A Data-Adaptive Targeted Learning Approach of Evaluating Viscoelastic Assay Driven Trauma Treatment Protocols

Linqing Wei* Lucy Z. Kornblith† Alan Hubbard‡

Abstract

Estimating the impact of trauma treatment protocols is complicated by the high-dimensional yet finite sample nature of trauma data collected from observational studies. Viscoelastic assays are highly predictive measures of hemostasis.[1] However, the effectiveness of thromboelastography(TEG) based treatment protocols has not been statistically evaluated. To conduct robust and reliable estimation with sparse data, we built an estimation "machine" for estimating causal impacts of candidate variables using the collaborative targeted maximum loss-based estimation(CTMLE) framework.[7] The computational efficiency is achieved by using the scalable version of CTMLE such that the covariates are pre-ordered by summary statistics of their importance before proceeding to the estimation steps.[20] To extend the application of the estimator in practice, we used super learning in combination with CTMLE to flexibly choose the best convex combination of algorithms. By selecting the optimal covariates set in high dimension and reducing constraints in choosing pre-ordering algorithms, we are able to construct a robust and data-adaptive model to estimate the parameter of interest. Under this estimation framework, CTMLE outperformed the other doubly robust estimators(IPW,AIPW,stabilized IPW,TMLE) in the simulation study. CTMLE demonstrated very accurate estimation of the target parameter (ATE). Applying CTMLE on the real trauma data, the treatment protocol (using TEG values immediately after injury) showed significant improvement in trauma patient's hemostasis status (control of bleeding), and a decrease in mortality rate at 6h compared to standard care. The estimation results did not show significant change in mortality rate at 24h after arrival.

Key Words: non-parametric, machine learning, high-dimension,finite sample inference

1. Introduction

Globally, trauma is a leading cause of death and poses both clinical and statistical challenges.[1] Using highly predictive measures to optimize treatment assignment is of great current interest. Given the most common preventable cause of death after trauma is bleeding, important predictor variables are related to controlling bleeding and treating impaired coagulation that can be diagnosed with standard plasma-based lab test like INR(International Normalized Ratio)/PTT(Partial Thromboplastin Time), as well as viscoelastic assays like TEG (Thromboelastography).[2] Viscoelastic assays are used to identify real time abnormalities of clot formation and fibrinolysis, termed trauma-induced coagulopathy.[1] However, the association of abnormalities identified by TEG with TIC (Trauma-Induced-Coagulopathy) are incompletely defined, and specifically the impact of treatment rules assigned based on TEG measures has not been closely examined. The goal of our study was to employ robust, semiparametric data-adaptive modeling procedures (specifically, scalable collaborative targeted minimum loss-based estimation) to estimate the potential impact of various protocols for achieving hemostasis (control of bleeding) and

*Division of Biostatistics, University of California, Berkeley
†Department of Surgery, University of California, San Francisco
‡Division of Biostatistics, University of California, Berkeley

avoiding death.

We focused on estimating the impact of blood products assigned based on TEG measures on patient's hemostasis and mortality status. Since the covariates are high-dimensional, we deployed ensemble machine learning (SuperLearning) methods for modeling the prediction of outcomes versus both adjustment covariates and intervention variables.[4] Given the common problem of high dimension and small sample, standard doubly-robust estimators (such as estimating equation and targeted learning approaches) can break down due to lack of experimentation in the data (positivity violations), caused by high correlation between covariates and exposures in trauma data.[3] We developed estimators of our estimands of interest within the collaborative targeted minimum loss-based estimation framework(CTMLE), which optimizes the variance-bias trade-off when modeling the so-called propensity score, not with respect to prediction intervention, but to the estimand/parameter of interest.[7] This allows for more automated estimation using machine learning in situations with limited data. Also,we built upon the standard CTMLE framework by adding pre-ordering step to boost the computational efficiency. CTMLE demonstrated robust and accurate estimation to the target parameter in the simulation study compared to the other doubly robust estimatros(IPW,AIPW,stablized IPW,TMLE). Also, the super learner version of CTMLE (SL-CTMLE) achieved excellent computational efficiency for high-dimensional adjustment sets. [20]

2. Method

2.1 Data

1671 trauma patients at San Francisco General Hospital were enrolled in an observational cohort study as part of the ongoing prospective longitudinal examination of the activation of coagulation and inflammation after injury. Injury, demographic, clinical and outcome data was collected on arrival and out to 28 days for all patients. TEG parameters from the ED were solely for study purposes and no treatment decisions were based upon their parameters. Massive transfusion (MT) was defined as the transfusion of 10 units of packed red blood cells (pRBCs)/fresh frozen plasma (FFP) in the first 24 hours, and transfusion decisions were at the discretion of the attending trauma surgeon. Blood samples were drawn in citrated vacuum tubes upon arrival to the ED, and then again at 6 hours. Citrate rapid thromboelastography (CRT TEG) and citrated kaolin TEG (CK TEG) were performed with the TEG 5000 immediately after sample collection. The patients were statistically treated with blood product transfusion based on a protocol defined as: 2 units of plasma for ACT(activated clotting time) $>$ 128 seconds, 10 units of cryo (cryoprecipitate) for alpha angle$<$ 65 degrees, and one 1 unit of plt(platelets) for MA(maximum amplitutde) $<$ 55mm.The demographics and baseline measurements are summarized in Table 1.

Table 1: Baseline Covariates and Missingness

Covariates	Mean(sd)	Number of Missing Values(%)
(Injury Survey Score)iss	17.7(15.63)	2(0.12)
Admit base excess	-4.0(6.31)	433(25.91)
Gender, n(%)		
Female	310(18.55)	0(0)
Male	1361(81.45)	0(0)
Age	41(18.6)	3(0.18)
Mechanism, n(%)		
Blunt	960(57.45)	5(0.30)
Penetrating	706(42.25)	0(0)
Systolic blood pressure (sbp, mmHg)	135.9(33.26)	43(2.57)
Heart rate (hr, bpm)	97.6(24.93)	42(2.51)
Race, n(%)		1(0.06)
White	949 (56.79)	
Black	375 (22.44)	
Asian	214(12.81)	
Native American	8(0.48)	
Pacific Islander	9(0.54)	
Other	36(2.15)	
Unknown	79(4.73)	

Table 1: Summary of covariates and missingness of the data.

2.2 Estimation Problem Definition

We aimed to analyze the average treatment effect of patients being on versus off protocol. We modeled the trauma data as following: a collection of baseline covariates, W, containing imputed values and missingness indicators (summarized in Table 1).

Binary outcome Y represented the outcome (e.g. 6h hemostasis, 6h mortality or 24h mortality) of each patient.We used notation $O_i = (W_i, A_i, Y_i)$ to represent each observation unit. The parameter of interest is defined as:

$$\Psi(P_0) = E_0[E_0(Y|A = 1, W) - E_0(Y|A = 0, W)] \quad (1)$$

where Ψ can be interpreted as a causal risk difference under assumptions (W contains all confounders, there is sufficient experimentation of A within strata of W [6], time orders of $W \to A \to Y$). However, it can also be thought of more generally as a measure of importance (akin to comparing adjusted means), without appeal to a causal model. Instead of treating the target estimand as a causal parameter, we take a more conservative perspective by treating the estimand as a statistical association controlling for the baseline covariates W given in Table 1.

For future sections, we will introduce a few new notations: the treatment model is defined as $g_0(a, W) = P_0(A = a|W)$, the conditional mean of the outcome is: $\bar{Q}_0(A, W) = E_0(Y|A, W)$. The empirical version of the above notations are represented as: $g_n(A, W)$ and $\bar{Q}_n(A, W)$

2.3 Model Trauma Data

2.3.1 Treatment Variable

The original data was a sequence of continuous measures (act,alpha, ma,refer to section 3.1) representing the units of each blood product given to patients from 0h

to 24h. By comparing the continuous measures and predefined protocols (cases in Algorithm 1 below), we were able to map the measures into on/off protocol status. We examined three separate rules based upon different sets of variables. Note, that $A^* = d(V)$ could imply both $A^* = 0$ or $A^* = 1$, depending on what the rule indicates given the variables V. In essence, being defined as an intervention where a subject receives a specific level of an intervention, it means that the patient receives the treatment as indicated by the particular rule. One could estimate a parameter equivalent to 1, by defining $A = I(A^* = d(V))$, that is the indicator of receiving the treatment returned by rule, $d(V)$.

Each of the three rules investigated, $d(V)$, are defined as following:

Algorithm 1: Mapping continuous measures to binary on/off protocol status

case 1
if act $\geq$ 128 **and** plasma $>$ 0, A = 1
if act $\geq$ 128 **and** plasma = 0, A = 0
if act = 0 **or** act $<$ 128 **and** plasma = 0, A = 1
else A = 0
case 2
if alpha $\leq$ 65 **and** cryo $>$ 0, A = 1
if alpha $\leq$ 65 **and** cryo = 0, A = 0
if alpha = 0 **and** cryo = 0, A = 1
else A = 0
case 3
if ma $\leq$ 55 **and** plt $>$ 0, A = 1
if ma $\leq$ 55 **and** plt $>$ 0, A = 0
if ma = 0 **and** plt = 0, A = 1
else A = 0

2.3.2 Outcome Variable

We analyzed three different types of outcomes: 1) hemostasis 2) mortality at 6h 3) mortality at 24h. Hemostasis is defined by whether or not the patient received packed red blood cells(RBC) after 6h (e.g. 7-12h interval). If the patient received treatments in 0-6h but did not receive packed RBC in 7-12h, then hemostasis was achieved. If the patient received packed RBC in 7-12h, hemostasis is coded as non-hemostasis. Mortality outcomes were given in the original data (e.g. dead = 1, alive = 0).

2.3.3 Covariates

Baseline covariates used for the analysis are shown in Data section(3.1). Only for theses variables, we imputed the missing values and also added, for each covariate with missing values, an additional basis function as the indicator that a particular observation had the covariate observed versus missing. Thus, one can think of our final list of covariates, W, as the information sufficient to consistently estimate the marginal adjusted (for all the confounders) associations of interest.

2.4 Doubly Robust Estimators

TMLE based estimator has its root in doubly robust estimators.[6] Double Robust(DR) estimators generally refer to estimators of the target parameter of interest that have either $\bar{Q}_0$ or $\bar{g}_0$ part or both being consistent.[8] To find the best DR, we

aim to find the estimator with efficient influence curve. The efficient influence curve of the target parameter in a semiparametric model model is defined as: [11]

$$D^*(\bar{Q}_0, g_0)(O) = \frac{a}{g_0(1,W)} - \frac{1-a}{g_0(0,W)}[Y - \bar{Q}_0(A,W)] + \bar{Q}_0(1,W) - \bar{Q}_0(0,W) - \Psi(P_0) \tag{2}$$

The augumented inverse probability of treatment weighted (AIPTW) estimator is one example of a doubly robust estimator.[16] Using the propensity score and conditional mean of outcome in sample $g_n, \bar{Q}_n$ to define the sample efficient influence influence curve, we have:[16]

$$\sum_{i=1}^{n} \frac{a}{g_n(1,W)} - \frac{1-a}{g_n(0,W)}[Y_i - \bar{Q}_n(A_i,W_i)] + \bar{Q}_n(1,W_i) - \bar{Q}_n(0,W_i) - \Psi_n \tag{3}$$

By setting the equation to 0 and solve for the target parameter directly, we get:

$$\Psi^{AIPTW} = \sum_{i=1}^{n} \frac{a}{g_n(1,W)} - \frac{1-a}{g_n(0,W)}[Y_i - \bar{Q}_n(A_i,W_i)] + \bar{Q}_n(1,W_i) - \bar{Q}_n(0,W_i) \tag{4}$$

However, under the AIPTW setting, the propensity score may inflate the value of the estimator by producing estimations outside of the reality constraints.[10] This in turn, given no inherent constraints on the estimator, values of parameter estimate that can be -1 or 1. Given how sensitive the estimator is to g_n, particularly in small sample sizes, one can benefit from finite sample robustness by using a substitution (plug-in) estimator.

2.5 TMLE

Different from the AIPTW esitmator, TMLE is a plug-in estimator which targets specifically towards the parameter of interest by first generating an initial estimate and then fluctuating the estimator to minimize the bias-variance tradeoff.[15] The estimation procedure can be roughly divided into two steps. The first step involves using specified confounders and intervention variable of interest as regressors to get the initial expectation of outcome, we call $\bar{Q}_n^0(A,W)$.[15] The updating step takes both the initial predicted value as offset and the clever covariates to regress on the expected conditional outcome. Variance can be calculated based on the efficient influence curve and confidence intervals can be constructed accordingly.[15] The detailed algorithm is shown as follwing:

Algorithm 2: TMLE Algorithm

Estimating Step:

1.Estimate initial $\bar{Q}_n^0$

2.Estimate g_n

3.Construct cleaver covariates $H_n(a, W)$ using formula:

$$H_n(a, W) = \frac{a}{g_n(1, W)} - \frac{1 - a}{g_n(0, W)} \tag{5}$$

Targeting Step:

4.Run $Y_i \sim \mathrm{H}_n(A_i, W_i) + logit(\bar{Q}_n^0(A_i, W_i))$, treating the last term as an offset in a regression.

5.Update initial estimator using ϵ_n from step 4

$$\bar{Q}_n^*(A, W) = expit(logit[\bar{Q}_n^0(A, W)] + \epsilon_n H_n(A, W))$$

6.

$$\Psi^{TMLE} = \frac{1}{n} \sum_{i=1}^{n} (\bar{Q}_n^*(1, W_i) - \bar{Q}_n^*(0, W_i))$$

Thus, the estimator augments the original fit of the regression model, by including a "clever covariate", which optimally adjusts for any residual confounding remaining in the original $\bar{Q}_n$ fit and smooths the estimator sufficiently so that it is asymptotically normal. However, though TMLE can be an attractive alternative to estimating equation approaches, it still can suffer from overfitting of g_n. Fortunately, the fact that it is a maximum likelihood-based (or minimum loss-based) procedure, one can use the likelihood (or risk) in the esitmation of $\bar{Q}$ to choose the model for g_n, to insure the estimator does not "blow-up". Generally, procedures based upon TMLE, but using clever selection of the model for g_n based upon the likelihood (or more generally risk) are called *collaborative*.

2.6 Greedy CTMLE

CTMLE is motivated by the sparse data setting. When data is sparse and lack of good support (e.g. lack of positivity),[7] the efficient influence curve can take extremely large value such that the TMLE will have non-linear behavior. One cause for this is that confounders might be instrumental variables, which have no impact on the outcome variable, but are highly correlated with the intervention variable of interest.[9] However, removing the variables with prescreening procedures could hurt the asymptotic behavior of TMLE because there exists statistically significant covariates which lie in the noise level.[7] Most prescreening procedures are based on marginal regression, which result in a bias of the order $n^{-1/2}$. [20] CTMLE does not affect the asymptotic behavior of TMLE, as the "full" model will be chosen as $n \to \infty$.

CTMLE algorithm constructs a sequence of estimators $g_{n,h}, Q_{n,h}^*$ such that the loss functions of both $g_{n,h}, Q_{n,h}^*$ have decreasing values when increasing h. The construction of sequences uses forward greedy search.[7] CTMLE reduces the dimension of covariates that are used to construct the external estimator g_n of g_0 thereby reduces the bias introduced by positivity violations, but one that avoids dropping strong confounders.[9]

2.7 Scalable CTMLE

The original CTMLE uses greedy approach for selecting covariates sequentially. Greedy CTMLE does have robust performance [9] and the time complexity of running greedy CTMLE is $O(p^2)$ when the dimension of covariates is p.[20] In addition, when the dimension of prediction basis (p) approaches to infinity, the large time complexity could lead to memory storage issue and slow down the computation process. We endeavored to utilize an estimator that is both statistically and computationally robust. In this case, if the CTMLE template is constructed by pre-ordering covariates rather than by greedy searching, we could largely reduce the time complexity to O(p).[21] Therefore, we used this *scalable CTMLE* to estimate the average treatment effect based on the data modeled in sections 3.2 and 3.3.

The pre-ordering step is in between of the estimation of $\bar{Q}_0$ and the iterative construction step of $\bar{Q}^*_{n,k}$.[20] We ordered the covariates by the relative decrease in cross-validated risk resulting from the model fit when these variables are included in the clever covariate (5). Specifically, for each covairte W_k as a predictor, we used logistic model to construct $g_{n,k}$ of g_0. Then we defined clever covarite as a function of A and W_k. Then we run regression of Y on clever covariate and offset $\bar{Q}^0(A, W)$ to fit ϵ_k. After defining $\bar{Q}^*_{n,k}$ in the standard way, we computed the empirical cross-validated loss L_k. Each covariate will get a rank based on the increasing value of L_k.[21]

After setting up the rank, we add the covariates one by one to estimate $g_{n,k}$ for k-th covariate.[20] Then we followed the standard CTMLE step to estimate $\bar{Q}^k_n$ and evaluate the empirial loss L_k of $\bar{Q}^*_{n,k}$. If the empirical loss L_k decreased compared to L_{k-1}, we kept adding the next covariate. Otherwise, we replaced $\bar{Q}_{n,k}$ by $\bar{Q}^*_{n,k-1}$. Then we re-add the k-th covariate. The final step was to use cross-validation to find the best estimator from the sequence $\bar{Q}^*_{n,0}$ to $\bar{Q}^*_{n,p}$.[13] The reason such a procedure leads to greater robustness in finite samples it tends to remove covariates in the model for g that behave like instrumental variables, in that they are very highly predictive of A but of little importance in the model for $\bar{Q}$. Such variables result in very small estimated g and since g enters as an inverse in the targeted learning update, they tend to introduce large variance to the estimator with little bias-reduction.

2.8 Super Learner Extension of Scalable CTMLE

The pre-ordering algorithm described in the above scalable CTMLE section used logistic model.[20]To further optimize the performance of scalable CTMLE, we combined multiple ordering algorithms, each generating a unique templates of CTMLE algorithm. Then all the CTMLE algoritms out of mutliple ordering schemes were merged and the one with the lowest cross-validated risk whas chosen estimator.

To combine various algorithms without bringing computational burdens, we relied on super learning, which is an ensemble learning strategy that combines algorithms in a convex manner, such that the time complexity stays as O(p).[14] The convex nature of super learner not only preserves running time efficiency but also preserves scalability as long as each algorithm in the learner library is scalable.[14]

The super learner version of scalable CTMLE is only a small extension added

to the previous algorithm. We followed (Ju, Gruber and Lendle 2016) SL-C-TMLE framework by defining N pre-ordering schemes and for each n-th scheme, we run the same algorithm as in scalable CTMLE. Finally, we used cross-validation to select the best $\bar{Q}_n^*$ out of all (n, k) combinations. [13] We derived the variance and 95% confidence interval coverage based on the efficient influence curve.

We used (Ju 2017)CMTLE package for the analysis.[21] (`https://cran.r-project.org/web/packages/ctmle/ctmle.pdf`). Results can be reproduced from: `https://github.com/WaverlyWei/TEG-CTMLE`

3. Simulation

3.1 Parametric Bootstrap

To demonstrate the robust performance of SL-CTMLE under finite sample and close to positivity violation situations, we used parametric bootstrap to run the simulation. In parametric bootstrap, samples were drawn from a data generating distribution that mimics the true data generating distribution. The bias estimate in parametric bootstrap is defined as: [22]

$$\hat{Bias}_{PB}(\hat{\Psi}, \hat{P}_0, n) = E_{\hat{P}_0}\hat{\Psi}(P_n^{\#}) - \Psi(\hat{P}_0) \tag{6}$$

$\hat{P}_0$ is an estimate of the true data generating distribution P_0. $P_n^{\#}$ is the parametric bootstrap distribution sampled from $\hat{P}_0$. $\Psi(\hat{P}_0)$ is the true parameter applied on $\hat{P}_0$.[24] In practice, since we do not know the true target parameter applied on $\hat{P}_0$, we would replace $\Psi(\hat{P}_0)$ with the estimator of the target parameter applied on the observed data:

$$\Psi(\hat{P}_0) = \hat{\Psi}(P_n) \tag{7}$$

The parametric bootstrap simulation procedure is implemented by first estimating P_0. P_0 includes the estimation of Q_0 and g_0 which further breaks down to estimation of: $P_0(W = w)$, $P(A = a|W = w)$, $P_0(Y = y|A = a, W = w)$. Then we generated bootstrap samples P^n from $\hat{P}_0$. We sampled W with random draws and generated A as $g_n(1|W)$. We sampled Y as $Q_n(Y|A, W)+N(0, 1)$.Then we applied our estimator to the bootstrapped samples P^n and computed the empirical mean across the sample estimates: $E_{\hat{P}_0}\hat{\Psi}(P^n)$

This simulated dataset consisted of W's generated from bivariate normal distribution. [22]

$$W = (W_1, W_2), W_i \sim N(\mu, \Sigma)$$

$$\mu_1 = 1, \mu_2 = 2, \Sigma = \begin{vmatrix} 1 & 1 \\ 2 & 3 \end{vmatrix}$$

$$P(A = 1|W) = \Phi(0.23 + 0.1 * W_1 + 0.35 * W_2)$$

$$P(Y|A, W) = 5.5 + A + 0.5 * W_1 + 4.5 * W_2 + N(0, 1)$$

We generated 200 parametric bootstrap samples each with 1000 obserbations. Then we evaluated the performance of AIPW, IPW, stablized IPW TMLE and SL-CTMLE on the parametric bootstrap samples. The performance of the five estimators are shown in Figure 1 and Table 2:

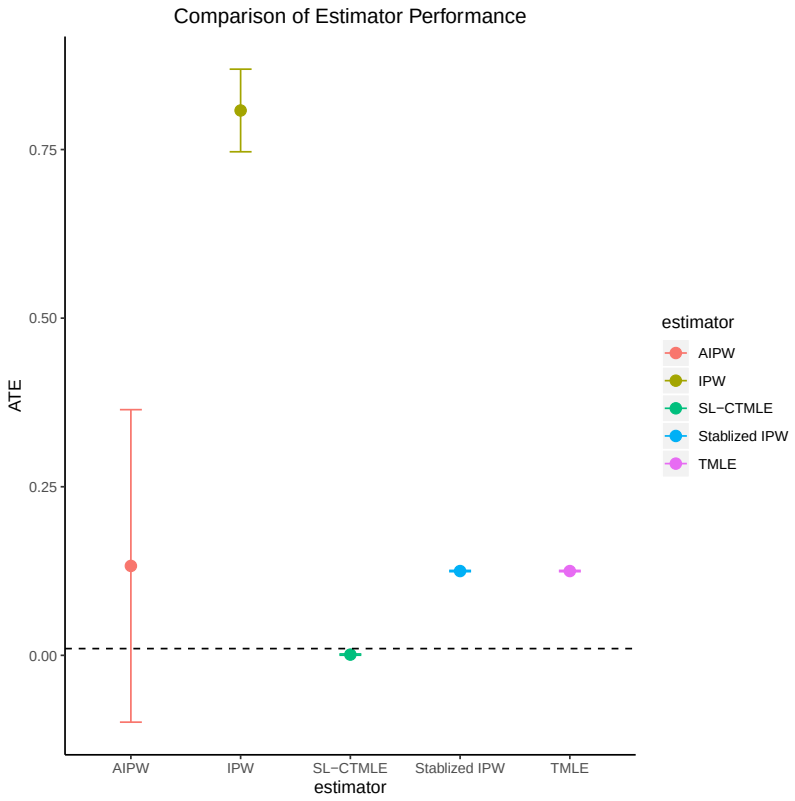

Figure 1: Estimator comparison. Dashed line indicates the true ATE value.

	ATE	SE	Bias	MSE	CI
AIPW	0.133	0.118	0.121	0.0286	(-0.0997, 0.364)
IPW	0.808	0.031	0.796	0.635	(0.747, 0.869)
SL-CTMLE	1.133e-3	4.76e-4	-0.012	1.44e-4	(2.00e-4,2.066e-3)
TMLE	0.125	4.75e-4	0.113	0.0128	(0.124, 0.126)
Stablized IPW	0.127	4.67e-4	0.115	0.0132	(0.124, 0.134)

Table 2: Estimator performance comparison over 200 parametric bootstrap samples

Doubly robust estimator IPW has the largest bias but with relatively narrow confidence interval compared to AIPW. AIPW has the largest standard error out of all five estimators. stablized IPW and TMLE has similar estimation performance but both have larger MSE than SL-CTMLE. SL-CTMLE has the smallest bias, variance ans MSE out of all five estimators. The estimated value from SL-CTMLE was cloest to the true ATE.

Then we compared the time complexity of greedy CTMLE and SL-CTMLE which has pre-ordering embedded. SL-CTMLE took less running time than greedy CTMLE across three different magnitude of sample sizes. The advantage of SL-CTMLE became more prominent when the data became high-dimensional. When p = 100, SL-CTMLE was 10 folds faster than greedy CTMLE. When p = 1000, SL-CTMLE was 50 folds faster than greedy CTMLE.

	N = 10	N = 100	N = 1000
greedy CTMLE	656.25	674.20	1308.75
SL-CTMLE	373.67	442.23	813.82
	p = 10	p = 100	p = 1000
greedy CTMLE	2534.33	480,000.21	5829,613.59
SL-CTMLE	934.85	47,265.28	101,375.31

Table 3: Estimator time complexity benchmarks (in milliseconds) over 100 evaluations. When sample size N = 10, N = 100 and N = 1000, dimension p was fixed (p = 5). When p = 10, p = 100 and p = 1000, sample size N was fixed (N = 100).

4. Results of Data Analysis

Table 4 to Table 6 report the average treatment effect (ATE), p-value and 95% confidence interval coverage of each blood product transfusion based on predefined protocol. Table 4 reports hemostasis as the outcome. Table 5 and Table 6 report mortality at 6 hour and mortality at 24 hour as outcome respectively. Platelet transfusion had 13% significant increase in hemostasis ($p < 0.01$) in the on versus off-protocol, but no significant effect in either mortality at 6h or mortality at 24h. Plasma transfusion showed significant 11% increase in hemostasis and 17% decrease in mortality at 6h ($p<0.01$) but no significant effect in mortality at 24h. Cryoprecipitate transfusion resulted in 20% increase in hemostasis and 16% decrease in mortality at 6h ($p<0.01$) but no significant impact on mortality at 24h.

	Treatment	Adjusted ATE(95% CI)	Unadjusted ATE
Hemostasis	Platelet Protocol	0.133(0.0430, 0.223)	0.2005 (0.183,0.221)
	Plasma Protocol	0.111(0.0914, 0.131)	0.09723 (0.0324,0.127)
	Cryo Protocol	0.203(0.1042, 0.3009)	0.2418(0.1318,0.2845)
Mortality at 6h	Platelet Protocol	-0.1365(-0.3239, 0.0508)	-0.1196(-0.3598,0.0357)
	Plasma Protocol	-0.1692(-0.2159, -0.1224)	-0.2111(-0.3912,-0.1523)
	Cryo Protocol	-0.1584(-0.2347, -0.08216)	-0.09389(-0.1456, 0.1721)
Mortality at 24h	Platelet Protocol	0.00501(-0.2951, 0.3051)	0.02597(-0.0817,0.0945)
	Plasma Protocol	-0.07946 (-0.4196, 0.2606)	-0.1476(-0.3128,0.4289)
	Cryo Protocol	0.04573(-0.1912, 0.2827)	0.04367(-0.0815,0.1023)

Table 4: The ATE of each blood product transfusion with change in hemostasis status as the outcome.Hemostasis is defined based on transfusion of packed RBC (achieved hemostasis == 1), positive value implies increase in the percentage of patients who achieved hemostasis.Death == 1, Alive == 0, negative value implies decrease in death rate

5. Discussion

The treatment protocol based on TEG showed significant improvement in 1671 trauma patient's hemostasis status and decrease in mortality at 6h. The effect of intervention on mortality at 24h is statistically insignificant for all blood product transfusions. In the data used, there are insufficient observations recorded after the first 6 hours to estimate the relative benefits of being on protocol during later periods, for those patients not yet in hemostasis. Between 6 hours to 24 hours, the intervention was possibly confounded by many other treatments received during the same time interval, but the positivity and independence assumptions under causal framework are no longer well-defined due to the sparsity of data between 6 and 24 hours. Although CTMLE is less prone to posivity bias, it can err on the side of insufficient adjustment of real confounding if the sample size insufficient. In any case, it provides essentially as much adjustment as the data will "tolerate."

The estimation of hemostasis and mortality at 6h as outcomes proves the robustness of CTMLE under the high-dimensional, sparse data setting. Traditionally, doubly robust estimators relie on getting a consistent estimation of either g_0 or Q_0 (ideally both) to construct a well-behaved estimator in an asymptotic manner.[17] Standard TMLE updates initial estimation of $\bar{Q}_0$ into $\bar{Q}_n^*$ based on estimator g_n of g_0. This method handles most of the estimation problems well, but breaks down when the data is extremely sparse in high dimensions.[15] To achieve robust estimation even in high dimensions, we optimized the construction step of g_n by incorporating the idea of "dimension reduction." Instead of including all the covariates to build propensity score model, we selected covariates based on a targeted loss function. The additional selection step makes the estimation pipeline more sequential and flexible in a data-driven manner.

To make the CTMLE procedure more computationally efficient, we add an additional layer, ranking, to the selection step. The core idea is that an ordered data structure generally reduces the computational complexity on a large scale. The previous greedy selection algorithm has a second-order time complexity. By pre-ranking the covariates, we could reduce the time complexity to $O(p)$, with p representing the dimension of covariates. [20]

The pre-ordering step leaves room for further improvement. We would like to select the most optimal pre-ordering algorithm. To achieve the goal, we extended the pre-ordering step to a convex combination of learners in learner library, which is called super learning.[14] Super learning does not disrupt the estimation step or the computational efficiency. Instead, it truly makes the estimation procedures non parametric by letting the machine try out all possible combinations of algorithms and eventually select the best estimator using cross-validation.[21]

The current work focuses on analyzing whether the protocol has any impact on trauma patient's hemostasis and mortality. The future work will extend to design the optimal treatment regime based on the scalable CTMLE framework to find out if we can theoretically achieve better performance than the current protocol. Also, the future work will extend the binary treatment/outcome variables to continuous variables to make the estimator even more data-driven and more adaptive in various practical settings.

REFERENCES

1. Hess JR, Holcomb JB, Hoyt DB. Damage control resuscitation: the need for specific blood products to treat the coagulopathy of trauma. Transfusion. 2006; 46:685–686. [PubMed: 16686833]
2. Lesko MM, Jenks T, O'Brien S, Childs C, Bouamra O, Woodford M, Lecky F. Comparing Model Performance for Survival Prediction Using Total GCS and Its Components in Traumatic Brain Injury. J Neurotrauma. 2012
3. Petersen ML, Porter KE, Gruber S, et al. Diagnosing and responding to violations in the positivity assumption. Stat Method Med Res 2012; 21: 31–54.
4. Van der Laan MJ. Targeted maximum likelihood based causal inference: Part I. Int J Biostat 2010; 6: Article 2.
5. Pirracchio R, Yue, JK, Manley, GT. Collaborative targeted maximum likelihood estimation for variable importance measure: Illustration for functional outcome prediction in mild traumatic brain injuries. Statistical Methods in Medical Research 2018; 27: 286–297.
6. Van der Laan MJ and Rose S. Targeted learning: causal inference for observational and experimental data. New York: Springer, 2011.
7. Gruber, S. and M. J. van der Laan (2010a): "An application of collaborative targeted maximum likelihood estimation in causal inference and genomics," The International Journal of Biostatistics, 6, Article 18.
8. Freedman, D. A. and R. A. Berk (2008): "Weighting regressions by propensity scores," Evaluation Review, 32, 392–409
9. Gruber, S. and M. J. van der Laan (2011): "C-tmle of an additive point treatment effect," in Targeted Learning, Springer, 301–321.
10. Porter, K. E., S. Gruber, M. J. van der Laan, and J. S. Sekhon (2011): "The relative performance of targeted maximum likelihood estimators," The International Journal of Biostatistics, 7, Article 31.
11. Robins, J. (2000a): "Robust estimation in sequentially ignorable missing data and causal inference models," in Proceedings of the American Statistical Association: Section on Bayesian Statistical Science, 6–10.
12. Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994): "Estimation of regression coefficients when some regressors are not always observed," Journal of the American statistical Association, 89, 846–866.
13. van der Laan, M. J. and S. Dudoit (2003): "Unified cross-validation methodology for selection among estimators and a general cross-validated adaptive epsilonnet estimator: Finite sample oracle inequalities and examples," , U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 130.
14. van der Laan, M. J., E. C. Polley, and A. E. Hubbard (2007): "Super learner," Statistical Applications in Genetics and Molecular Biology, 6, Article 25.
15. van der Laan, M. J. and D. Rubin (2006): "Targeted maximum likelihood learning," The International Journal of Biostatistics, 2.
16. Robins, J. M. (1986): "A new approach to causal inference in mortality studies with sustained exposure periods - application to control of the healthy worker survivor effect," Mathematical Modelling, 7, 1393–1512.
17. Robins, J. M. and A. Rotnitzky (2001): "Comment on the Bickel and Kwon article, 'Inference for semiparametric models: Some questions and an answer'," Statistica Sinica, 11, 920–936.
18. B. K. Lee, J. Lessler, and E. A. Stuart. Improving propensity score weighting using machine learning. Statistics in Medicine, 29(3):337–346, 2010.
19. Rosenblum M and Van Der Laan MJ. Targeted maximum likelihood estimation of the parameter of a marginal structural model. Int J Biostat 2010; 6: 1557–4679.
20. Ju,C, Gruber S, M.J. van der laan. Scalable Collaborative Targeted Learning for High-dimensional Data. Statisitcal Methods in Medical Research 2017.
21. Ju,C,Wyss,R,Franklin,J.M,et al. Collaborative-controlled LASSO for Constructing Propensity Score-based Estimators in High-Dimensional Data.Statisitcal Methods in Medical Research 2017.
22. Maya,P,Kristin,P,Susan,G,et al. Diagnosing and Responding to Violations in the Positivity Assumption.Statisitcal Methods in Medical Research 2010.
23. Freedman DA, Berk RA. Weighting regressions by propensity scores. Eval Rev. 2008;32(4):392–409.
24. Romain,P,John,Y,Geoffrey,M,et al. Collaborative Targeted Maximum Likelihood Estimation for Variable Importance Measure: Illustration for Functional Outcome Prediction in Mild Traumatic Brain Injuries.Statisitcal Methods in Medical Research 2018.