only** occurs at 90 days
- **Interpretability of Models**
 - Random Forests are not as interpretable as Cox Regression
- **Single Center Study**
 - Only from Beth Israel Deaconess Medical Center

Going Forward: Future Directions

- Replicate on larger data set
- Incorporate other centers into our study
- Include time-dependent covariates
- Explore other machine learning models
- Causal Inference with Cox Regression



Estimating Causal Treatment Effects for Brain Bleed Patients

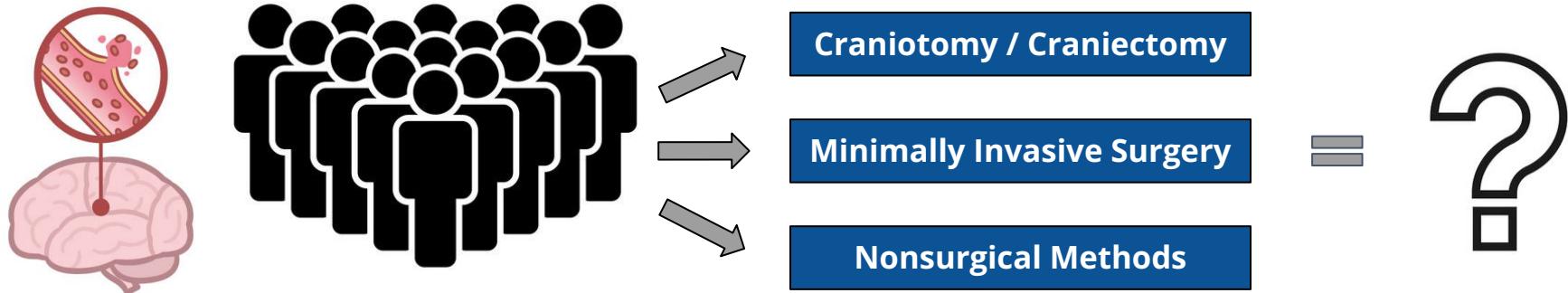
Albert Zhang¹, Hodan Ali², Ryan Kwok³, Sylvia Pesch⁴

Advisors: Dr. Rahul Ladhania⁵, Dr. Katherine Brumberg⁵, Abby Loe⁵

Carnegie Mellon University¹, Smith College², Johns Hopkins University³, University of Minnesota-Morris⁴, University of Michigan⁵



Introduction



Research Topic: Health outcomes of brain bleed patients who received one of three possible treatments during hospitalization

Research Question

Main Question

What is the effect of treatment type on the 90-day survival of brain bleed patients?



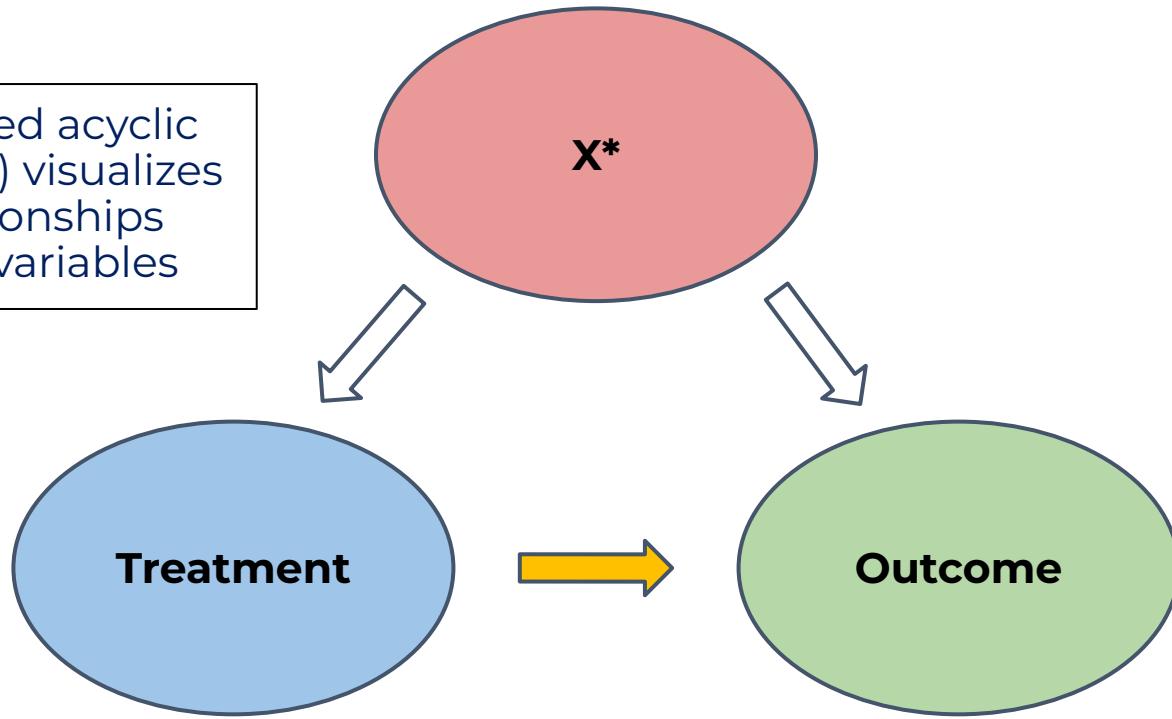
Estimands of Interest

1. Average Treatment Effect (ATE)
→ What is the average effect of treatment type on survival?
2. Conditional Average Treatment Effect (CATE)
→ How does the effect of treatment type on survival vary across patient subgroups?



Causal Inference Background

This directed acyclic graph (DAG) visualizes key relationships between variables

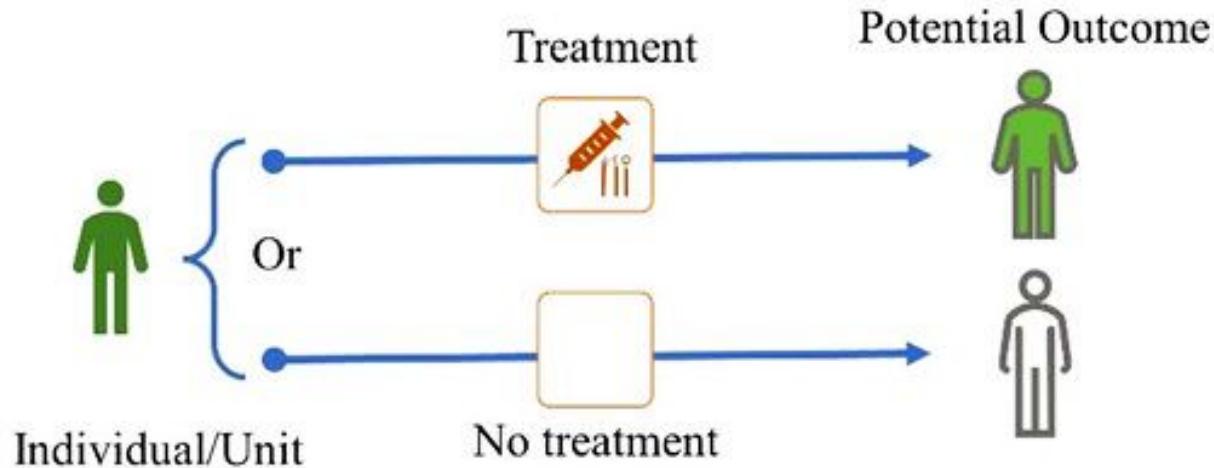


* Covariates

- Patient demographics (age, gender)
- Baseline health measurements (glucose, white blood cell count)

Fundamental Problem of Causal Inference

Explained with the Neyman-Rubin Potential Outcomes Framework:



We cannot observe both the outcome of a patient receiving a treatment and that same patient receiving the control

Causal Inference Assumptions

- **Overlap**
 - For all types of individuals in the population, we can find some portion of individuals in the treatment and some in the control
- **Conditional Unconfoundedness**
 - The treatment assignment is independent of an individual's potential outcomes, conditional on observed covariates
- **Consistency**
 - The observed outcome equals the potential outcome under the treatment the person actually received



Methods and Results

Part I: Average Treatment Effect

Estimand of Interest

Average Treatment Effect (ATE)

$$\tau_a := E[Y_i(a) - Y_i(0)]$$

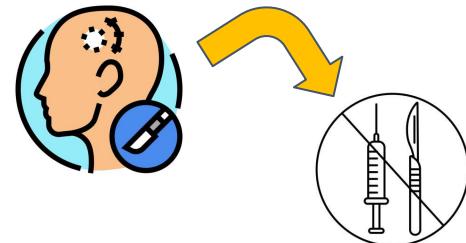
where $Y_i(a)$ is individual i 's outcome under treatment a

where $a := \begin{cases} 1 & \text{Craniotomy/Craniectomy} \\ 2 & \text{Minimally Invasive Surgery} \end{cases}$

Outcome Model

$$\mu(x, a) := E[Y_i \mid X_i = x, A_i = a]$$

Given a person with characteristics x , what is their probability of 90-day mortality if they receive treatment a ? If they received the *control*?



- Predicts counterfactuals $\hat{\mu}(x, a)$
 - What would have happened if a craniotomy patient had received nonsurgical treatment instead?

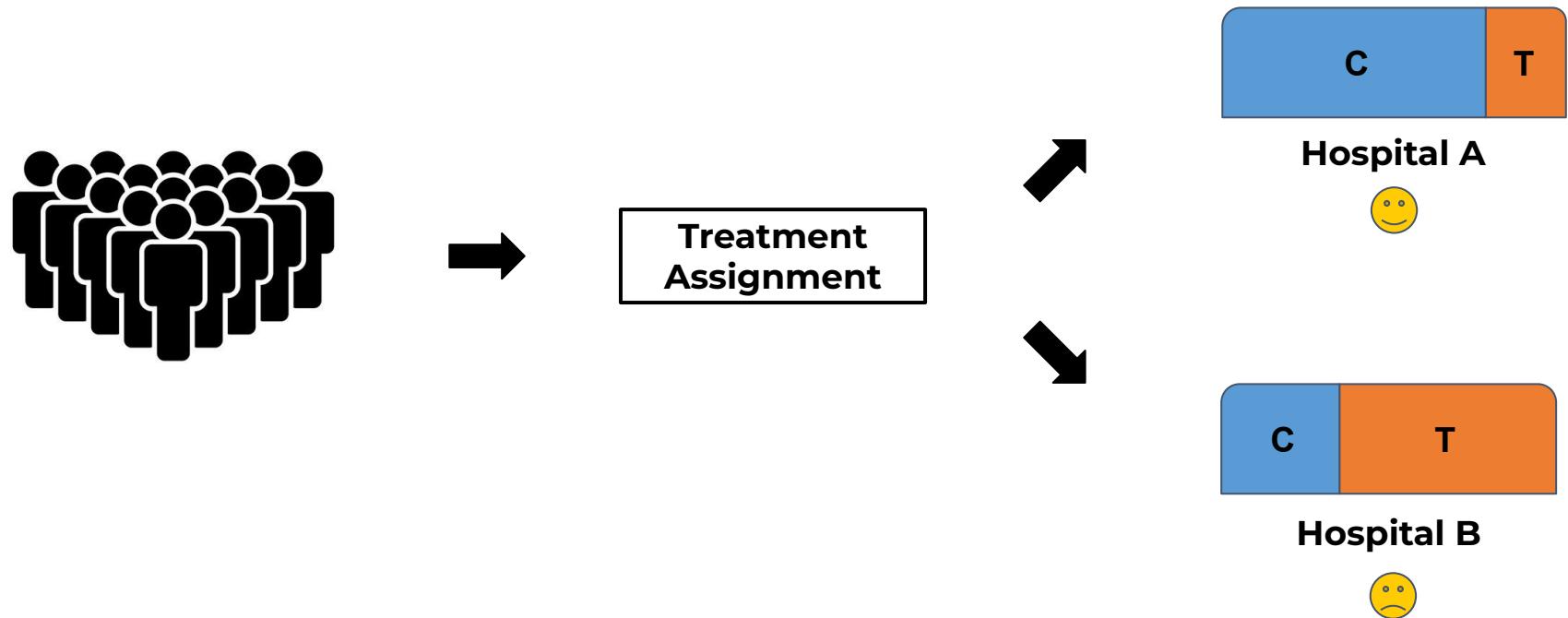
Outcome Model Methods



Outcome Model	Method
Nonsurgical; $\hat{\mu}(x, 0)$	Extra Trees Classifier
Craniotomy/Craniectomy; $\hat{\mu}(x, 1)$	Extra Trees Classifier
Minimally Invasive Surgery; $\hat{\mu}(x, 2)$	Support Vector Machine (SVM)

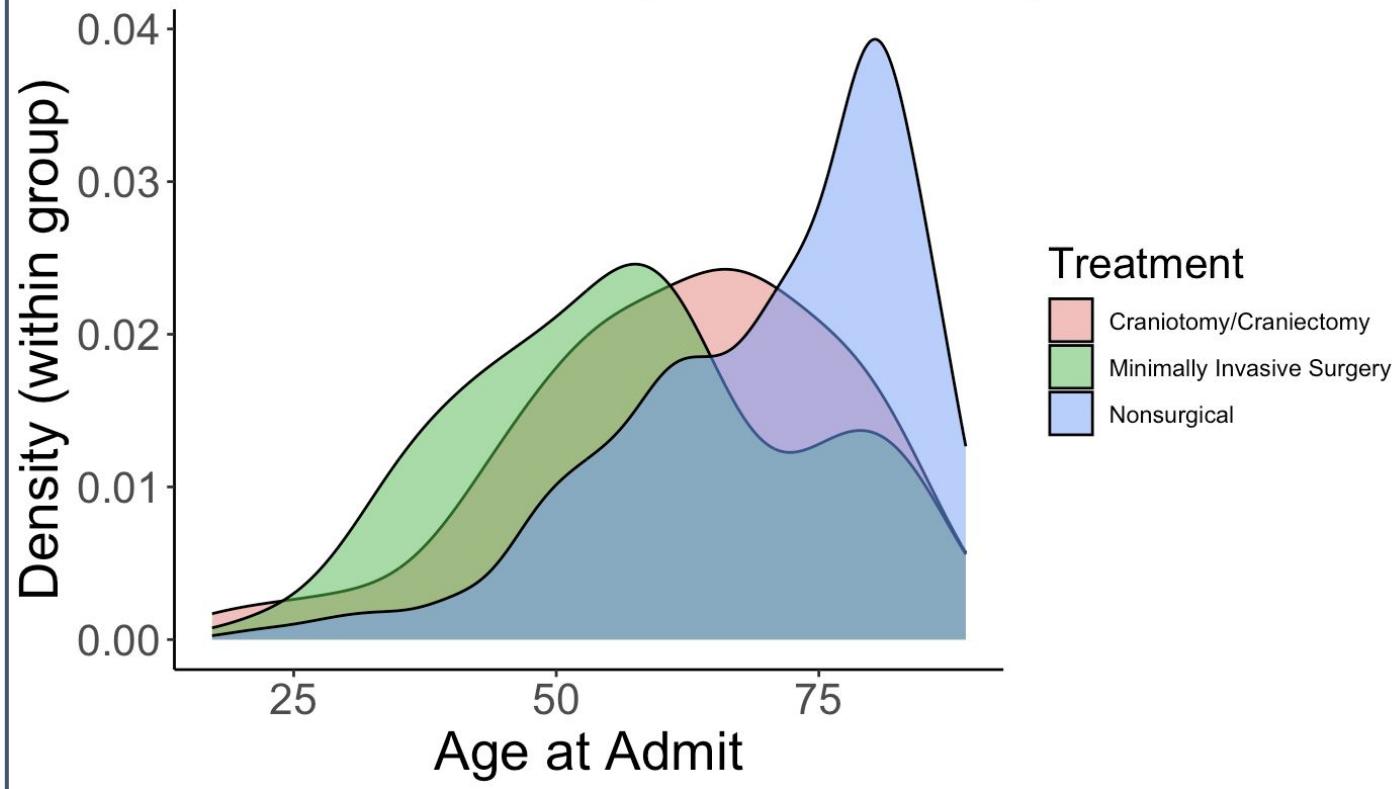


Motivating Propensity



We need to control for unequal treatment probabilities (propensities)

Distribution of Age Stratified by Treatment



M

Treatment propensities differ across patient characteristics

Propensity Score Model

$$e_a(x) := P[A_i = a | X_i = x]$$

Given a person with characteristics x , what is their probability of receiving treatment a ?



Observational setting: propensity scores must be estimated

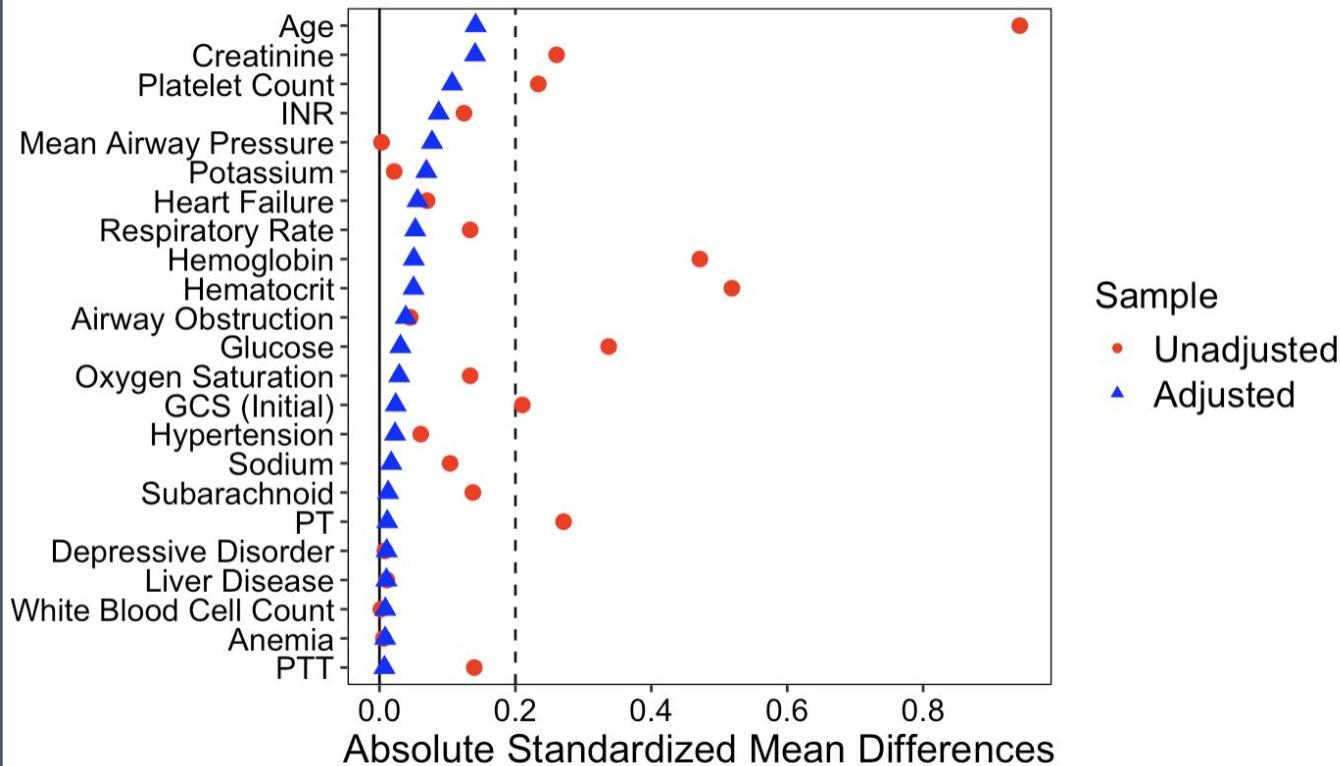
$$\hat{e}_a(x)$$



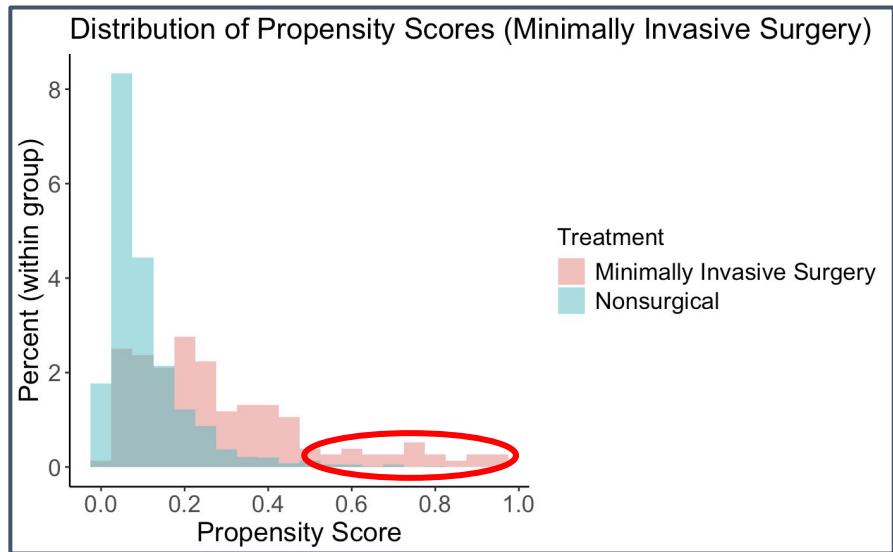
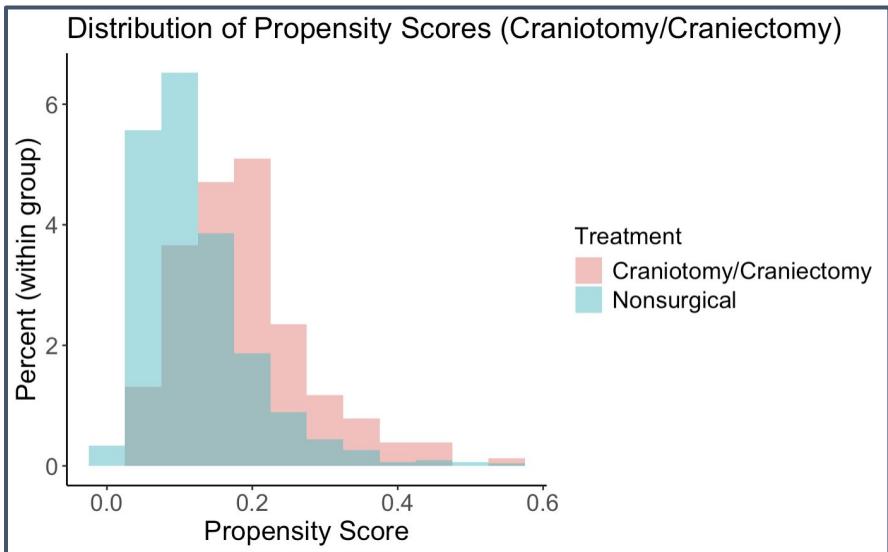
Method of choice: logistic regression model

Assessing Covariate Balance

Covariate Balance (Minimally Invasive Surgery and Nonsurgical)



Propensity Score Distributions



- Overlap of Craniotomy/Craniectomy propensity scores is acceptable
- Upper-end of Minimally Invasive Surgery propensity scores aren't ideal

Augmented Inverse-Propensity Weighted Estimator Workflow

$\hat{\mu}(x, a)$

Choose **outcome models** from 13 candidate models

Choose **propensity score models** from 2 candidate models

$\hat{e}_a(x)$

$\hat{\tau}_a$

“Plug” model estimates into AIPW equation

“Doubly Robust”

Interpret results

Estimator of Interest

Augmented Inverse Propensity-Weighted Estimator (AIPW)

$$\hat{\tau}_a := \frac{1}{n} \sum_{i=1}^n [\hat{\mu}(X_i, a) - \hat{\mu}(X_i, 0) + \frac{(\mathbb{1}_{A_i=a})}{\hat{e}_a(X_i)} (Y_i - \hat{\mu}(X_i, a)) - \frac{(\mathbb{1}_{A_i=0})}{(1-\hat{e}_a(X_i))} (Y_i - \hat{\mu}(X_i, 0))]$$

where $\hat{\mu}(X_i, a) := E[Y_i | X_i = x, A_i = a]$

(outcome model)

where $\hat{e}_a(X_i) = P[A_i = a | X_i]$

(propensity score model)

AIPW Results and Interpretation

τ	Estimate	Standard Error	p-value
<u>Craniotomy/Craniectomy</u>	-0.274	0.0596	0.0003
<u>Minimally Invasive Surgery</u>	-0.132	0.0758	0.4612

**Average difference in mortality across the population is -0.274
for Craniotomy/Craniectomy patients**

(57.7% → 30.3%)



Methods and Results

Part II: Heterogeneous Treatment Effects

Estimand of Interest

Conditional Average Treatment Effect (CATE)

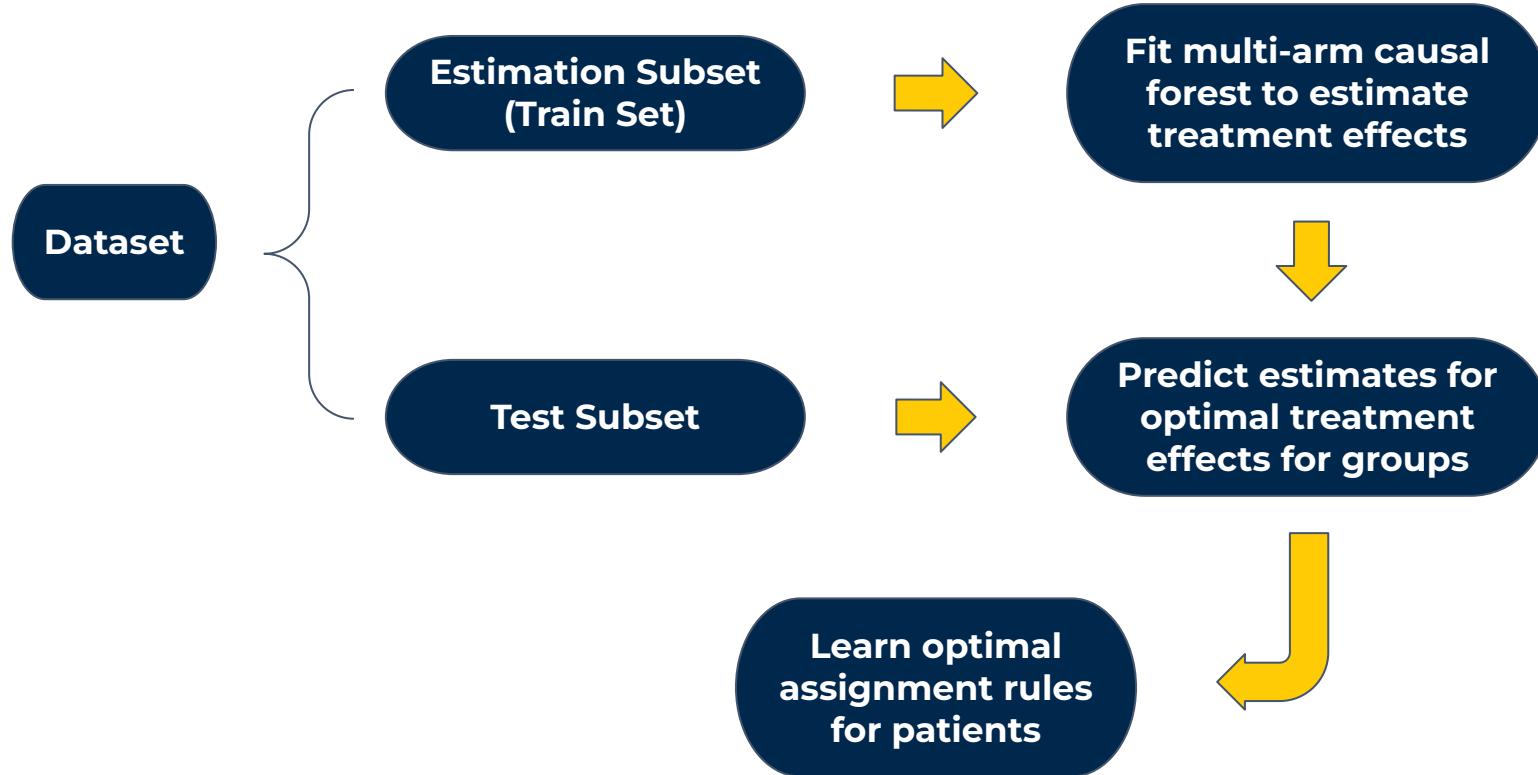
$$\tau_a := \mathbb{E}[Y_i(a) - Y_i(0) | X_i = x]$$

where $Y_i(a)$ is individual i 's outcome under treatment a



What is the treatment effect for a **particular individual** or **group of individuals** with characteristics x ?

HTE Workflow



Patient Characteristics by Model-Optimal Treatment Group

Mean Covariate Values by Model-Optimal Treatment
Rounded, with p-value significance

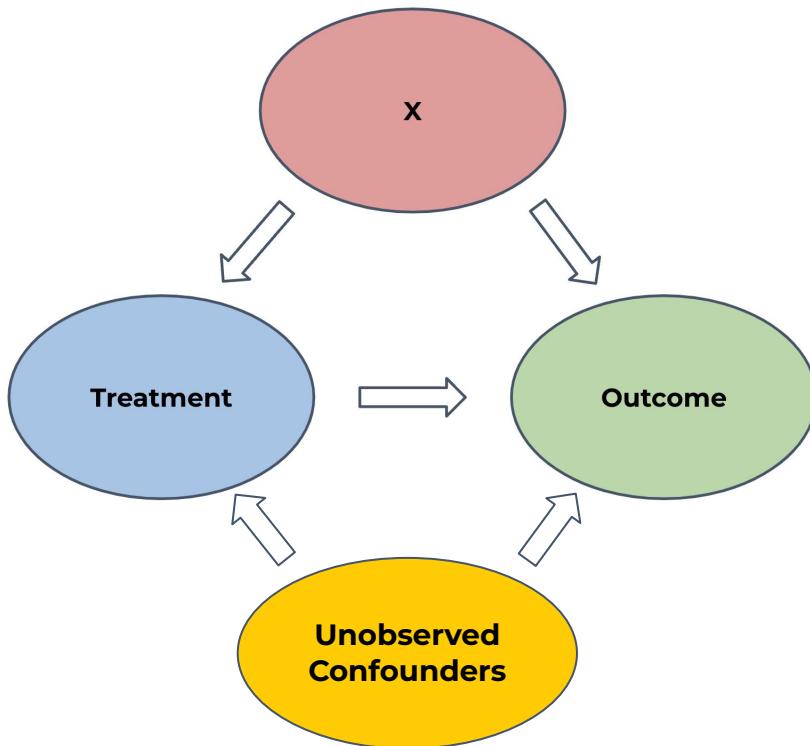
Covariate	Craniotomy/Craniectomy	Minimally Invasive Surgery	p_value
Age at Admission	74	59	0.000
Initial GCS	6	8	0.000
White Blood Cell Count	18	14	0.014
Glucose	168	159	0.019
Platelet Count	234	242	0.126
Hematocrit	30	30	0.428

The model assigns treatments based on meaningful differences in patient characteristics which suggests that it is learning clinically relevant patterns that guide optimal treatment decisions

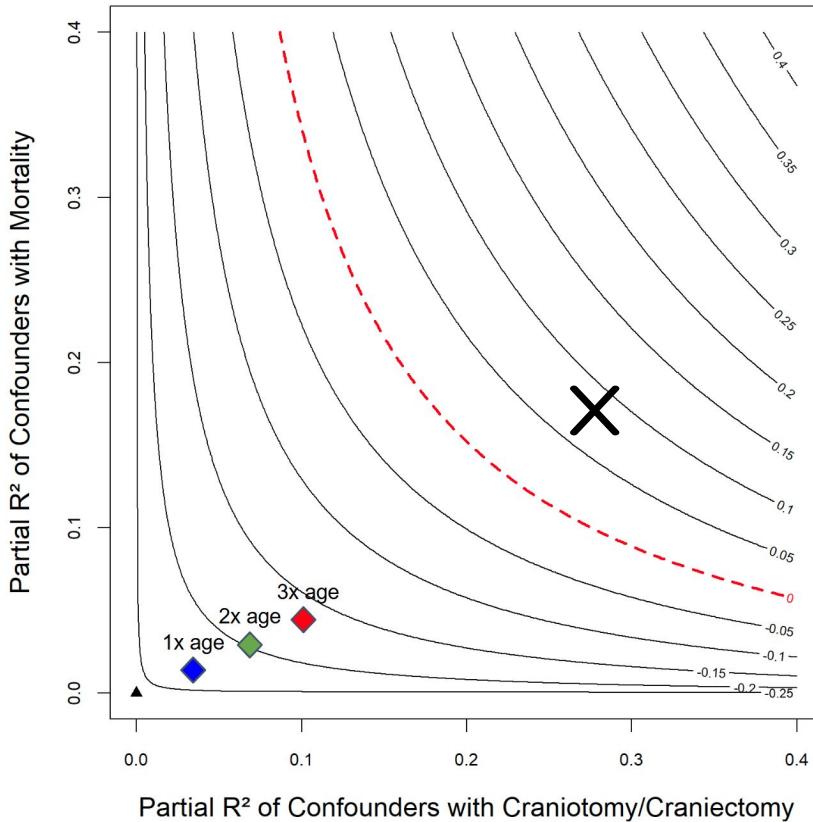


Sensitivity Analysis

Why do we need sensitivity analysis?



Sensitivity Analysis



Blue Diamond → Strength of Age

Green Diamond → 2x Strength of Age

Red Diamond → 3x Strength of Age



Discussion



Conclusions

- **Average Treatment Effect**

- AIPW suggests that receiving a craniotomy/craniectomy reduces mortality compared to nonsurgical care by about 27.4%
- No statistically significant effect for Minimally Invasive Surgery

- **Heterogeneous Treatment Effects**

- Age, GCS, WBC are significant covariates in determining the model-optimal treatment

- **Moving Forward**

- Consider estimating ATT (Average Treatment Effect on the Treated) in addition to ATE

Limitations

- Some violations were made to the overlap assumption at high propensity scores for Minimally Invasive Surgery
- Since the number of patients in the test set was 262, estimates may be noisy
- Imbalance of sample sizes in dataset (~1000 Nonsurgical | ~150 Craniotomy | ~150 Minimally Invasive Surgery) may affect reliability of subgroup estimates

Acknowledgements

We would like to thank our faculty advisors, **Dr. Rahul Ladhania** and **Dr. Katherine Brumberg**, and our graduate student advisor, **Abby Loe**, for their continuous support throughout this project.

We would also like to thank **all BDSI faculty and staff** who made this opportunity possible for all of us!





Thank you!

SOURCES

1. Wang, Z., Wan, M. P., Tai, J. H., & Wang, Y. (2025). Global regional, and national burden of intracerebral hemorrhage between 1990 and 2021. *Scientific Reports*, 15, 3624.
<https://doi.org/10.1038/s41598-025-88017-0>
2. Abulhasan, Y. B., Teitelbaum, J., Al-Ramadhan, K., Morrison, K. T., & Angle, M. R. (2023). Functional Outcomes and Mortality in Patients With Intracerebral Hemorrhage After Intensive Medical and Surgical Support. *Neurology*, 100(19), e1985–e1995.
<https://doi.org/10.1212/WNL.00000000000207132>
3. Boland, P. J. (1997). A reliability comparison of basic systems using hazard rate functions. *Applied stochastic models and data analysis*, 13(3-4), 377-384.
4. Stensrud, M. J., & Hernán, M. A. (2020). Why test for proportional hazards?. *Jama*, 323(14), 1401-1402.

SOURCES

5. Gunay, D. (2023, September 14). Random Forest. Medium.
<https://medium.com/denizgunay/random-forest-af5bde5d7e1e>
6. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the Yield of Medical Tests. *JAMA*. 1982;247(18):2543–2546.
doi:10.1001/jama.1982.03320430047030
7. Park, S. Y., Park, J. E., Kim, H., & Park, S. H. (2021). Review of Statistical Methods for Evaluating the Performance of Survival or Other Time-to-Event Prediction Models (from Conventional to Deep Learning Approaches). *Korean journal of radiology*, 22(10), 1697–1707.
<https://doi.org/10.3348/kjr.2021.0223>