TMLE for Causal Effects of Multiple-Time-Point Interventions

Mark J. van der Laan

Division of Biostatistics & Department of Statistics
University of California, Berkeley

laan@berkeley.edu
stat.berkeley.edu/∼laan/
targetedlearningbook.com

in collaboration with:
O Stitelman (Media6Degrees), S Gruber (Harvard), V De Gruttola (Harvard)

September 29, 2011

Observed Data Structure

We observe n i.i.d. copies of a longitudinal data structure

$$O = (L(0), A(0), \ldots, L(K), A(K), Y = L(K+1)),$$

where $A(j)$ denotes a discrete valued intervention node, $L(j)$ is an intermediate covariate realized after $A(j-1)$ and before $A(j)$, $j = 0, \ldots, K$, and Y is a final outcome of interest.

For example, $A(j) = (A_1(j), A_2(j))$ could be a vector of two binary indicators of censoring and treatment, respectively.

Factorization of Likelihood

The probability distribution P_0 of O can be factorized according to the time-ordering as

$$\begin{aligned} P_0(O) &= \prod_{k=0}^{K+1} P_0(L(k) \mid Pa(L(k))) \prod_{k=0}^{K} P_0(A(k) \mid Pa(A(k))) \\ &\equiv \prod_{k=0}^{K+1} Q_{0,L(k)}(O) \prod_{k=0}^{K} g_{0,A(k)}(O) \\ &\equiv Q_0 g_0, \end{aligned}$$

where $Pa(L(k)) \equiv (\bar{L}(k-1), \bar{A}(k-1))$ and $Pa(A(k)) \equiv (\bar{L}(k), \bar{A}(k-1))$ denote the parents of $L(k)$ and $A(k)$ in the time-ordered sequence, respectively. The g_0-factor represents the intervention mechanism: e..g, treatment and right-censoring mechanism.

G-computation Formula for Post-Intervention Distribution

Let

$$P^a(l) = \prod_{k=0}^{K+1} Q^a_{L(k)}(\bar{l}(k)), \tag{1}$$

where $Q^a_{L(k)}(\bar{l}(k)) = Q_{L(k)}(l(k) \mid \bar{l}(k-1), \bar{A}(k-1) = \bar{a}(k-1))$.

This is the so called G-computation formula for the post-intervention distribution corresponding with the intervention that sets all intervention nodes $\bar{A}(K)$ equal to $\bar{a}(K)$ in the NPSEM.

Let $L^a = (L(0), L^a(1), \ldots, Y^a = L^a(K+1))$ denote the random variable with probability distribution P^a, and let Y^a be its final component.

Our statistical target parameter is the mean of Y^a: $\Psi(P) = E_{P^a} Y^a$, where $\Psi : \mathcal{M} \to \mathbb{R}$.

This target parameter only depends on P through $Q = Q(P)$. Therefore, we will also denote the target parameter mapping with $\Psi : \mathcal{Q} = \{Q(P) : P \in \mathcal{M}\} \to \mathbb{R}$ so that $\psi_0 = \Psi(Q_0)$.

Other Target Parameters

We can define target parameter as projection of the true dose-response curve $(E_{P^a} Y^a : a \in \mathcal{A})$ onto a working model $\{a \to m_\beta(a) : \beta\}$.

We can also define $E_{P^d} Y^d$ as the treatment-specific mean defined by the G-computation formula for a dynamic treatment d.

Target parameters can then be defined as the projection of the true dose-response curve $(E_{P^d} Y^d : d \in \mathcal{D})$, $\mathcal{D}$ a collection of dynamic treatment rules, onto a working model $\{d \to m_\beta(d) : \beta\}$.

Similarly, we can define target parameters as summary measures of conditional dose-response curves $(E_{P^d}(Y^d \mid V) : d \in \mathcal{D})$, conditioning on some baseline-covariates of interest.

Related classes of target parameters can be defined by history-adjusted marginal structural working models for history adjusted treatment specific means conditional on a history.

One can also define treatment effects of stochastic interventions, and intention to treat interventions, and so on.

A Sequential Regression G-computation Formula

By the iterative conditional expectation rule (tower rule), we have

$$E_{P^a} Y^a = E \ldots E(E(Y^a \mid \bar{L}^a(K)) \mid L^a(K-1)) \ldots \mid L(0)).$$

In addition, the conditional expectation, given $\bar{L}^a(K)$ is equivalent with conditioning on $\bar{L}(K), \bar{A}(K-1) = \bar{a}(K-1)$.

This yields the following sequential regression G-computation formula:
Compute $\bar{Q}^a_Y = E_{Q^a_Y} Y \equiv E(Y \mid \bar{A}(K) = \bar{a}(K), \bar{L}(K))$.

Given $\bar{Q}^a_Y$, we compute
$\bar{Q}^a_{L(K)} = E_{Q^a_{L(K)}}(\bar{Q}^a_Y \mid \bar{L}(K-1), \bar{A}(K-1) = \bar{a}(K-1))$.

This process is iterated: Given $\bar{Q}^a_{L(k+1)}$, we compute
$\bar{Q}^a_{L(k)} = E_{Q^a_{L(k)}}(\bar{Q}^a_{L(k+1)} \mid \bar{L}(k-1), \bar{A}(k-1) = \bar{a}(k-1))$, starting at
$k = K+1$ and moving backwards till the final step $\bar{Q}^a_{L(0)} = E_{Q_{L(0)}} \bar{Q}^a_{L(1)}$.

The Efficient Influence Curve

The pathwise derivative for a path $\{P(\epsilon) : \epsilon\} \subset \mathcal{M}$ through P at $\epsilon = 0$ is defined by $\frac{d}{d\epsilon}\Psi(P(\epsilon))\Big|_{\epsilon=0}$. If for all paths through P, this derivative can be represented as $PD^*(P)S \equiv \int D^*(P)(o)S(o)dP(o)$, where S is the score of the path at $\epsilon = 0$, and $D^*(P)$ is an element of the tangent space at P, then the target parameter mapping is pathwise differentiable at P and its canonical gradient is $D^*(P)$.

An estimator is efficient if and only if it has influence curve equal to canonical gradient:

$$\hat{\Psi}(P_n) - \Psi(P_0) = \frac{1}{n} \sum_{i=1}^{n} D^*(P_0)(O_i) + o_P(1/\sqrt{n}).$$

The canonical gradient forms a crucial ingredient for the construction of double robust semiparametric efficient estimators, and, in particular, for the construction of a TMLE.

Representation of efficient influence curve of target parameter as sum of iteratively defined scores of iteratively defined conditional means

Recall the definition $\bar{Q}^a_{L(k)} = E(Y^a \mid \bar{L}^a(k-1))$, and the recursive relation $\bar{Q}^a_{L(k)} = E_{Q^a_{L(k)}} \bar{Q}^a_{L(k+1)}$.

The efficient influence curve is given by $D^* = \sum_{k=0}^{K+1} D^*_k$, where

$$D^*_{K+1} = \frac{I(\bar{A}(K) = \bar{a}(K))}{g_{0:K}}(Y - \bar{Q}^a_{K+1}),$$

and

$$\begin{aligned} D^*_k &= \frac{I(\bar{A}(k-1) = \bar{a}(k-1))}{g_{0:k-1}} \left\{ \bar{Q}^a_{L(k+1)} - E_{Q^a_{L(k)}} \bar{Q}^a_{L(k+1)} \right\}, \\ &= \frac{I(\bar{A}(k-1) = \bar{a}(k-1))}{g_{0:k-1}} \left\{ \bar{Q}^a_{L(k+1)} - \bar{Q}^a_{L(k)} \right\}, \ k = K, \ldots, 1, \end{aligned}$$

and

$$D^*_0 = \bar{Q}^a_{L(1)} - E_{L(0)} \bar{Q}^a_{L(1)} = \bar{Q}^a_{L(1)} - \Psi(\bar{Q}^a).$$

Formal TMLE algorithm

We assume that $Y \in [0,1]$. A special case would be that Y is binary valued with values in $\{0,1\}$.

We already obtained an initial estimator $\bar{Q}^a_{k,n}$, $k = 0, \ldots, K+1$ and g_n. Let $\bar{Q}^{a,*}_{K+2,n} \equiv Y$.

For $k = K+1$ to $k = 1$, we compute

$$\epsilon_{k,n} \equiv \arg\min_{\epsilon_k} P_n \mathcal{L}_{k,\bar{Q}^{a,*}_{k+1,n}}(\bar{Q}^a_{k,n}(\epsilon_k, g_n)),$$

and the corresponding update $\bar{Q}^{a,*}_{k,n} = \bar{Q}^a_{k,n}(\epsilon_{k,n}, g_n)$.

Finally, $\bar{Q}^{a,*}_{L(0),n} = \frac{1}{n}\sum_{i=1}^n \bar{Q}^{a,*}_{1,n}(L_i(0))$, which is the TMLE of $E_0 Y^a$.

Here $\mathcal{L}_{k,\bar{Q}^a_{k+1}}(\bar{Q}^a_k)$ is the logistic regression log-likelihood for outcome $\bar{Q}^a_{k+1}$:

$$-I(\bar{A}(k-1)=\bar{a}(k-1))\left\{\bar{Q}^a_{k+1}\log\bar{Q}^a_k+(1-\bar{Q}^a_{k+1})\log\{1-\bar{Q}^a_k\}\right\},$$

and $\bar{Q}^a_k(\epsilon_k, g)$ is the logistic-regression submodel

$$\text{Logit}\bar{Q}^a_{L(k)}(\epsilon_k, g)=\text{Logit}\bar{Q}^a_{L(k)}+\epsilon_k\frac{1}{g_{0:k-1}},\ k=K+1,\ldots,0.$$

This choice of loss and submodel does indeed satisfy the generalized score-condition

$$\frac{d}{d\epsilon_k}\mathcal{L}_{k,\bar{Q}^a_{k+1}}(\bar{Q}^a_k(\epsilon_k, g))\Big|_{\epsilon_k=0} = D^*_k, \ k = 0, \ldots, K+1.$$

Consequently, the TMLE $(\bar{Q}^{a,*}_n = (\bar{Q}^{a,*}_{k,n} : k)$ solves the efficient influence curve estimating equation $P_n D^*(\bar{Q}^*_n, g_n) = 0$.

Practical Implementation of the TMLE

- Let g_n be an estimator of g_0.
- Firstly, we carry out a logistic regression regressing Y onto $\bar{A}(K) = \bar{a}(K), \bar{L}(K)$.
- Subsequently, we use this initial estimator of $\bar{Q}^a_{Y,0} = E_0(Y \mid \bar{A}(K) = \bar{a}(K), \bar{L}(K))$ as an off-set in a univariate logistic regression with clever covariate $I(\bar{A}(K) = \bar{a}(K))/g_{0:K,n}$, and fit the corresponding univariate logistic regression of Y among the observations with $\bar{A}(K) = \bar{a}(K)$. This yields the TMLE $\bar{Q}^{a,*}_{Y,n}$ of $\bar{Q}^a_{Y,0} = E_0(Y^a \mid \bar{L}^a(K))$.

- Run a logistic regression of $\bar{Q}^{a,*}_{Y,n}$ onto $\bar{A}(K-1) = \bar{a}(K-1), \bar{L}(K-1)$. Use this initial estimator of $\bar{Q}^{a}_{L(K)} = E(Y^a \mid \bar{L}^a(K-1))$ as an off-set in a univariate logistic regression of $\bar{Q}^{a,*}_{Y,n}$ with clever covariate $I(\bar{A}(K-1) = \bar{a}(K-1))/g_{0:K-1,n}$. The resulting fit $\bar{Q}^{a,*}_{L(K),n}$ is the TMLE of $\bar{Q}^{a}_{L(K),0} = E_0(Y^a \mid \bar{L}^a(K-1))$.
- This process of subsequent estimation of the next conditional mean, given the TMLE-fit of the previous conditional mean, is iterated.

- Thus, for any $k \in \{K+1, \ldots, 1\}$, run a logistic regression of the previous TMLE fit $\bar{Q}^{a,*}_{L(k+1),n}$ onto $\bar{A}(k-1) = \bar{a}(k-1), \bar{L}(k-1)$, and use this fit as an off-set in a univariate logistic regression of $\bar{Q}^{a,*}_{L(k+1),n}$ with clever covariate $I(\bar{A}(k-1) = \bar{a}(k-1))/g_{0:k-1,n}$. Let $\bar{Q}^{a,*}_{L(k),n}$ be the resulting logistic regression fit of $\bar{Q}^{a}_{L(k)}$. This is the TMLE of $\bar{Q}^{a}_{L(k),0}$, $k = K+1, \ldots, 1$.
- Consider now the fit $\bar{Q}^{a,*}_{L(1),n}$ at the $k = 1$-step. This is a function of $L(0)$. We estimate $\bar{Q}^{a}_{L(0)}$ with the empirical mean $\frac{1}{n}\sum_{i=1}^{n} \bar{Q}^{a,*}_{L(1),n}(L_i(0))$.

Here $\bar{Q}_n^{a,*} = (\bar{Q}_{L(k),n}^{a,*}, k = 0, \ldots, K+1)$ is the TMLE of $\bar{Q}_0^a$. The last estimate $\frac{1}{n}\sum_{i=1}^n \bar{Q}_{L(1),n}^{a,*}(L_i(0))$ is the TMLE $\bar{Q}_{L(0),n}^* = \Psi(\bar{Q}_n^{a,*})$ of our target parameter $\bar{Q}_{L(0)}^a = \Psi(\bar{Q}_0^a)$.

Randomized Trials with Time to Event, Right-Censoring, and Time-Dependent Covariates

Drop-out can be informed by baseline covariates and/or time-dependent covariates.

Ignoring these covariates in the estimation procedure results in biased estimators.

Even if drop-out is independent, the TMLE utilizes the covariates by essentially imputing failure times for the censored observations, without inducing bias, even if the imputation models are misspecified.

Simulations

In our simulations we simulate a longitudinal data structure

$$O = (W(0), A(0), N(1), W_4(1), W_5(1), A(1)..., \\ N(K), W_4(K), W_5(K), A(K), N(K+1)).$$

Here $W(0) = (W_1(0), W_2(0), W_3(0), W_4(0), W_5(0))$ are the baseline covariates, $A(0)$ is the binary baseline treatment randomized with probability 0.5, $N(t)$ is the indicator of observing a failure time event at time t, $A(t)$ is the indicator of observing a censoring event at time t, and $W_4(t)$ and $W_5(t)$ are the continuous time-dependent covariates.

In each simulation, 500 simulated data sets with sample size $n = 500$ were generated.

The treatment specific survival curve $S_1(t_0) = 0.469$ at time point $t_0 = 3$ was estimated for six different estimators, and estimates of bias and MSE were reported.

All estimators were supplied consistent estimators of the conditional intensity of the censoring process, and failure-time process, while the conditional distributions of the time-dependent covariates were estimated inconsistently by discretizing the continuous covariates $W_4(t)$, $W_5(t)$, coding these discretized covariates with binary indicators, and estimating the conditional distribution of the binary indicators with logistic parametric regression.

Informative Censoring

The two time-dependent covariates $W_4(t)$ and $W_5(t)$ are generated as follows:

$$W_4(t) = .2A(0) + .5W_1(0) - .4W_2(0) - .4W_3(0) + 2W_4(t-1) + 2W_5(t-1) + U_4$$
$$W_5(t) = .1A(0) + .1W_1(0) + .1W_2(0) - .4W_3(0) + 2W_4(t) + 2W_5(t-1) + U_5,$$

where U_4 and U_5 are i.i.d. $N(0, \sigma = 0.4)$.

The event indicators, $N(t)$, were generated as Bernoulli-indicators with the probability defined by the following conditional intensity of time to failure T:

$$\lambda_T(t) = \text{expit}(-3 + .3A(0) + .3W_1(0) - .3W_2(0) - .3W_3(0) + 2W_4(t-1) + 2W_5(t-1)).$$

The censoring indicators, $A(t)$, were generated as Bernoulli-indicators with the probability defined by the following conditional intensity for censoring for the low and high informative censoring case, respectively:

$$\lambda_C(t) = \text{expit}(-4 + .8A(0) + .3W_1(0) - .3W_2(0) - .3W_3(0) + .1W_4(t) + .1W_5(t-1))$$
$$\lambda_C(t) = \text{expit}(-4 + .8A(0) + .3W_1(0) - .3W_2(0) - .3W_3(0) + 1W_4(t) + 1W_5(t-1)).$$

Informative Censoring

Low Informative Censoring Scenario

	Time Dependent			Baseline		
	TMLE	EE	IPCW	TMLE	A-IPCW	IPCW
Mean	0.469	0.469	0.486	0.475	0.475	0.475
Mean SD	0.027	0.027	0.041	0.027	0.027	0.040
MSE	0.00070	0.00070	0.00113	0.00076	0.00076	0.00077
Coverage	0.942	0.942	0.986	0.940	0.938	0.996

High Informative Censoring Scenario

	Time Dependent			Baseline		
	TMLE	EE	IPCW	TMLE	A-IPCW	IPCW
Mean	0.479	0.470	0.475	0.587	0.585	0.595
Mean SD	0.029	0.035	0.039	0.034	0.034	0.059
MSE	0.00112	0.00440	0.00073	0.01485	0.01453	0.01740
Coverage	0.898	0.898	0.996	0.066	0.074	0.352

Table: Simulation Results For Informative Censoring: Mean of Estimates and Mean Square Error for All Six Estimators

Informative Censoring

In our modified simulation, we generated the censoring events for the low and high informative censoring case as follows:

$$\lambda_C(t) = \text{expit}(-4 + .8A(0) + .3W_1(0) - .3W_2(0) - .3W_3(0) - .01W_4(t) - .01W_5(t-1)),$$
$$\lambda_C(t) = \text{expit}(-4 + .8A(0) + .3W_1(0) - .3W_2(0) - .3W_3(0) - .1W_4(t) - .1W_5(t-1)).$$

Informative Censoring

Low Informative Censoring Scenario

	Time Dependent			Baseline		
	TMLE	EE	IPCW	TMLE	A-IPCW	IPCW
Mean	0.470	0.470	0.452	0.469	0.469	0.470
Mean SD	0.027	0.027	0.040	0.027	0.027	0.042
MSE	0.00065	0.00066	0.00105	0.00068	0.00067	0.00077
Coverage	0.960	0.960	0.974	0.956	0.958	1.000

High Informative Censoring Scenario

	Time Dependent			Baseline		
	TMLE	EE	IPCW	TMLE	A-IPCW	IPCW
Mean	0.468	0.468	0.174	0.432	0.433	0.396
Mean SD	0.027	0.027	0.026	0.033	0.033	0.067
MSE	0.00068	0.00067	0.08732	0.00251	0.00241	0.00731
Coverage	0.960	0.960	0.000	0.798	0.810	0.836

Table: Simulation Results For Informative Censoring: Mean of Estimates and Mean Square Error for All Six Estimators

Independent Censoring

The hazard of censoring was now only a function of time, so that censoring is independent of the evolving processes, but three different hazards were considered representing different levels of independent censoring: no censoring, medium censoring, and high censoring.

In the first scenario every individual was left uncensored. In the second and third scenario each subject was censored with 20 percent probability (Medium Censoring Scenario) and 60 percent probability (High Censoring Scenario), respectively.

No Censoring Scenario

	Time Dependent			Baseline		
	TMLE	EE	IPCW	TMLE	A-IPCW	IPCW
Mean	0.468	0.468	0.468	0.468	0.468	0.468
Mean SD	0.027	0.027	0.038	0.027	0.027	0.038
MSE	0.00067	0.00068	0.00073	0.00069	0.00069	0.00073
Coverage	0.952	0.952	0.990	0.950	0.950	0.990

Medium Censoring Scenario

	Time Dependent			Baseline		
	TMLE	EE	IPCW	TMLE	A-IPCW	IPCW
Mean	0.469	0.470	0.471	0.469	0.469	0.470
Mean SD	0.028	0.028	0.051	0.029	0.029	0.051
MSE	0.00070	0.00072	0.00120	0.00081	0.00081	0.00106
Coverage	0.960	0.960	0.996	0.952	0.952	1.000

High Censoring Scenario

	Time Dependent			Baseline		
	TMLE	EE	IPCW	TMLE	A-IPCW	IPCW
Mean	0.474	0.481	0.474	0.467	0.467	0.466
Mean SD	0.044	0.047	0.114	0.043	0.042	0.112
MSE	0.00110	0.00248	0.00712	0.00196	0.00197	0.00496
Coverage	0.988	0.988	0.978	0.940	0.940	0.984

Misspecified models for both censoring and failure times

For the study here we evaluate what happens to the simulation results when the time dependent covariates, W_4, W_5, and then both W_4 and W_5 are removed from the models for the initial estimates of Q_n and g_n.

This allows us to observe how the different estimators behave when the initial estimates for Q_n and g_n are both initially mis-specified.

Correctly Specifying Initial Models

	Time Dependent			Baseline		
	TMLE	EE	IPCW	TMLE	A-IPCW	IPCW
Mean	0.468	0.468	0.174	0.432	0.433	0.396
Mean SD	0.027	0.027	0.026	0.033	0.033	0.067
MSE	0.00068	0.00067	0.08732	0.00251	0.00241	0.00731
Coverage	0.960	0.960	0.000	0.798	0.810	0.836

Removing $W_4(t)$ From Initial Model Specification

	Time Dependent			Baseline		
	TMLE	EE	IPCW	TMLE	A-IPCW	IPCW
Mean	0.457	0.455	0.172	0.420	0.421	0.411
Mean SD	0.034	0.063	0.026	0.035	0.036	0.067
MSE	0.00133	0.01211	0.08893	0.00360	0.00359	0.00512
Coverage	0.900	0.900	0.000	0.740	0.740	0.910

Table: Simulation Results For Independent Censoring: Mean of Estimates and Mean Square Error for All Six Estimators

Removing $W_5(t)$ From Initial Model Specification

	Time Dependent			Baseline		
	TMLE	EE	IPCW	TMLE	A-IPCW	IPCW
Mean	0.459	0.461	0.173	0.411	0.411	0.396
Mean SD	0.034	0.066	0.026	0.038	0.038	0.065
MSE	0.00133	0.01649	0.08840	0.00467	0.00465	0.00725
Coverage	0.920	0.920	0.000	0.640	0.650	0.810

Removing $W_4(t)$ and $W_5(t)$ From Initial Model Specification

	Time Dependent			Baseline		
	TMLE	EE	IPCW	TMLE	A-IPCW	IPCW
Mean	0.462	1.243	0.357	0.405	0.405	0.403
Mean SD	0.616	0.619	0.056	0.038	0.038	0.063
MSE	0.00472	1.02729	0.01415	0.00549	0.00549	0.00604
Coverage	1.000	1.000	0.440	0.590	0.600	0.870

Table: Simulation Results For Independent Censoring: Mean of Estimates and Mean Square Error for All Six Estimators

Analysis of Tshepo Study

Our initial analysis of the Tshepo study was based on TMLEs of the causal effect of the treatment on survival, and corresponding effect-modification parameters, only adjusting for the baseline covariates.

We extend here this TMLE to account for potential bias due to informative censoring by time-dependent covariates CD4 and viral load that have an effect on both time to drop-out and the time to event of interest.

We will directly compare results using this TMLE that only incorporates the baseline covariates to the TMLE that accounts for time dependent confounding in the form of informative censoring due to the time-dependent covariates. Moreover, we will compare these results to results based on an IPCW estimator and a locally efficient double robust estimating equation based estimator.

For the analysis performed here we evaluate the effect modification of gender on the two cART treatments for two outcomes of interest:

1. Time to death censored by treatment modification or end of study (DEATH).
2. Time to minimum of virologic failure, death, or treatment modification censored by end of study (TLOVR).

For each of the two time to event outcomes we will estimate the difference in additive risk by gender at 36 months after randomization to cART therapy. We will estimate this parameter using the six estimators examined in the simulation analysis.

Prior to doing this analysis we expected that utilizing the time-dependent covariates should have a small effect on the estimates for the TLOVR outcome since censoring is independent for this time to event outcome.

On the other hand, the time to death is subject to censoring by time to treatment modification which is expected to be informed by CD4 and viral load, so that one might expect a bias reduction for the new TMLE relative to the previously implemented TMLE that only incorporated the baseline covariates.

Results

Risk Difference @ 36 Months

	Time Dependent			Baseline		
	TMLE	DR-EE	IPCW	TMLE	DR-EE	IPCW
Est	20.0%	20.1%	18.3%	18.9%	18.9%	18.3%
SE	5.0%	4.9%	10.3%	4.9%	4.9%	10.3%
p	<0.001	<0.001	0.074	<0.001	<0.001	0.074

Table: Gender Effect Modification on TLOVR

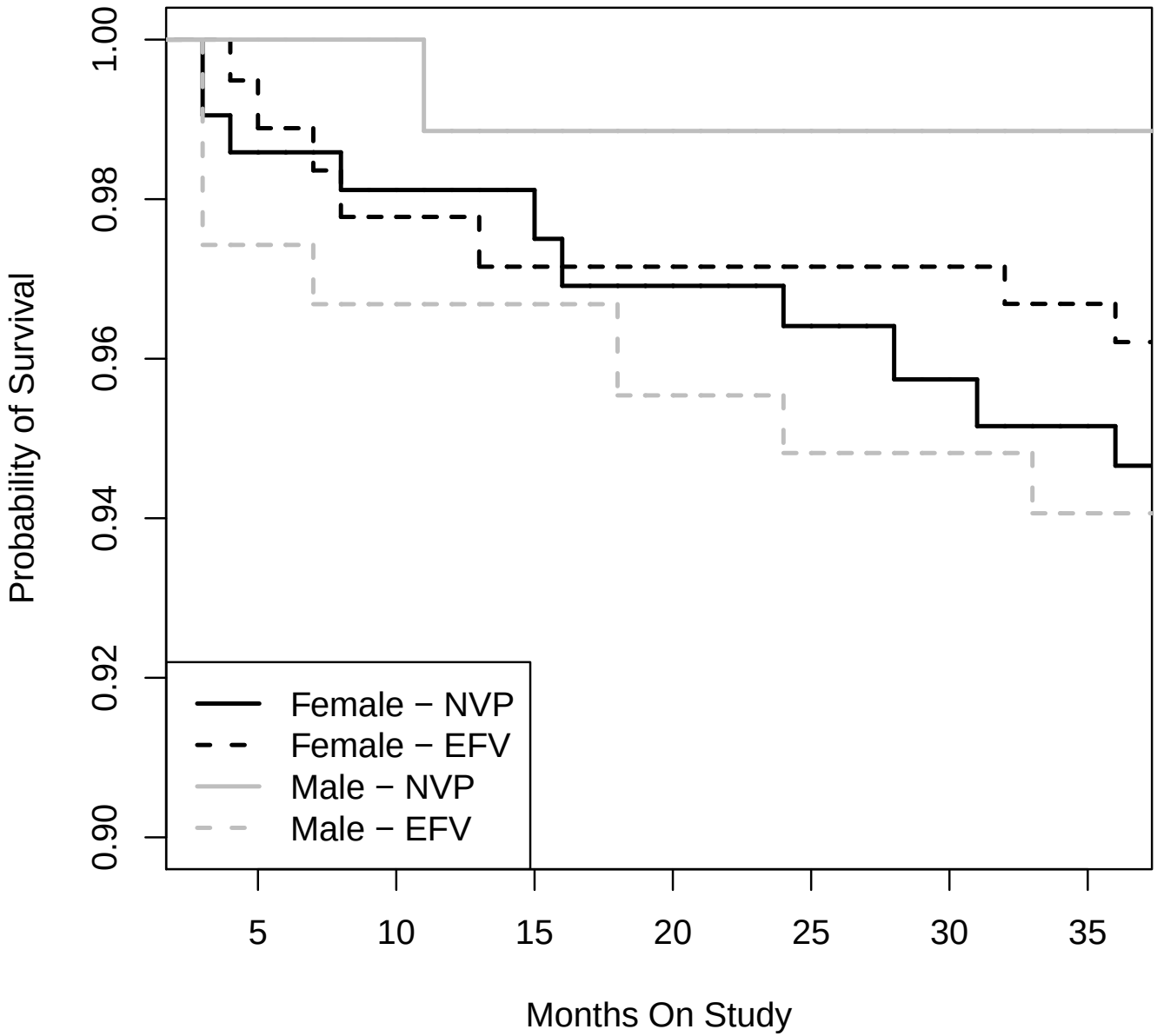
Results

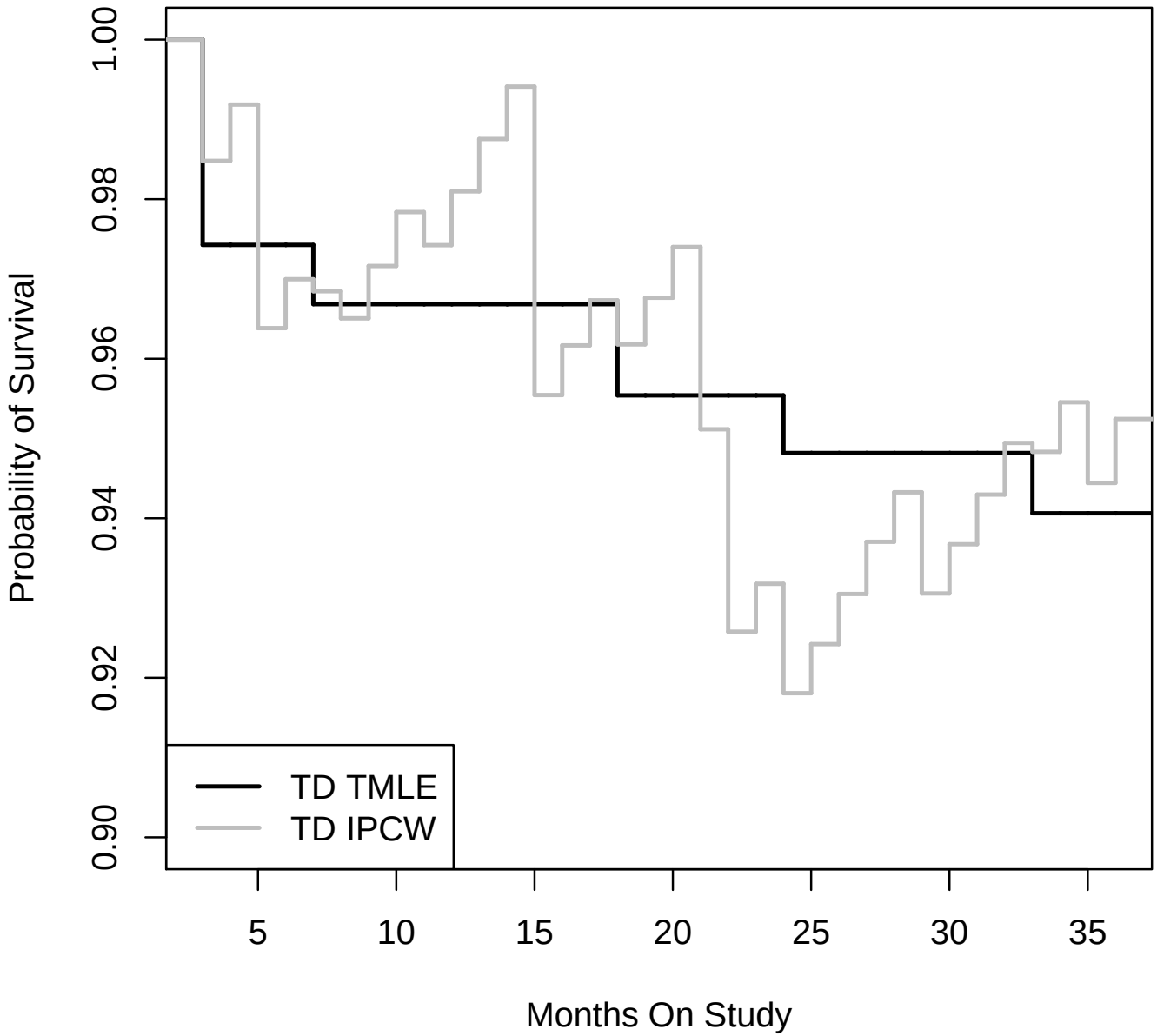
Risk Difference @ 36 Months

	Time Dependent			Baseline		
	TMLE	DR-EE	IPCW	TMLE	DR-EE	IPCW
Est	6.3%	6.5%	5.2%	5.1%	5.1%	5.2%
SE	2.3%	2.3%	12.5%	2.4%	2.4%	12.5%
p-value	0.005	0.004	0.680	0.029	0.030	0.680

Table: Gender Effect Modification on Death

The TD TMLE results indicate that gender does in fact modify the effect of the drug treatment EFV/NVP and the difference in the effect between males and females at 36 months is 6.3 percent.


Probability of Survival
1.00
0.98
0.96
0.94
0.92
0.90
5
10
15
20
25
30
35
Months On Study
Female − NVP
Female − EFV
Male − NVP
Male − EFV


Probability of Survival
1.00
0.98
0.96
0.94
0.92
0.90
5
10
15
20
25
30
35
Months On Study
TD TMLE
TD IPCW

Concluding Remarks

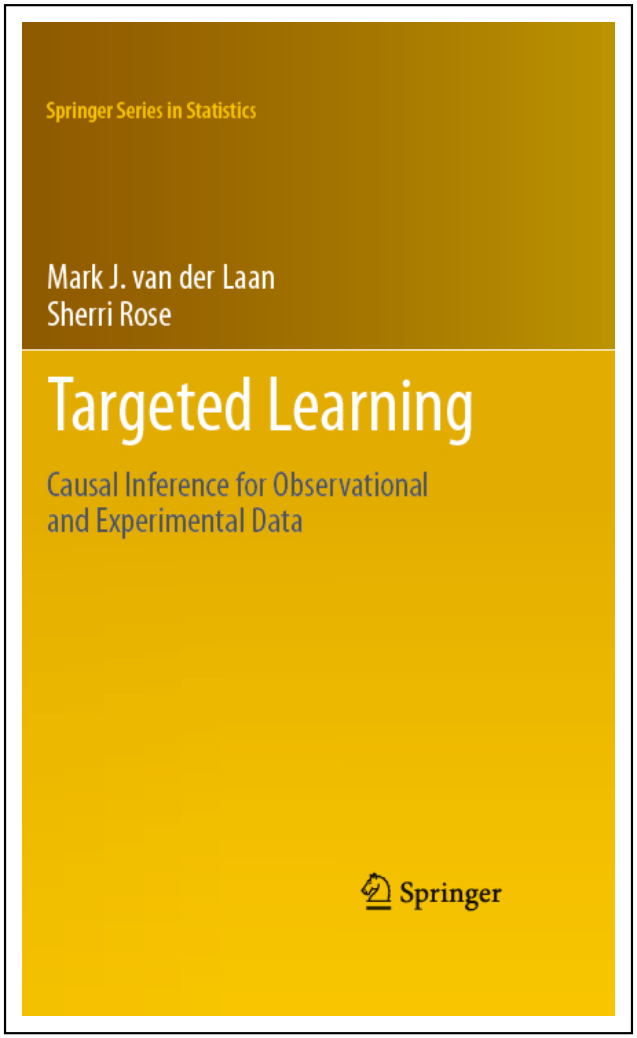
- TMLE provides a template for construction of efficient substitution estimators.
- The basis of TMLE is a loss-function, a submodel for fluctuation, and a procedure for iteratively minimizing the empirical risk along the fluctuation model through a current estimator.
- The submodel is typically chosen so that its score spans the efficient influence curve of the target parameter.

- Sequential regression provides a particularly effective way of representing post-intervention distributions and thereby causal effects. The TMLE for a sequential regression substitution estimator was presented above.
- The TMLE for a maximum likelihood based substitution estimator has also been developed, and the above simulation and data analysis results were for this TMLE. It is harder to implement.
- All TMLE are double robust and efficient, but different types of TMLE can differ in their finite sample performance, and the sequential regression TMLE might have finite sample advantages by only estimating what is really needed.

Targeted Learning Book

`targetedlearningbook.com`

MJ van der Laan, S Rose (2011), Targeted Learning: Causal Inference for Observational and Experimental Data, Springer: New York.



Key References

- MJ van der Laan, S Gruber (2011), "Targeted Minimum Loss Based Estimation of an Intervention Specific Mean Outcome," UC Berkeley Division of Biostatistics Working Paper Series, `http://www.bepress.com/ucbbiostat/paper290/`.
- OM Stitelman, V De Gruttola, MJ van der Laan (2011). "A General Implementation of TMLE for Longitudinal Data Applied to Causal Inference in Survival Analysis," UC Berkeley Division of Biostatistics Working Paper Series, `http://www.bepress.com/ucbbiostat/paper281/`.
- H Bang, J Robins (2005), "Doubly robust estimation in missing data and causal inference models," *Biometrics*, 61:692-972.