Targeted Learning for Variable Importance in Precision Medicine

Yue You[1], Alan Hubbard[1], Rachael Callcut[2], Lucy Kornblith[2], and Sabrinah Christie[2]

[1]*Division of Biostatistics and Epidemiology, University of California, Berkeley*
[2]*Department of Surgery, Zuckerberg San Francisco General Hospital, University of California, San Fransisco*

September 4, 2018

Abstract

Determining the relative importance of a variable is typically done as a byproduct of fitting algorithms, such as coefficients in large-scale regression (e.g., LASSO), leave-one-variable-out changes in risk (random forest) and other methods that might yield insight, but are tied to specific fitting procedures and even parametric models. We proposed a parameter within a non-parametric model that measures the importance of each variable as the amount of attribution of that variable towards changes in the mean outcome. Specifically, for each of the candidate competing causes of the outcome, we utilized an estimate of this attribution using a statistical approach based upon a combination of machine learning and causal inference via Targeted Learning. This approach allows for 1) variable importance comparisons at the same scale regardless of the original scale of the variable, 2) estimation not dependent on arbitrary parametric assumptions, 3) and asymptotically linear (locally efficient) estimator for which robust asymptotic inference is available. We implemented this approach to determine the variable importance of clinical measures in trauma patients in predicting the probability of mortality at different time periods (from time of injury) using data from three independent trauma studies. This approach allowed comparisons of variable importance within and between trauma cohorts and identified variables with the biggest potential "intervention" impact for mortality. Our results showed that the most important variables across all time intervals is initial International Normalized Ratio, while importance of other variables varied by time. These findings were similar across three trauma cohorts. This method can serve as an alternative to more standard variable importance procedures that lack both broad clinical interpretability and mechanisms for accurate statistical inference in the context of data-adaptive estimation.

1 Introduction

Nowadays, researchers have access to overwhelmingly complex high-dimensional electronic health record (EHR) data. And it became our primary goal to establish a measure of variable importance at each given time point in terms of impacting certain health outcomes, so that we are able to assist clinicians to make optimal care decisions using the entirety of the patient's data. However, there is not a universal definition of "most important" variables. One type of definition is motivated by prediction power, where the "most important" variables are those which accurately predict the outcome (like those generated from Random Forest), and another type by causal association, where the "most important" variables are those whose change cause most changes in outcomes (like coefficients in Lasso regression). Here, we focus on the latter one and refer it as Variable Importance Measure (VIM) in the following parts of article.

There are plenty of medical literature measuring variable importance and building associations between these variables and outcomes of interest. Yet most of them only applies parametric models or evaluate the importance of parameters as a fixed value over the whole intervention [1][2][3][4][5]. In comparison, our method makes no parametric modeling assumptions and relies on automated machine learning to estimate the data-generating distribution.

In this paper we applied a novel variable importance measure based upon a combination of machine learning and causal inference via Targeted Learning[6], to determine the variable importance of trauma study with regard to patients' mortality rates at different time periods. Specifically, We applied the NPVI estimator [7] to three independent trauma studies, which include two observational studies (ACIT[8][9][10], PROMMTT[11]) and one randomized trial (PROPPR[12]). We used the methods to examine whether to estimate importance of seven variables measured in emergency department, adjusting for twelve demographic and trauma-specific variables, regarding patients' mortality rates within 2 hours, in 2 to 6 hours, and in 6 to 24 hours respectively. Our results showed that we could use this novel method of variable importance to rank the variables that contribute to poor outcomes among trauma patients. Consistently across the three studies, the most important variable across all time intervals is initial International Normalized Ratio (*inr*), and the importance of other blood-clotting related variables (such as initial PLaTelet count, or *plt*, and initial Partial Thromboplastin Time, or *ptt*) vary by time. Thus, the paper demonstrates the centrality of coagulopathy and its resolution on patient survival in acute trauma.

In section 2, we illustrated the methodology, including source of data, parameter of interest, and implementation details. In section 3, We presented the variable lists, summary statistics, and results. In section 4, we discussed our findings and how they would assist clinicians in emergency department.

2 Methodology

We compared the results across three prospective studies of acute trauma, among which two are multi-center studies and one is single-center. We specified a common set of predictors across the three studies and examined the relative VIM as defined in section 2.2. The predictor variables include both demographic factors, measures of severity of injury, and lab values related to coagulation. As explained below, some of the variables we considered representing potential interventions of interest, others as adjustment variables for all variable importance measures constructed.

2.1 Data

We describe the three independent data sources below, the first two of which are prospective observational studies, the last a clinical trial, where ratio of blood products was the randomized intervention.

2.1.1 PROMMTT

The PRospective, Observational Multi-center Major Trauma Transfusion (PROMMTT) study enrolled 1,245 individuals at ten level I trauma centers from around the United States. Patients had to survive at least 30 minutes and receive at least one unit of red blood cells within 6 hours of arrival in the emergency department[11]. Once enrolled they were followed for diagnostic and therapeutic procedures and subsequent outcome data were collected . The PROMMTT study was approved at each study site and the Data Coordinating Center by the local institutional review boards. The US Army Human Research Protections Office also provided secondary level review and approval. Patient records were deidentified prior to their use in this analysis.

2.1.2 ACIT

The Activation of Coagulation and Inflammation in Trauma (ACIT) study is a prospective cohort study of 1,671 severe trauma patients admitted to a single level I trauma center.[8][9][10] Several physiological and clinical measurements were recorded at several time points for each patient after arrival to the emergency department. These variables include demographic variables, baseline risk factors, longitudinally measured variables that account for the patients' exposure and health status history, and an indicator of the occurrence of death at each time interval. Because these data are often collected in a highly dynamic environment, it is common that some variables are missing for some patients at a given time point[13].

2.1.3 PROPPR

The Pragmatic, Randomized Optimal Platelet and Plasma Ratios (PROPPR) trial was designed to address the effectiveness and safety of a 1:1:1 transfusion ratio compared with a

1:1:2 transfusion ratio in patients with trauma for whom massive transfusion were indicated[12]. Specifically, PROPPR was a pragmatic, phase 3, multi-site, randomized clinical trial of 680 severely injured patients who arrived at 1 of 12 level I trauma centers in North America, between August 2012 and December 2013. This study measured patient demographics, injury mechanism, clinical scoring measurements related to trauma injuries, a baseline test of whole blood coagulation, and data were collected multiple time points.

2.2 Parameter of Interest

Suppose the observed data structure is $O = (W, X, Y)$, where confounder $W \in W$ is a vector of baseline covariates, exposure $X \in R$ and outcome $Y \in R$ respectively quantify an exposure and a response, and we wish to investigate the relationship between X on Y, accounting for W. Taking W into account is desirable because we know (or cannot rule out the possibility) that it contains confounding factors, i.e., common factors upon which the exposure X and the response Y may simultaneously depend. One classical approach to study such relationship is the Average exposure Effect (ATE) estimator [14], which investigates the causal effect of binary exposure X on outcome Y. Specifically, the estimand for the ATE for binary X is:

$$E_0\{E_0(Y \mid X = 1, W) - E_0(Y \mid X = 0, W)\}$$

However, if one can define a target level that is a priori to be the goal value of medical interventions, then an interesting and very relevant measure of the variable can be estimated. It requires that the target level, x_0, has positive mass ($P(X = x_0) > 0$, that is, there will a growing number of people in the data with the same value x_0 as sample size grows) and a continuum of other levels. Thus, we propose to estimate an alternative parameter, one that measures the causal effect of pseudo-continuous exposures on health outcomes, by comparing the current mean of the outcome to what it would have been if the variables was set at its target level. This allows one to compare the estimates across differently explanatory variables, as now, regardless of the scale and distribution of the variable, the estimate has the same interpretation, and hence evaluate the relative importance of several variables in regard to post-trauma outcomes.

Thus, we could use an estimator. Such estimator is named Targeted Minimum Loss Estimation Methodology Non-Parametric Variable Importance (tmle.npvi) estimator [7] and introduced briefly here:
For all distributions P of O compatible with the above description of O, for f, a user-supplied function such that $f(0) = 0$,

$$\Psi_f(P) = \frac{E_P\{f(X - x_0)[E_P(Y|X,W)) - E_P(Y|X = x_0, W)]\}}{E_P\{f(X - x_0)^2\}} \tag{1}$$

In this paper, we set $f(X) = |X|$ for $X \in R$, and set x_0 as the low-risk range, defined in 3. And our targeted parameter becomes:

$$\Psi_{|X|}(P) = \frac{E_P\{|X - x_0|[E_P(Y|X,W)) - E_P(Y|X = x_0, W)]\}}{E_P\{(X - x_0)^2\}} \tag{2}$$

Developed based on the semi-parametric estimation methodology called TMLE [6]. the tmle.npvi estimator has several desirable statistical properties (convergence of the iterative procedure in TMLE methodology; consistency and asymptotic normality of the estimator) [7].

2.3 implementation

Each time, we take each of the 7 exposures (A), all of 12 confounders (W), and one of the 3 outcomes (Y), to fit model 2, and obtain one estimated $\hat{\Psi}_{|X|}(P)$. So we have $7\,A \times 3\,Y = 21\,\hat{\Psi}_{|X|}(P)$ results in total.

We implemented our estimation using tmle.npvi R-package [7], where we fitted the SuperLearner[15] using the algorithms implemented by the following R-packages:

1. Generalized additive models by gam R-package[16], with its default values.
2. Generalized linear models: glm R-function.
3. Piecewise linear splines by polymars R-function from the polspline R-package[17] with its default values.
4. Random forests by randomForest R-package[18], with its default values.
5. Support vector machines by svm R-function from the e1071 R-package[19], with its default values.

3 Results

3.1 Variable Tables

Here we listed all the confounders used in modeling each of exposure variables as below:
Confounder W:

	Variable Name	Description
1	bmi	Body mass index
2	latino	Indicator for Hispanic ethnicity
3	age	Age in year
4	penetrating	Injury type (blunt vs. penetrating)
5	sex	Gender
6	race	Race
7	anticoag	Indicator for anticoagulant use
8	Ibmi	Indicator for missing body mass
9	Ilatino	Indicator for missing Hispanic ethnicity
10	Iage	Indicator for missing age
11	Ianticoag	Indicator for missing anticoagulant use
12	Irace	Indicator for missing race

Table 1: *Variable names of confounder W*

Here we listed all the exposures as below:
Exposure A:

	Variable	Description
1	hgb	Initial ED hemoglobin results
2	hr	Initial ED heart rate
3	inr	Initial ED international normalized ratio
4	iss	Injury severe score
5	plt	Initial ED platelet count
6	ptt	Initial ED partial thromboplastin time results
7	sbp	Initial ED systolic blood pressure

* ED = Emergency Department

Table 2: *Variable names of exposure A*

And we defined the ”low-risk range” for each of our exposures as below. The cutoff is provided by clinicians working in the emergency department.
Low-risk ranges of exposure::

	Variable name	Low-risk range
1	hgb	[12, 17]
2	hr	[60, 80]
3	inr	[0, 1.4]
4	iss	[0, 10]
5	plt	[150, 450]
6	ptt	[0, 35]
7	sbp	[120, 159]

Table 3: *Low-risk Ranges of exposure*

Here we listed all three outcomes as below:
Outcome Y:

	Variable	Description
1	Mortality 2h	Indicator of death in 2 hours
2	Mortality 2to6h	Indicator of death in 2 to 6 hours
3	Mortality 6to24h	Indicator of death in 6 to 24 hours

Table 4: *Variable names of outcome Y*

3.2 Descriptive Statistics

Variable	Overall		
	ACIT	PROPPR	PROMMTT
	Overall	Overall	Overall
Body mass index (kg/m^2), M (SD)	26.77 (5.21)	28.20 (14.20)	27.94 (6.87)
Ethnicity, n (%)			
Hispanic	432 (27.12)	120 (17.65)	933 (79.61)
Age, M (SD)	40.96 (18.60)	38.69 (17.45)	40.78 (18.67)
Penetrating Injury, n (%)			
Penetrating	706 (42.38)	330 (48.53)	438 (35.27)
Sex, n (%)			
female	310 (18.55)	134 (19.71)	320 (25.76)
Race, n (%)			
White	949 (59.65)	434 (65.46)	941 (75.76)
Black	375 (23.57)	186 (28.05)	223 (17.95)
Asian/Pacific Islander	223 (14.02)	33 (4.98)	47 (3.78)
Other	44 (2.77)	10 (1.51)	31 (2.50)
Anticoagulant use, n (%)			
Yes	95 (8.22)	24 (4.59)	145 (15.15)
Deficit	894 (72.21)	439 (86.59)	884 (71.18)
Initial ED hemoglobin results, M (SD)	13.63 (1.99)	11.97 (6.54)	11.21 (3.20)
Initial ED heart rate, M (SD)	97.60 (24.93)	109.12 (25.18)	103.64 (31.73)
Initial ED international normalized ratio, M (SD)	1.20 (0.46)	1.42 (0.44)	1.26 (1.19)
Injury severe score, M (SD)	17.70 (15.63)	28.94 (15.04)	26.13 (15.23)
Initial ED platelet count, M (SD)	228.55 (93.04)	218.56 (82.62)	218.83 (96.78)
Initial ED partial thromboplastin time results, M (SD)	30.40 (12.79)	35.71 (22.75)	26.76 (19.96)
Initial ED systolic blood pressure, M (SD)	135.87 (33.26)	107.94 (32.50)	105.34 (35.51)

Table 5: *Summary Statistics on overall datasets*

Variable	0-2h					
	ACIT		PROPPR		PROMMTT	
	$> 2h$	$\leq 2h$	$> 2h$	$\leq 2h$	$> 2h$	$\leq 2h$
Body mass index (kg/m^2), M (SD)	26.77 (5.22)	26.94 (4.63)	28.18 (14.39)	28.55 (8.19)	27.91 (6.77)	29.25 (11.03)
Hispanic ethnicity, n (%)	427 (27.34)	5 (16.13)	114 (17.87)	6 (14.29)	893 (79.24)	40 (88.89)
Age, M (SD)	40.92 (18.49)	42.84 (22.88)	38.83 (17.50)	36.55 (16.64)	40.82 (18.68)	39.64 (18.64)
Penetrating injury, n (%)	684 (42.01)	22 (57.89)	308 (48.28)	22 (52.38)	414 (34.67)*	24 (50.00)*
Female, n (%)	303 (18.55)	7 (18.42)	124 (19.44)	10 (23.81)	309 (25.88)	11 (22.92)
Race, n (%)						
White	938 (60.17)*	11 (34.38)*	408 (65.49)	26 (65.00)	911 (76.30)	30 (62.50)
Black	364 (23.35)*	11 (34.38)*	172 (27.61)	14 (35.00)	209 (17.50)	14 (29.17)
Asian/Pacific Islander	215 (13.79)*	8 (25.00)*	33 (5.30)	0 (0.00)	45 (3.77)	2 (4.17)
Other	42 (2.69)*	2 (6.25)*	10 (1.61)	0 (0.00)	29 (2.43)	2 (4.17)
Anticoagulant use, n (%)	93 (8.10)	2 (25.00)	24 (4.74)	0 (0.00)	144 (15.35)	1 (5.26)
Initial ED hemoglobin results, M (SD)	13.68 (1.95)***	11.55 (2.49)***	12.02 (6.66)	10.89 (2.70)	11.34 (3.09)***	8.03 (4.28)***
Initial ED heart rate, M (SD)	97.30 (24.61)*	112.84 (35.23)*	109.22 (24.70)	107.06 (33.92)	103.55 (31.05)	105.79 (45.96)
Initial ED international normalized ratio, M (SD)	1.19 (0.44)**	1.89 (0.98)**	1.41 (0.44)**	1.80 (0.45)**	1.25 (1.16)	1.60 (1.72)
Injury severe score, M (SD)	17.28 (15.30)	35.82 (18.62)	28.72 (14.97)	32.31 (15.80)	25.70 (14.86)***	36.90 (19.97)***
Initial ED platelet count, M (SD)	229.82 (92.96)	167.88 (76.37)	220.90 (82.15)***	163.96 (75.60)***	222.35 (95.09)***	131.33 (98.33)***
Initial ED partial thromboplastin time results, M (SD)	29.79 (9.93)**	67.56 (53.14)**	35.28 (22.54)	50.39 (26.02)	26.37 (18.81)	36.36 (37.94)
Initial ED systolic blood pressure, M (SD)	136.17 (32.77)	120.94 (50.53)	108.83 (32.14)**	85.96 (34.12)**	106.09 (34.78)**	86.75 (47.22)**

* p-value ≤ 0.05, ** p-value ≤ 0.01, *** p-value ≤ 0.001

Table 6: *Summary statistics in 0 to 2 hours*

Variable	2-6h					
	ACIT		PROPPR		PROMMTT	
	$> 6h$	$(2h, 6h]$	$> 6h$	$(2h, 6h]$	$> 6h$	$(2h, 6h]$
Body mass index (kg/m^2), M (SD)	26.79 (5.21)	25.14 (5.40)	28.19 (14.64)	27.98 (6.26)	27.85 (6.80)**	30.93 (4.32)**
Hispanic ethnicity, n (%)	421 (27.57)	6 (17.14)	108 (17.94)	6 (16.67)	854 (79.00)	39 (84.78)
Age, M (SD)	40.69 (18.24)*	50.42 (25.49)*	38.57 (17.32)	43.28 (20.11)	40.69 (18.51)	43.72 (21.95)
Penetrating injury, n (%)	670 (42.19)	14 (35.00)	295 (49.00)	13 (36.11)	393 (34.47)	21 (38.89)
Female, n (%)	291 (18.27)	12 (30.00)	119 (19.77)	5 (13.89)	297 (26.05)	12 (22.22)
Race, n (%)						
White	917 (60.17)	21 (60.00)	383 (65.14)	25 (71.43)	873 (76.58)	38 (70.37)
Black	357 (23.43)	7 (20.00)	163 (27.72)	9 (25.71)	198 (17.37)	11 (20.37)
Asian/Pacific Islander	208 (13.65)	7 (20.00)	33 (5.61)	0 (0.00)	43 (3.77)	2 (3.70)
Other	42 (2.76)	0 (0.00)	9 (1.53)	1 (2.86)	26 (2.28)	3 (5.56)
Anticoagulant use, n (%)	91 (8.01)	2 (16.67)	23 (4.70)	1 (5.88)	142 (15.52)	2 (8.70)
Initial ED hemoglobin results, M (SD)	13.71 (1.92)**	12.36 (2.57)**	12.10 (6.82)**	10.72 (2.36)**	11.40 (3.03)**	9.98 (3.80)**
Initial ED heart rate, M (SD)	97.27 (24.58)	98.55 (25.85)	109.08 (24.47)	111.53 (28.47)	103.84 (29.86)	97.43 (49.90)
Initial ED international normalized ratio, M (SD)	1.18 (0.43)***	1.66 (0.65)***	1.41 (0.45)	1.50 (0.31)	1.24 (1.17)	1.40 (1.09)
Injury severe score, M (SD)	16.78 (15.02)***	37.64 (12.87)***	27.96 (14.47)***	41.36 (17.69)***	25.33 (14.70)***	33.35 (16.28)***
Initial ED platelet count, M (SD)	231.59 (92.32)***	163.65 (93.64)***	222.35 (81.71)	197.11 (86.91)	225.05 (94.54)***	165.17 (89.15)***
Initial ED partial thromboplastin time results, M (SD)	29.25 (8.00)***	51.67 (31.93)***	34.52 (21.89)	46.58 (28.91)	25.82 (17.85)**	38.04 (31.21)**
Initial ED systolic blood pressure, M (SD)	136.45 (32.48)	125.33 (41.63)	110.19 (31.87)***	86.14 (28.27)***	106.75 (33.53)	92.15 (53.46)

* p-value ≤ 0.05, ** p-value ≤ 0.01, *** p-value ≤ 0.001

Table 7: *Summary statistics in 2 to 6 hours*

Variable	6-24h					
	ACIT		PROPPR		PROMMTT	
	$> 24h$	$(6h, 24h]$	$> 24h$	$(6h, 24h]$	$> 24h$	$(6h, 24h]$
Body mass index (kg/m^2), M (SD)	26.83 (5.23)	25.95 (4.68)	28.22 (14.84)	27.20 (5.83)	27.95 (6.81)*	25.35 (5.95)*
Hispanic ethnicity, n (%)	409 (27.82)	12 (21.05)	105 (18.13)	3 (13.04)	819 (78.75)	35 (85.37)
Age, M (SD)	40.29 (17.96)***	50.65 (22.11)***	38.16 (16.89)*	48.70 (24.13)*	40.60 (18.16)	42.67 (25.71)
Penetrating Injury, n (%)						
Penetrating injury, n (%)	654 (42.86)*	16 (25.81)*	288 (49.74)	7 (30.43)	382 (34.92)	11 (23.91)
Female, n (%)	270 (17.64)**	21 (33.87)**	111 (19.17)	8 (34.78)	282 (25.78)	15 (32.61)
Race, n (%)						
White	893 (60.87)***	24 (42.11)***	365 (64.60)	18 (78.26)	835 (76.33)	38 (82.61)
Black	346 (23.59)***	11 (19.30)***	159 (28.14)	4 (17.39)	193 (17.64)	5 (10.87)
Asian/Pacific Islander	187 (12.75)***	21 (36.84)***	32 (5.66)	1 (4.35)	42 (3.84)	1 (2.17)
Other	41 (2.79)***	1 (1.75)***	9 (1.59)	0 (0.00)	24 (2.19)	2 (4.35)
Anticoagulant use, n (%)	91 (8.01)*	2 (16.67)*	23 (4.70)	1 (5.88)	142 (15.52)	2 (8.70)
Initial ED hemoglobin results, M (SD)	13.75 (1.91)**	12.82 (2.12)**	12.13 (6.90)	11.30 (4.39)	11.43 (3.00)	10.70 (3.68)
Initial ED heart rate, M (SD)	97.04 (24.18)	102.92 (32.79)	109.07 (24.38)	109.33 (27.58)	103.95 (29.73)	101.33 (33.21)
Initial ED international normalized ratio, M (SD)	1.16 (0.35)*	1.57 (1.27)*	1.38 (0.41)*	1.89 (0.79)*	1.24 (1.18)	1.38 (0.82)
Injury severe score, M (SD)	16.15 (14.76)***	32.18 (13.17)***	27.54 (14.26)**	38.48 (15.91)**	24.93 (14.56)***	35.02 (14.71)***
Initial ED platelet count, M (SD)	233.98 (90.58)***	172.48 (114.04)***	224.24 (81.57)**	177.09 (72.98)**	227.01 (94.56)***	178.57 (82.34)***
Initial ED partial thromboplastin time results, M (SD)	28.83 (7.38)***	39.96 (13.95)***	32.74 (17.03)*	74.80 (57.20)*	25.40 (17.41)**	35.76 (24.34)**
Initial ED systolic blood pressure, M (SD)	135.65 (31.46)**	155.97 (47.87)**	110.44 (31.72)	103.40 (36.00)	107.11 (32.90)	98.26 (45.70)

* p-value ≤ 0.05, ** p-value ≤ 0.01, *** p-value ≤ 0.001

Table 8: *Summary statistics in 6 to 24 hours*

3.3 Results

The results for ACIT cohort is

variable	Ψ_{2h}	95% CI_{2h}	Ψ_{2-6h}	95%CI_{2-6h}	Ψ_{6-24h}	95% CI_{6-24h}
hgb	0.0117*	(0.0008, 0.0225)	0.0081	(-0.0013, 0.0175)	0.0005	(-0.0092, 0.0101)
hr	0.0004**	(0.0002, 0.0006)	0.0001	(-0.0002,0.0003)	0.0001	(-0.0002, 0.0005)
inr	0.2791***	(0.1859, 0.3722)	0.3391***	(0.2253, 0.4529)	0.3456***	(0.2424, 0.4489)
iss	0.0014***	(0.0007, 0.0022)	0.0016***	(0.0011, 0.0021)	0.0019***	(0.0011, 0.0028)
plt	0.0000	(-0.0001, 0.0001)	0.0000	(-0.0001, 0.0002)	0.0001	(-0.0002, 0.0003)
ptt	0.0187***	(0.0143, 0.0231)	0.0174***	(0.0128, 0.022)	0.0015**	(0.0004, 0.0025)
sbp	0.0001	(-0.0001, 0.0003)	0.0000	(-0.0003, 0.0002)	0.0005**	(0.0001, 0.0009)

* p-value $\leq$ 0.05, ** p-value $\leq$ 0.01, *** p-value $\leq$ 0.001

Table 9: *ACIT*

The results for PROPPR cohort is

variable	Ψ_{2h}	95%CI_{2h}	Ψ_{2-6h}	95%CI_{2-6h}	Ψ_{6-24h}	95%CI_{6-24h}
hgb	NA	NA	0.0011	(-0.0007, 0.0029)	-0.0004	(-0.0041, 0.0033)
hr	0.0000	(-0.0006, 0.0006)	0.0003	(-0.0001, 0.0008)	0.0001	(-0.0005, 0.0008)
inr	0.0444**	(0.01800, 0.0709)	0.0486*	(0.0159, 0.0814)	0.1316***	(0.0856, 0.1776)
iss	-0.0001	(-0.0011, 0.0010)	0.0017***	(0.001, 0.0024)	0.0010	(-0.0001, 0.0021)
plt	0.0005	(-0.0001, 0.0011)	0.0002	(-0.0003, 0.0006)	0.0006	(-0.0001, 0.0013)
ptt	0.0012*	(0.0004, 0.0020)	0.0032***	(0.0017, 0.0047)	0.0101***	(0.0073, 0.0130)
sbp	0.001*	(0.0003, 0.0017)	0.0012*	(0.0005, 0.0020)	0.001*	(0.0000, 0.0020)

* p-value $\leq$ 0.05, ** p-value $\leq$ 0.01, *** p-value $\leq$ 0.001

Table 10: *PROPPR*

The results for PROMMTT cohort is

variable	Ψ_{2h}	95%CI_{2h}	Ψ_{2-6h}	95%CI_{2-6h}	Ψ_{6-24h}	95%CI_{6-24h}
hgb	0.0117*	(0.0011, 0.0224)	0.0073	(-0.0018, 0.0164)	0.0018	(-0.0088, 0.0125)
hr	0.0005***	(0.0002, 0.0007)	0.0000	(-0.0004, 0.0003)	0.0003	(0.0000, 0.0007)
inr	0.2838***	(0.1894, 0.3783)	NA	NA	0.2451***	(0.1737, 0.3164)
iss	0.0016***	(0.0009, 0.0024)	0.0015***	(0.0010, 0.0020)	0.0018***	(0.0008, 0.0028)
plt	0.0000	(-0.0001, 0.0001)	0.0000	_-0.0001, 0.0002_	0.0002	(-0.0001, 0.0004)
ptt	NA	NA	0.0171***	(0.0126, 0.0216)	0.0143***	(0.0102, 0.0183)
sbp	0.0002	(-0.0001, 0.0005)	0.0001	(-0.0002, 0.0005)	0.0005	(-0.0001, 0.0011)

* p-value $\leq$ 0.05, ** p-value $\leq$ 0.01, *** p-value $\leq$ 0.001

Table 11: *PROMMTT*

The variable that most clearly affect survival outcome across all time intervals is *inr*, or International Normalized Ratio. *inr* is derived in part from the *ptt*, or Prothrombin time.

The *ptt*, or Initial ED partial thromboplastin time results, is another factor in clotting. We can see from the variable importance measure that these variables all have significant positive changes for mortality across all three time intervals and across three different studies.

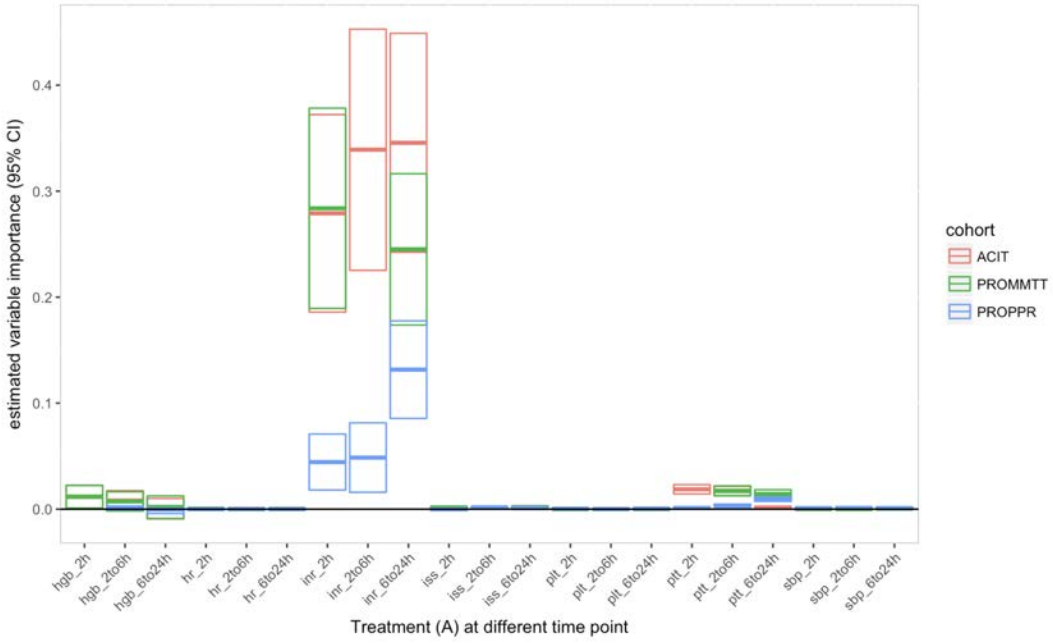

Figure 1: *Variable importance across cohorts*

The statistical interpretation for *inr* in PROPPR is:

In PROPPR cohort, if *inr*, or initial International Normalized Ratio at Emergency Room, is out of the "low-risk range" of [0, 1.4], the mortality rate within 2 hours will increase by 4.44%(1.80%, 7.09%), the mortality rate from 2 to 6 hours will increase by 4.86%(1.59%, 8.14%), and the mortality rate from 2 to 6 hours will increase by 13.16%(8.56%, 17.76%) . We can interpret other treatments using the similar framework.

4 Discussion

When we looked at these results we noticed a few patterns.

- *inr* is the overwhelmingly significant variable in predicting mortality rate across all 3 time periods and all 3 cohorts.
- The two variables *inr*and *ptt* that have the highest degree of association with poor outcomes are signs of abnormality in the blood clotting pathway. This is a noticeable effect at all three time ranges and across all three different studies.
- The variable *hgb* is significant contributors to mortality prior to 6 hours after admission. These are both indicators of oxygen transport capacity.
- The variables *hr* and *sbp* are both indicators for blood volume. The fact that these variables have less impact on mortality suggests that there is little effect of blood volume on mortality after admission. This makes sense since the patient can be given a transfusion which will stabilize fluid volumes assuming no other abnormalities. Abnormalities in oxygen transport can be overcome with proper treatment such as transfusion to replace lost hemoglobin and exogenous oxygen. This would help explain why these variables are less significant 6 hours after admission. In contrast, the factors affecting blood clotting seem to have an effect on mortality in the short, medium and long term. It is interesting that there is such similarity in the correlations of the different variables we have identified for each clinical consideration.

Our study meets the practical need of clinicians to find the important variables related to mortality from trauma in emergency department. For future study, we plan to use data-adaptive methods to determine the practical low-risk range of all treatments in emergency department.

References

[1] Nicole J Krumrei, Myung S Park, Bryan A Cotton, and Martin D Zielinski. Comparison of massive blood transfusion predictive models in the rural setting. *Journal of Trauma and Acute Care Surgery*, 72(1):211–215, 2012.

[2] Mehdi Moazzez Lesko, Tom Jenks, Sara J O'Brien, Charmaine Childs, Omar Bouamra, Maralyn Woodford, and Fiona Lecky. Comparing model performance for survival prediction using total glasgow coma scale and its components in traumatic brain injury. *Journal of neurotrauma*, 30(1):17–22, 2013.

[3] Lisa N MacFadden, Philemon C Chan, Kevin H-H Ho, and James H Stuhmiller. A model for predicting primary blast lung injury. *Journal of trauma and acute care surgery*, 73(5):1121–1129, 2012.

[4] Timothy C Nunez, Igor V Voskresensky, Lesly A Dossett, Ricky Shinall, William D Dutton, and Bryan A Cotton. Early prediction of massive transfusion in trauma: simple as abc (assessment of blood consumption)? *Journal of Trauma and Acute Care Surgery*, 66(2):346–352, 2009.

[5] Herbert Schöchl, Bryan Cotton, Kenji Inaba, Ulrike Nienaber, Henrik Fischer, Wolfgang Voelckel, and Cristina Solomon. Fibtem provides early prediction of massive transfusion in trauma. *Critical Care*, 15(6):R265, 2011.

[6] Mark J Van Der Laan and Daniel Rubin. Targeted maximum likelihood learning. *The International Journal of Biostatistics*, 2(1), 2006.

[7] Antoine Chambaz, Pierre Neuvial, and Mark J van der Laan. Estimation of a nonparametric variable importance measure of a continuous exposure. *Electronic Journal of Statistics*, 6:1059, 2012.

[8] Nastasha Bir, Mathieu Lafargue, Marybeth Howard, Arnaud Goolaerts, Jeremie Roux, Michel Carles, Mitchell J Cohen, Karen E Iles, José A Fernández, John H Griffin, et al. Cytoprotective-selective activated protein c attenuates pseudomonas aeruginosa–induced lung injury in mice. *American journal of respiratory cell and molecular biology*, 45(3):632–641, 2011.

[9] Mitchell J Cohen, Karim Brohi, Carolyn S Calfee, Pamela Rahn, Brian B Chesebro, Sarah C Christiaans, Michel Carles, Marybeth Howard, and Jean-François Pittet. Early release of high mobility group box nuclear protein 1 after severe trauma in humans: role of injury severity and tissue hypoperfusion. *Critical care*, 13(6):R174, 2009.

[10] Mitchell J Cohen, Natasha Bir, Pamela Rahn, Rachel Dotson, Karim Brohi, Brian B Chesebro, Robert Mackersie, Michel Carles, Jeanine Wiener-Kronish, and Jean-François Pittet. Protein c depletion early after trauma increases the risk of ventilator-

associated pneumonia. *Journal of Trauma and Acute Care Surgery*, 67(6):1176–1181, 2009.

[11] John B Holcomb, Deborah J Del Junco, Erin E Fox, Charles E Wade, Mitchell J Cohen, Martin A Schreiber, Louis H Alarcon, Yu Bai, Karen J Brasel, Eileen M Bulger, et al. The prospective, observational, multicenter, major trauma transfusion (prommtt) study: comparative effectiveness of a time-varying treatment with competing risks. *JAMA surgery*, 148(2):127–136, 2013.

[12] John B Holcomb, Barbara C Tilley, Sarah Baraniuk, Erin E Fox, Charles E Wade, Jeanette M Podbielski, Deborah J Del Junco, Karen J Brasel, Eileen M Bulger, Rachael A Callcut, et al. Transfusion of plasma, platelets, and red blood cells in a 1: 1: 1 vs a 1: 1: 2 ratio and mortality in patients with severe trauma: the proppr randomized clinical trial. *Jama*, 313(5):471–482, 2015.

[13] Iván Díaz, Alan Hubbard, Anna Decker, and Mitchell Cohen. Variable importance and prediction methods for longitudinal problems with missing variables. *PloS one*, 10(3):e0120031, 2015.

[14] Mark J Van der Laan and Sherri Rose. *Targeted learning: causal inference for observational and experimental data*. Springer Science & Business Media, 2011.

[15] EC Polley, AE Hubbard, et al. Super learner. *Statistical applications in genetics and molecular biology*, 6:Article25–Article25, 2007.

[16] Trevor Hastie. *gam: Generalized additive models*, version 1.15, 2018. URL `https://CRAN.R-project.org/package=gam`.

[17] Charles Kooperberg. *polspline: Polynomial Spline Routines*, version 1.1.12, 2015. URL `https://CRAN.R-project.org/package=polspline`.

[18] Andy Liaw and Matthew Wiener. Classification and regression by randomforest. *R News*, 2(3):18–22, 2002. URL `https://CRAN.R-project.org/doc/Rnews/`.

[19] David Meyer et al. *Misc Functions of the Department of Statistics, Probability Theory Groupp (Formerly: E1071), TU Wien*, version 1.6-8, 2017. URL `https://CRAN.R-project.org/package=e1071`.