Targeted Learning

Causal Inference for Observational and Experimental Data

Mark van der Laan
http://www.stat.berkeley.edu/~laan/
University of California, Berkeley
INSERM workshop, Bordeaux,
June 6-8, 2011

Complications of Human Art in Statistics

1. The parametric model is misspecified.
2. The target parameter is interpreted as if parametric model is correct.
3. The parametric model is often data-adaptively (or worse!) selected, and this part of the estimation of procedure is not accounted for in the variance.

Estimation is a Science, *Not an Art*

1. **Data:** realizations of random variables with a probability distribution.
2. **Model:** actual knowledge about the data-generating probability distribution.
3. **Target Parameter:** a feature of the data-generating probability distribution.
4. **Estimator:** an *a priori*-specified algorithm, benchmarked by a dissimilarity-measure (e.g., MSE) w.r.t. target parameter.

Targeted Learning

- Avoid reliance on human art and non-realistic (parametric) models
- Define interesting parameters
- Target the fit of data-generating distribution to the parameter of interest
- Statistical Inference

TMLE/SL

Targeted Maximum Likelihood
coupled with Super Learner methodology

TMLE/SL Toolbox

Targeted effects

- Effect of static or dynamic treatments (e.g. on survival time)
- Direct and Indirect Effects
- Parameters of Marginal Structural Models
- Variable importance analysis in genomics

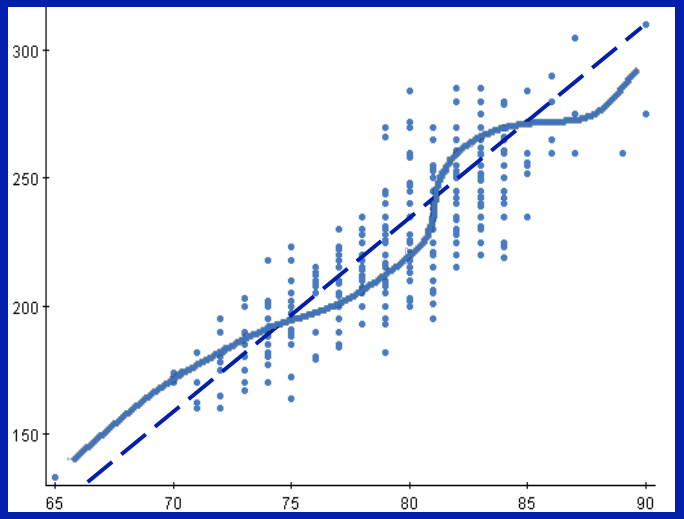
Types of data

- Point treatment
- Longitudinal/Repeated Measures
- Censoring/Missingness/Time-dependent confounding.
- Case-Control
- Randomized clinical trials and observational data

Two-stage Methodology: SL/TMLE

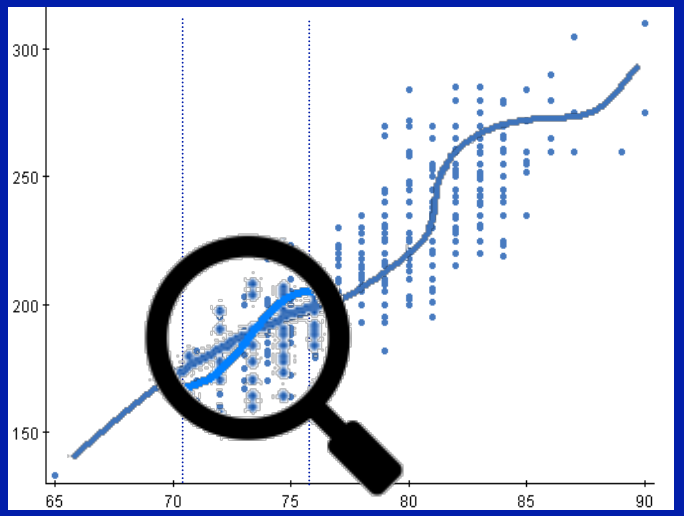
1. Super Learning

- Uses a library of estimators
- Builds data-adaptive weighted combination of estimators
- Weights are optimized based on loss-function specific cross-validation to guarantee best overall fit



2. Targeted Maximum Likelihood Estimation

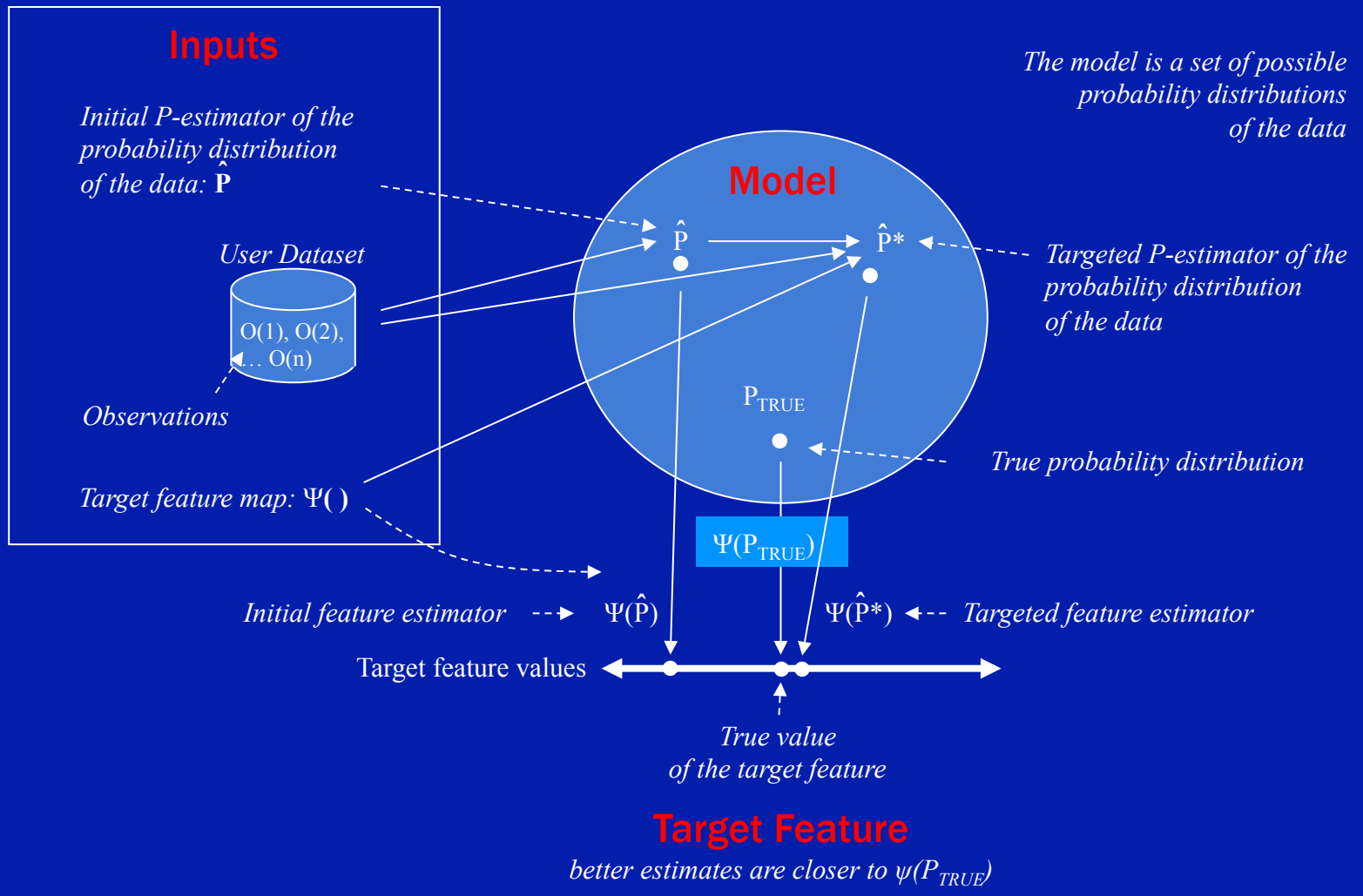
- Zooms in on one aspect of the estimator—the target feature
- Removes bias for the target.



Targeted Maximum Likelihood

- MLE/SL aims to do well estimating whole density
- Targeted MLE aims to do well estimating the parameter of interest
  - General decrease in bias for parameter of Interest
  - Fewer false positives
  - Honest *p*-values, inference, multiple testing

Targeted Maximum Likelihood Estimation Flow Chart



Targeted MLE

1. Identify optimal parametric model for fluctuating initial $\hat{P}$
   - Small “fluctuation” -> maximum change in target
2. Given strategy, identify optimum amount of fluctuation by MLE
3. Apply optimal fluctuation to $\hat{P}$ -> 1st-step targeted maximum likelihood estimator
4. Repeat until the incremental “fluctuation” is zero
   - Some important cases: 1 step to convergence
5. Final probability distribution solves efficient influence curve equation

→ T-MLE is double robust & locally efficient

Targeted Minimum Loss Based Estimation (TMLE)

$\Psi(Q_0)$ target parameter

$Q_0 = \arg\min_Q P_0 L(Q) \equiv \int L(Q)(o) dP_0(o)$

$\hat{Q}(P_n)$: Initial estimator, Loss-based SL

$\{\hat{Q}_g(\epsilon) : \epsilon\}$ fluct. model for fitting ψ_0

$\hat{g} = \hat{g}(P_n)$ loss based SL of treatment/cens mech

$\frac{d}{d\epsilon} L(\hat{Q}_{\hat{g}}(\epsilon))\Big|_{\epsilon=0} = D^*(\hat{Q}, \hat{g})$

$\epsilon_n = \arg\min_\epsilon P_n L(\hat{Q}_{\hat{g}}(\epsilon))$

Iterate till convergence: $\hat{Q}^*$

Solves efficient influence curve equation:

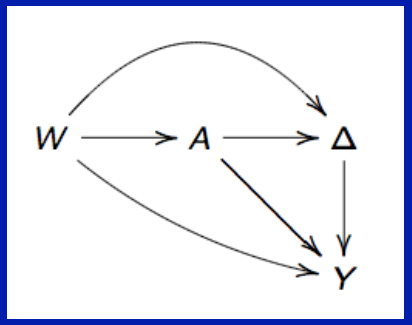
$P_n D^*(\hat{Q}^*, \hat{g}) = 0$

TMLE: $\Psi(\hat{Q}^*)$

TMLE for Average Causal Effect

Non-parametric structural equation model for a point treatment data structure with missing outcome.

$$
\begin{aligned}
W &= f_W(U_W) \\
A &= f_A(W, U_A) \\
\Delta &= f_A(W, A, U_\Delta) \\
Y &= f_Y(W, A, \Delta, U_Y).
\end{aligned}
$$



We can now define counterfactuals Y(1,1) and Y(0,1) corresponding with interventions setting A and Δ.

We assume U_A and U_Δ independent of U_Y given W.

The additive causal effect EY(1)-EY(0) equals:
Ψ(P)=E[E(Y|A=1, Δ=1, W)-E(Y|A=0, Δ=1, W)]

TMLE for Average Causal Effect

- Our first step is to generate an initial estimator P_n^0 of P; we estimate E(Y|A, Δ=1, W) with super learning.
- We fluctuate this initial estimator with a logistic regression:

$$\text{logit } P_n^0(\epsilon)(Y = 1 \mid A, \Delta = 1, W) = \text{logit } P_n^0(Y = 1 \mid A, \Delta = 1, W) + \epsilon h$$

where

$$h(A, W) = \frac{1}{\Pi(A, W)} \left(\frac{A}{g(1 \mid W)} - \frac{1 - A}{g(0 \mid W)} \right)$$

and

$g(1 \mid W) = P(A = 1 \mid W)$ Treatment Mechanism
$\Pi(A, W) = P(\Delta = 1 \mid A, W)$ Missingness Mechanism

- Let ε_n be the maximum likelihood estimator and $P_n^* = P_n^0(\varepsilon_n)$. The TMLE is given by $\Psi(P_n^*)$.

TMLE of Mean when Outcome is Missing at Random

Kang and Shafer debate

Kang and Schafer, 2007

n i.i.d. units of $O = (W, \Delta, \Delta Y) \sim P_0$

W is a vector of 4 baseline covariates

Δ is an indicator of whether the continuous outcome, Y, is observed.

Parameter of interest

$$\mu(P_0) = E_0(Y) = E_0(E_0(Y \mid \Delta = 1, W))$$

Observed covariates:

$$W_1 = \exp(Z_1 / 2)$$
$$W_2 = Z_2 / (1 + \exp(Z_1)) + 10$$
$$W_3 = (Z_1 Z_3 / 25 + 0.6)^3$$
$$W_4 = (Z_2 + Z_4 + 20)^2$$

where $Z_1, \ldots, Z_4 \sim N(0, 1)$ independent

$$Y = 210 + 27.4 Z_1 + 13.7 Z_2 + 13.7 Z_3 + 13.7 Z_4 + N(0, 1)$$

$$g_0(1 \mid W) = P(\Delta = 1 \mid W) = \text{expit}(-Z_1 + 0.5 Z_2 - 0.25 Z_3 - 0.1 Z_4)$$

$g_0(1 \mid W)$ between (0.01, 0.98)

TMLE for Binary Y

- A semi-parametric efficient substitution estimator that respects bounds:

$$\mu_{n,TMLE} = \frac{1}{n}\sum_{i=1}^{n} \overline{Q}_n^*(W_i).$$

$$\log it\overline{Q}_n^*(W) = \log it\overline{Q}_n^0(W) + \varepsilon h(1,W).$$

where $h(1,W) = \frac{1}{g_n(1 \mid W)}$.

  - ε is estimated by maximum likelihood,
  - Loss function:

$$-L(\overline{Q})(Oi) = \Delta\{Y \log \overline{Q}(W) + (1-Y)\log(1-\overline{Q}(W))\}$$

We use machine learning (preferably super learner) for $\overline{Q}_n^0$ and for g_n if the missingness mechanism is unknown.

TMLE for Continuous $Y \in [0,1]$

- If $Y \in [0,1]$, we can implement this same TMLE as we would for binary Y.

We use the same logistic fluctuation as defined on the previous slide, using standard software for logistic regression and simply ignoring that Y is not binary. The same loss function is still valid (Gruber and van der Laan, 2010).

- If Y is bounded between (a,b), then we transform it into Y*=(Y-a)/(b-a)

Kang and Schafer

Modification 1

Modification 2

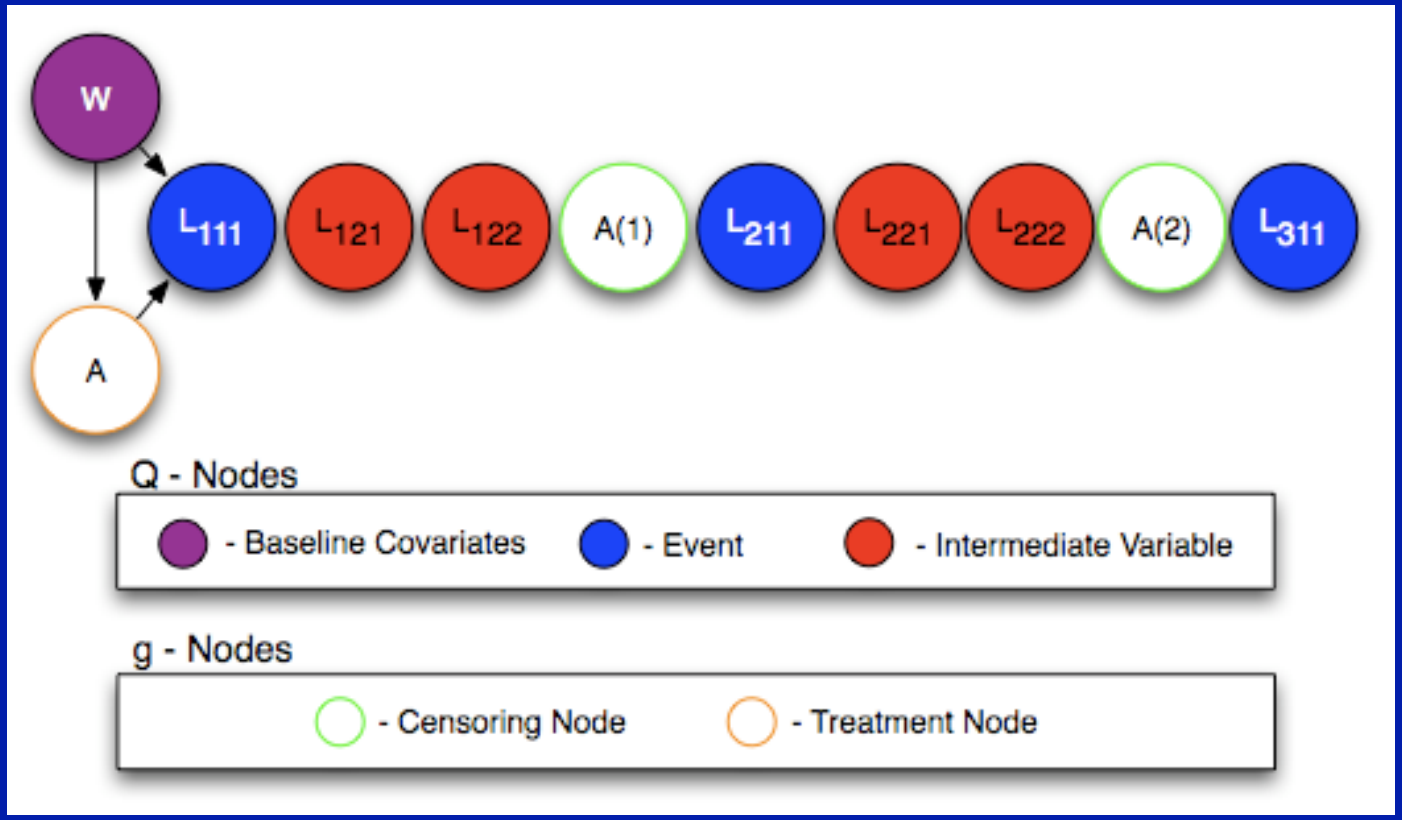
Targeted Maximum Likelihood Learning for Time to Event Data, Accounting for Time Dependent Variables: Analyzing the Tshepo RCT

Ori M. Stitelman, Victor DeGruttolas,
Mark J. van der Laan
Division of Biostatistics, UC Berkeley

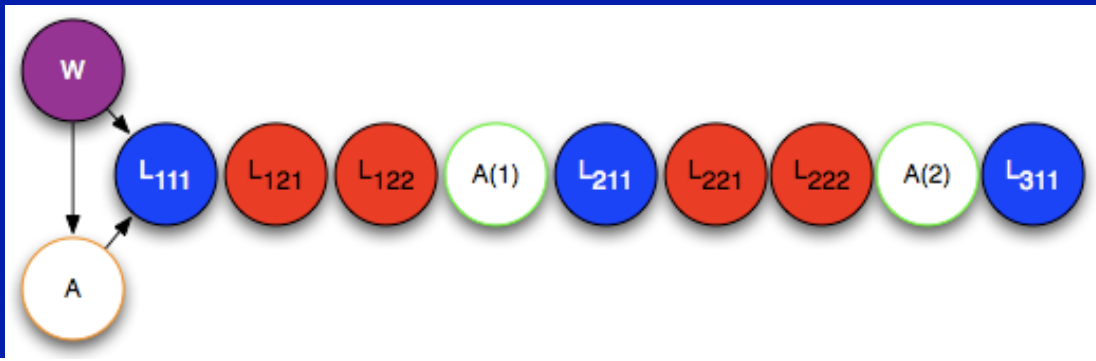
Data Structure

- n i.i.d copies of O = (A,W,(A(t):t),(L(t):t)) ~ p_0
- A – Treatment – HIV cART therapy (EFV/NVP)
- W=L(0) – Baseline Covariates – Sex, VL, BMI
- A(t) – Binary Censoring Variables
  - Equals 1 When Individual is Censored.
  - Equals 0 at all time when individual is not censored.
  - $\underline{A}(t)$ is equal to the history of A(t)
- L(t) – Failure time event process, and time-dependent process (CD4+, Viral Load)
  - L(t) is defined as (L(s):s $\leq$ t).
  - $\bar{W}$e code L(t) with binaries.

Causal Graph For 3 Time Points

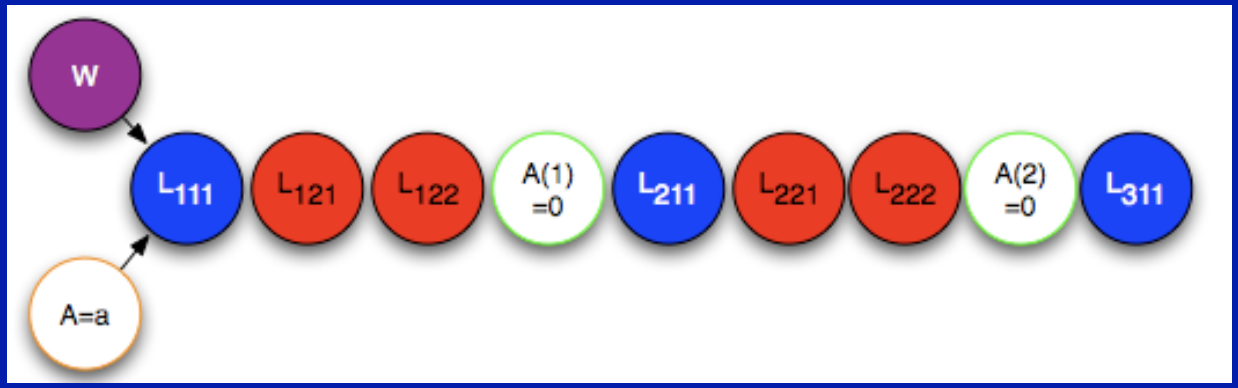


Likelihood of the Observed Data



$$p_0(O) = \overbrace{P(W)}^{Q_W} \overbrace{P(A \mid W)}^{g_A} \prod_{t=1}^{t_k} \prod_{j=1}^{n(t)} \prod_{l=1}^{n(t,j)-1} \overbrace{P(L(t,j,l) \mid Pa(L(t,j,l)))}^{Q_{L(t,j,l)}} \prod_{t=1}^{t_k-1} \overbrace{P(A(t) \mid Pa(A(t)))}^{g_{A(t)}}$$

G-computation Formula



$$p_{0,a,0}(O) = \overbrace{P(W)}^{Q_W} \prod_{t=1}^{t_k} \prod_{j=1}^{n(t)} \prod_{l=1}^{n(t,j)-1} \overbrace{P(L(t,j,l) \mid Pa(L(t,j,l)), A=a, A(t-1)=0)}^{Q_{L(t,j,l),a,0}}$$

Parameter of Interest

- Treatment specific survival curve:

$$\begin{aligned}\Psi(Q_0) &= P(T_{a,0} > t_k) \\ &= E_W P(T_{a,0} > t_k \mid W) \\ &= E_W P(L(t_k, 1, 1)_{a,0} = 0 \mid W)\end{aligned}$$

$$E_{L(0)} \sum_{L(t_k,1,1)=0} \prod_{t=1}^{t_k} \prod_{j=1}^{n(t)} \prod_{l=1}^{n(t,j)-1} Q_{L(t,j,l)}(l(t,j,l) \mid Pa(l(t,j,l)), A = a, \bar{A}(t-1) = 0)$$

Simulations of TMLE of causal effect of treatment on survival accounting for time-dependent covariates

- Compare TMLE with Estimating Equation (EE) and IPCW, both with and without the incorporation of time-dependent covariates

Simulations with informative censoring. The precise data-generating mechanism is described as follows.

(1) Drawing baseline covariates $W(0)$ involved first generating from a mean-zero multivariate normal and truncating any component from above by 2 and from below by -2. The covariance matrix was defined as 1 on the diagonal and 0.2 off the diagonal. The truncation was enforced to ensure that the censoring mechanisms were not suffering from practical violations of the positivity assumption, as required for identifiability of $S_1(t_0)$.

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

$$W_4(t) = 0.2A(0) + 0.5W_1(0) - 0.4W_2(0) - 0.4W_3(0) + 2W_4(t-1) + 2W_5(t-1) + U_4$$
$$W_5(t) = 0.1A(0) + 0.1W_1(0) + 0.1W_2(0) - 0.4W_3(0) + 2W_4(t) + 2W_5(t-1) + U_5,$$

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

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

$$\lambda_T(t) = \text{expit}(-3+0.3A(0)+0.3W_1(0)-0.3W_2(0)-0.3W_3(0)+2W_4(t-1)+2W_5(t-1)).$$

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

$$\lambda_C(t) = \text{expit}(-4 + 0.8A(0) + 0.3W_1(0) \\ -0.3W_2(0) - 0.3W_3(0) - 0.01W_4(t) - 0.01W_5(t-1)),$$

$$\lambda_C(t) = \text{expit}(-4 + 0.8A(0) + 0.3W_1(0) \\ -0.3W_2(0) - 0.3W_3(0) - 0.1W_4(t) - 0.1W_5(t-1)).$$

	Time-dependent			Baseline		
	TMLE	EE	IPCW	TMLE	A-IPCW	IPCW
Low informative						
Mean of Estimates	0.471	0.471	0.451	0.469	0.469	0.469
MSE	0.00070	0.00073	0.00127	0.00082	0.00081	0.00093
Highly informative						
Mean of Estimates	0.472	0.472	0.172	0.436	0.437	0.394
MSE	0.00066	0.00067	0.08864	0.00215	0.00210	0.00773

Simulation results for low and highly informative censoring

Simulations with independent censoring. The data-generating distribution was the same as above, except the censoring mechanism was modified. The hazard of censoring was only a function of time, such that censoring was 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, each individual was left uncensored. In the second and third scenario each subject was censored with either 20% probability (medium) or 60% probability (high).

	Time-dependent			Baseline		
	TMLE	EE	IPCW	TMLE	A-IPCW	IPCW
No censoring						
Mean of Estimates	0.469	0.469	0.469	0.468	0.468	0.469
MSE	0.00047	0.00047	0.00054	0.00048	0.00048	0.00054
Medium censoring						
Mean of Estimates	0.467	0.467	0.470	0.469	0.469	0.468
MSE	0.00063	0.00086	0.00203	0.00093	0.00093	0.00169
High censoring						
Mean of Estimates	0.476	0.477	0.477	0.464	0.464	0.466
MSE	0.00111	0.00315	0.00566	0.00180	0.00181	0.00417

Simulation results for independent censoring

Removing $W_4(t)$ From Initial Model Specification

	Time Dependent			Baseline		
	TMLE	EE	IPCW	TMLE	A-IPCW	IPCW
Mean of Estimates	0.457	0.455	0.172	0.420	0.421	0.411
Mean SE	0.034	0.063	0.026	0.035	0.036	0.067
Mean Square Error	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

Removing $W_5(t)$ From Initial Model Specification

	Time Dependent			Baseline		
	TMLE	EE	IPCW	TMLE	A-IPCW	IPCW
Mean of Estimates	0.459	0.461	0.173	0.411	0.411	0.396
Mean SE	0.034	0.066	0.026	0.038	0.038	0.065
Mean Square Error	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 of Estimates	0.462	1.243	0.357	0.405	0.405	0.403
Mean SE	0.616	0.619	0.056	0.038	0.038	0.063
Mean Square Error	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 5: Simulation Results For Independent Censoring: Mean of Estimates and Mean Square Error for All Six Estimators

Tshepo Results Incorporating Time Dependent Covariates

Effect of Treatment on Death

- Mean Risk Difference

	TD TMLE	TD DR-EE	TD IPW	BASE TMLE	BASE DR-EE	BASE IPW
Est	0.006	0.005	-0.013	0.004	0.004	0.006
SE	0.012	0.012	0.083	0.012	0.012	0.083
p	0.604	0.661	0.871	0.758	0.756	0.939

- Risk Difference @ 36 Months

	TD TMLE	TD DR-EE	TD IPW	BASE TMLE	BASE DR-EE	BASE IPW
Est	0.005	0.004	0.003	0.002	0.002	0.003
SE	0.017	0.017	0.088	0.017	0.017	0.088
p	0.750	0.821	0.973	0.925	0.924	0.973

Gender Effect Modification on Death

- Mean Risk Difference

	TD TMLE	TD DR-EE	TD IPW	BASE TMLE	BASE DR-EE	BASE IPW
Est	0.039	0.039	0.043	0.033	0.032	0.037
SE	0.017	0.017	0.117	0.017	0.017	0.117
p	0.021	0.019	0.717	0.055	0.058	0.753

- Risk Difference @ 36 Months

	TD TMLE	TD DR-EE	TD IPW	BASE TMLE	BASE DR-EE	BASE IPW
Est	0.063	0.065	0.052	0.051	0.051	0.052
SE	0.023	0.023	0.125	0.024	0.024	0.125
p	0.005	0.004	0.680	0.029	0.030	0.680

Gender Effect Modification on Death, Viral Failure, Drop-out

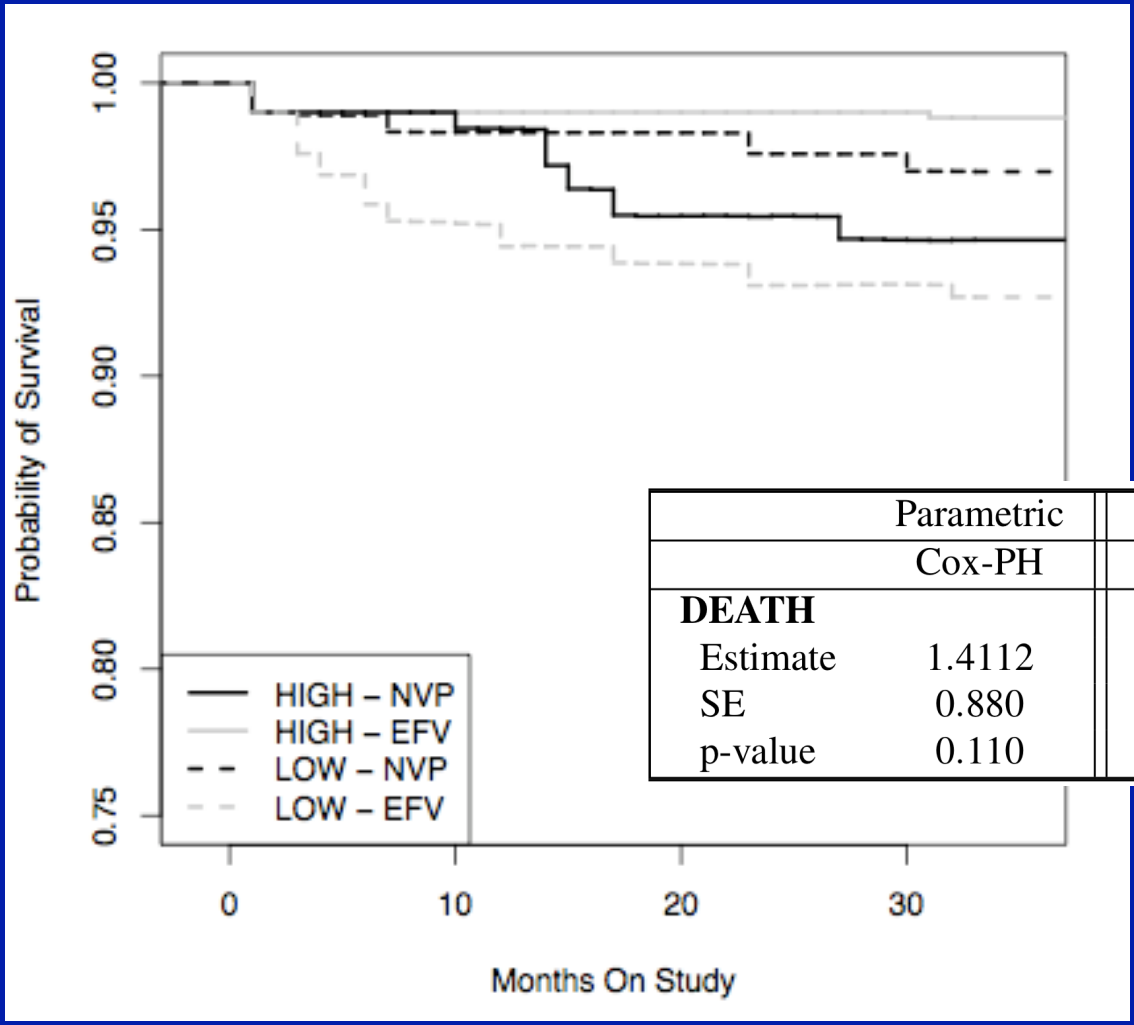
- Mean Risk Difference

	TD TMLE	TD DR-EE	TD IPW	BASE TMLE	BASE DR-EE	BASE IPW
Est	0.132	0.133	0.129	0.126	0.126	0.130
SE	0.039	0.038	0.101	0.038	0.038	0.100
p	0.001	0.001	0.199	0.001	0.001	0.196

- Risk Difference @ 36 Months

	TD TMLE	TD DR-EE	TD IPW	BASE TMLE	BASE DR-EE	BASE IPW
Est	0.200	0.201	0.183	0.189	0.189	0.183
SE	0.050	0.049	0.103	0.049	0.049	0.103
p	0.000	0.000	0.074	0.000	0.000	0.074

Causal Effect Modification By CD4 Level: Death



	Parametric	Targeted Maximum Likelihood		
	Cox-PH	Mean RH	Mean RD	RD at t = 34
DEATH				
Estimate	1.4112	2.182	−0.050	−0.084
SE	0.880	0.826	0.021	0.032
p-value	0.110	0.008	0.017	0.009

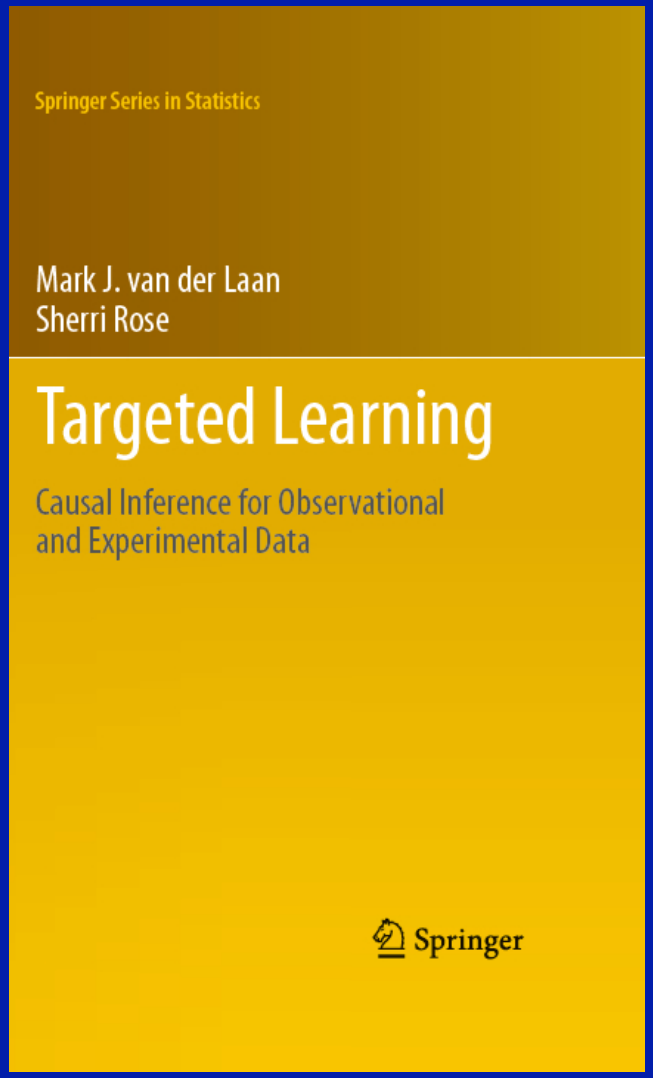
Closing Remarks

- True knowledge is embodied by semi or non-parametric models
- Define target parameter on realistic model
- Semi-parametric models require fully automated state of the art machine learning (super learning)
- Targeted bias removal is essential and is achieved by targeted MLE

Closing Remarks

- Targeted MLE is effective in dealing with sparsity by being substitution estimator, and having relevant criterion for fitting treatment/censoring mechanism (C-TMLE)
- TMLE is double robust and efficient.
- Statistical Inference is now sensible.

Forthcoming book *Targeted Learning* coming June 2011



www.**targetedlearningbook**.com

Acknowledgements

- UC Berkeley
  - Jordan Brooks
  - Paul Chaffee
  - Ivan Diaz Munoz
  - Susan Gruber
  - Alan Hubbard
  - Maya Petersen
  - Kristin Porter
  - Sherri Rose
  - Jas Sekhon
  - Ori Stitelman
  - Cathy Tuglus
  - Wenjing Zheng
- Johns Hopkins
  - Michael Rosenblum
- Stanford
  - Hui Wang
- Paris Descartes
  - Antoine Chambaz
- Kaiser
  - Bruce Fireman
  - Alan Go
  - Romain Neugebauer
- FDA
  - Thamban Valappil
  - Greg Soon
  - Dan Rubin
- Harvard
  - David Bangsberg
  - Victor De Gruttola
- NCI
  - Eric Polley

EXTRA SLIDES

Loss-Based Super Learning in Semi-parametric Models

- Allows one to combine many data-adaptive estimators into one improved estimator.
- Grounded by oracle results for loss-function based cross-validation (vdL&D, 2003). Loss function needs to be bounded.
- Performs asymptotically as well as best (oracle) weighted combination, or achieves parametric rate of convergence.

The Dangers of Favoritism

- Relative Mean Squared Error (compared to main terms least squares regression) based on the validation sample

Method	Study 1	Study 2	Study 3	Study 4
Least Squares	1.00	1.00	1.00	1.00
LARS	0.91	0.95	1.00	0.91
D/S/A	0.22	0.95	1.04	0.43
Ridge	0.96	0.9	1.02	0.98
Random Forest	0.39	0.72	1.18	0.71
MARS	0.02	0.82	0.17	0.61

Super Learning in Prediction

Method	Study 1	Study 2	Study 3	Study 4	Overall
Least Squares	1.00	1.00	1.00	1.00	1.00
LARS	0.91	0.95	1.00	0.91	0.95
D/S/A	0.22	0.95	1.04	0.43	0.71
Ridge	0.96	0.9	1.02	0.98	1.00
Random Forest	0.39	0.72	1.18	0.71	0.91
MARS	0.02	0.82	0.17	0.61	0.38
Super Learner	0.02	0.67	0.16	0.22	0.19

The Library in Super Learning: The Richer the Better

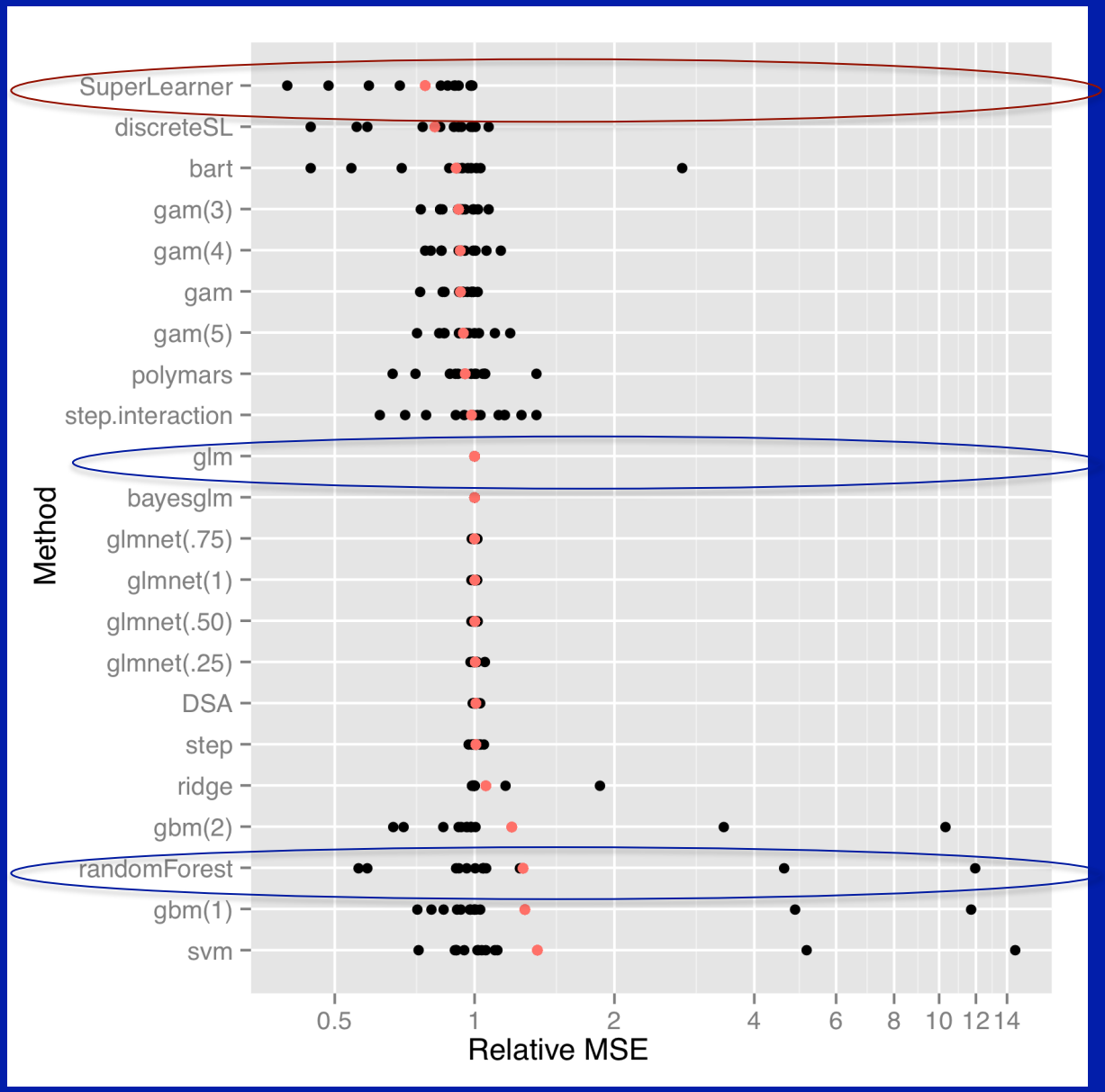
- The key is a vast library of machine learning algorithms to build your estimator
- Currently 40+ R packages for machine learning/prediction
- If we combine dimension-reduction algorithms with these prediction algorithms, we quickly generate a large library

Super Learner: Real Data

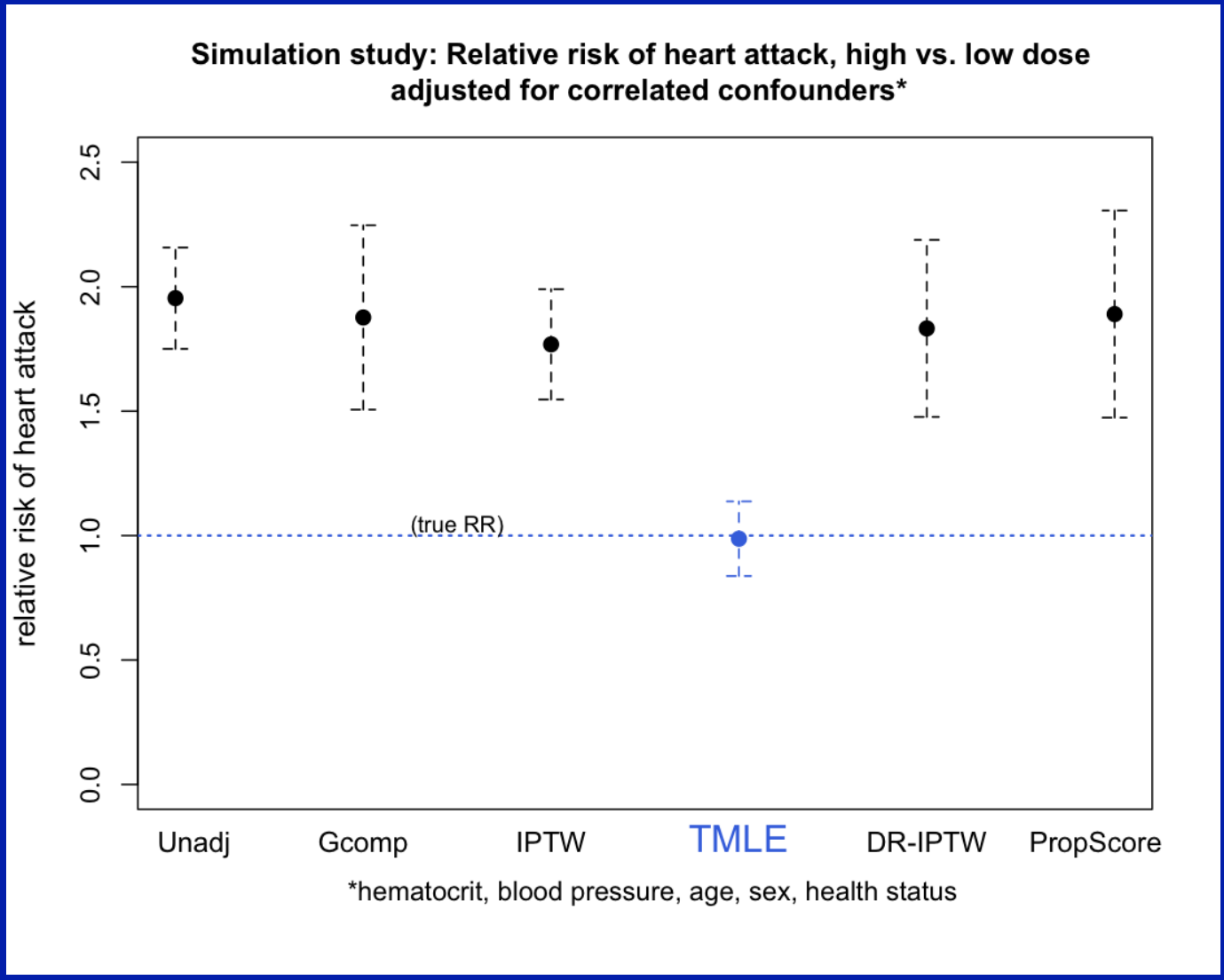
Super Learner- Best weighted combination of algorithms for a given prediction problem

Example algorithm : Linear Main Term Regression

Example algorithm: Random Forest



TMLE/SL: more accurate information from less data



Simulated Safety Analysis of Epogen (Amgen)

Example: Targeted MLE in RCT

Impact of Treatment on Disease

The Gain in Relative Efficiency in RCT is function of Gain in R^2 relative to unadjusted estimator

- We observe (W,A,Y) on each unit
- A is randomized, P(A=1)=0.5
- Suppose the target parameter is additive causal effect EY(1)-Y(0)
- The relative efficiency of the unadjusted estimator and a targeted MLE equals 1 minus the R-square of the regression

  0.5 Q(1,W)+0.5 Q(0,W), where Q(A,W) is the regression of Y on A,W obtained with targeted MLE.

TMLE in Actual Phase IV RCT

- Study: RCT aims to evaluate safety based on mortality due to drug-to-drug interaction among patients with severe disease
- Data obtained with random sampling from original real RCT FDA dataset
- Goal: Estimate risk difference (RD) in survival at 28 days (0/1 outcome) between treated and placebo groups

TMLE in Phase IV RCT

	Unadjusted	TMLE
Estimate	0.034	0.043
p-value (RE)	0.085 (1.000)	0.009 (1.202)

- TMLE adjusts for small amount of empirical confounding (imbalance in AGE covariate)
- TMLE exploits the covariate information to gain in efficiency and thus power over unadjusted
- TMLE Results significant at 0.05

TMLE in RCT: Summary

- TMLE approach handles censoring and improves efficiency over standard approaches
  - Measure strong predictors of outcome
- Implications
  - Unbiased estimates with informative censoring
  - Improved power for clinical trials
  - Smaller sample sizes needed
  - Possible to employ earlier stopping rules
  - Less need for homogeneity in sample
    - More representative sampling
    - Expanded opportunities for subgroup analyses

Targeted Maximum Likelihood Estimation for longitudinal data structures

The Likelihood for Right Censored Survival Data

- It starts with the marginal probability distribution of the baseline covariates.
- Then follows the treatment mechanism.
- Then it follows with a product over time points t
- At each time point t, one writes down likelihood of censoring at time t, death at time t, and it stops at first event
- Counterfactual survival distributions are obtained by intervening on treatment, and censoring.
- This then defines the causal effects of interest as parameter of likelihood.

TMLE with Survival Outcome

- Suppose one observes baseline covariates, treatment, and one observes subject up till end of follow up or death:

$$(W, A, \Delta = I(T \leq C), \tilde{T} = \min(T, C))$$

- One wishes to estimate causal effect of treatment A on survival T
- Targeted MLE uses covariate information to adjust for confounding, informative drop out and to gain efficiency

TMLE with Survival Outcome

- Target $\psi_1(t_0)=\Pr(T_1>t_0)$ and $\psi_0(t_0)=\Pr(T_0>t_0)$ – thereby target treatment effect, e.g.,

  1) Difference: $\Pr(T_1>t_0) - \Pr(T_0>t_0)$, 2) Log RH: $\log \frac{\log \psi_1(t_0)}{\log \psi_0(t_0)}$

- Obtain initial conditional hazard fit (e.g. super learner for discrete survival) and add two time-dependent covariates

$$h_\delta(t, A, W) = \frac{I(A=\delta)}{g(A \mid W)\bar{G}(t \mid A, W)} \frac{S(t_0 \mid A, W)}{S(t \mid A, W)} I(t \le t_0)$$

  - Iterate until convergence, then use updated conditional hazard from final step, and average corresponding conditional survival over W for fixed treatments 0 and 1

TMLE analogue to log rank test

- The parameter,

$$\psi = \frac{1}{\#t_0} \sum_{t_0} \log \left(\frac{\log S_1(t_0)}{\log S_0(t_0)} \right)$$

  corresponds with Cox ph parameter, and thus log rank parameter

- Targeted MLE targeting this parameter is double robust

TMLE in RCT with Survival Outcome

Difference at Fixed End Point

Independent Censoring

	% Bias	Power	95% Coverage	Relative Efficiency
KM	<1%	0.79	0.95	1.00
TMLE	<1%	0.91	0.95	1.44

→ **TMLE: gain in power over KM**

Informative Censoring

	% Bias	Power	95% Coverage	Relative Efficiency
KM	13%	0.88	0.92	1.00
TMLE	<1%	0.92	0.95	1.50

→ **TMLE: unbiased**

TMLE in RCT with survival outcome: Log rank analogue

Independent Censoring

	% Bias	Power	95% Coverage	Relative Efficiency
Log rank	<2%	0.13	0.95	1.00
TMLE (correct λ)	<1%	0.22	0.95	1.48
TMLE (mis-spec λ)	<1%	0.19	0.95	1.24

➢**TMLE: gain in power over log rank**

Informative Censoring

	% Bias	Power	95% Coverage	Relative Efficiency
Log rank	32%	0.20*	0.93	1.00
TMLE (correct λ, correct G)	<1%	0.18	0.95	1.44
TMLE (mis-spec λ, correct G)	<1%	0.15	0.95	1.24

➢**TMLE: unbiased**

Kang and Schafer Simulation

- Continuous Y and 4 baseline covariates W_1, W_2, W_3, W_4.
- The true population mean is 210, while the mean among respondents is 200.
- Covariates predict missingness and outcome
- Positivity violations: $g_0 \in [0.01, 0.98]$ and $g_n \in [4 \times 10^{-6}, 0.97]$.
- The estimators of regressions on Y and Delta are

  either miss-specified or correctly specified, as in KS.

Modifications to Kang and Schafer Simulation

Modification 1

- The true population mean is again 210, but now the mean among respondents is 184.
- More misspecification.
- Stronger Positivity violations.

$$g_0 \in [1.1 \times 10^{-5}, 0.99]$$
$$g_n \in [2.2 \times 10^{-16}, 0.87].$$

Modification 2

- Same as above, except one of the covariates no longer causally affects the outcome

Traditional Approach in Epidemiology

1. Fit several parametric logistic regression models, and select a favorite one.
2. Report point estimate of coefficient in front of treatment, confidence intervals, and *p*-value, as if this parametric model was a priori-specified.

Complications of Human Art in Statistics

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

The New York Times
nytimes.com

September 16, 2007

Do We Really Know What Makes Us Healthy?

By GARY TAUBES

AMSTATNEWS
The Membership Magazine of the American Statistical Association

Statistics Ready for a Revolution

1 SEPTEMBER 2010 503 VIEWS 2 COMMENTS

Next Generation of Statisticians Must Build Tools for Massive Data Sets

Mark van der Laan, Jiann-Ping Hsu/Karl E. Peace Professor in Biostatistics and Statistics at UC Berkeley, and Sherri Rose, PhD candidate at UC Berkeley

Complications of Human Art in Statistics

Debate over HRT

Professional groups gave HRT their stamp of approval 15 years ago.

Studies indicated HRT protective against osteoporosis and heart disease.

In 1998, a study demonstrated increased risk of heart attack among women with heart disease taking HRT.

In 2002 a study showed increased risk for breast cancer, heart disease, and stroke, among other ailments, for women on HRT.

Why were there inconsistencies in the study results?

Complications of Human Art in Statistics

Debate over mammography

Mammography gained widespread acceptance as effective tool for breast cancer screening in the 1980s.

The Health Insurance Plan trial and Swedish Two-County trial demonstrated mammography saved lives.

In 2009, surprise over new recommendations from the U.S. Preventive Services Task Force.

Among women without a family history, mammography now recommended for women aged 50 to 74. Previous guidelines started at age 40.

Why was there a seemingly sudden paradigm shift?

Kaiser Permanente Data Summary

Nested case-control sample (n=27,012) from a Kaiser Permanente database of persons over the age of 65 in 2003.

- **Outcome** Y was **death** the subsequent year (2004).
- **Covariates** $W = \{W_1, \ldots W_{186}\}$ were **184 medical flags** covering a variety of diseases, treatments, and conditions as well as **gender** and **age**.

Weighting

- Since we use a **two-stage design**, we need to account for this in our analysis with **weighting**.
- The weighting method involves simple observation weights $w_i = \Delta_i / P_n(\Delta_i = 1 \mid Y_i)$, where $\Delta_i = 1$ indicates inclusion in the nested case-control sample, to eliminate the bias of the sampling design, where these observation weights are determined by **inverse probability of missingness**.
- Thus cases were given observation weights equal to 1 and controls were each given an observation weight of $1/0.041 = 24$
- We incorporate inverse probability of missingness observation weighting into the super learner algorithm to generate a risk score for mortality in nested case-control data from a large Kaiser Permanente database.

Algorithm	RE	R^2
`SuperLearner`	-	0.113
`glm.1`	1.004	0.109
`glm.2`	1.004	0.109
`glm.3`	1.004	0.109
`glm.4`	1.004	0.109
`glm.5`	1.004	0.109
`glm.6`	1.004	0.109
`glm.7`	1.037	0.080
`glm.8`	1.032	0.084
`glm.9`	1.059	0.060
`bayesglm`	1.132	-0.005
`glmnet`,$\alpha = 0.50$	1.000	0.112
`glmnet`,$\alpha = 1.00$	1.000	0.112
`gam`, degree $= 2$	1.004	0.109
`gam`, degree $= 3$	1.004	0.109
`nnet`, size $= 2$	1.173	-0.041
`nnet`, size $= 4$	1.173	-0.041

Targeted Learning

1. Traditional approaches for prediction and effect estimation are biased
2. **Super Learning** allows researchers to combine multiple algorithms to build a prediction function
3. **Targeted MLE** provides bias reduction for efficient effect estimation of the target parameter

Summary of Simulation Results

- TMLE’s are more robust to violations of the positivity assumption, and outperform the other estimators.
- C-TMLE's perform better than TMLE when not all covariates are causally related to outcome.
- Even the case in which all covariates are causally related to the outcome, C-TMLE's still perform as well as TMLE.

$Q_{L(t,j,l)}$

- Convenient way of factorizing the Q part of the likelihood for the contributions of the binary variables L(t,j,l).
- Let $L(t) = (L(t,j) : j = 1, ..., n(t))$ and $L(t,j) = (L(t,j,l) : l = 1, ..., n(t,j))$
- $Q_{L(t)} = P(L(t) \mid Pa(L(t)))$ may be factorized in the following way:

$$Q_{L(t)} = \prod_{j=1}^{n(t)} Q_{L(t,j)}$$

- Furthermore, $Q_{L(t,j)} = P(L(t,j) \mid Pa(L(t,j)))$ may be factorized as:

$$Q_{L(t,j)} = \prod_{l=1}^{n(t,j)-1} Q_{L(t,j,l)}$$

$$Pa(L(t,j)) = \bar{L}(t-1), L(t,1), ..., L(t,j-1), \bar{A}(t-1))$$

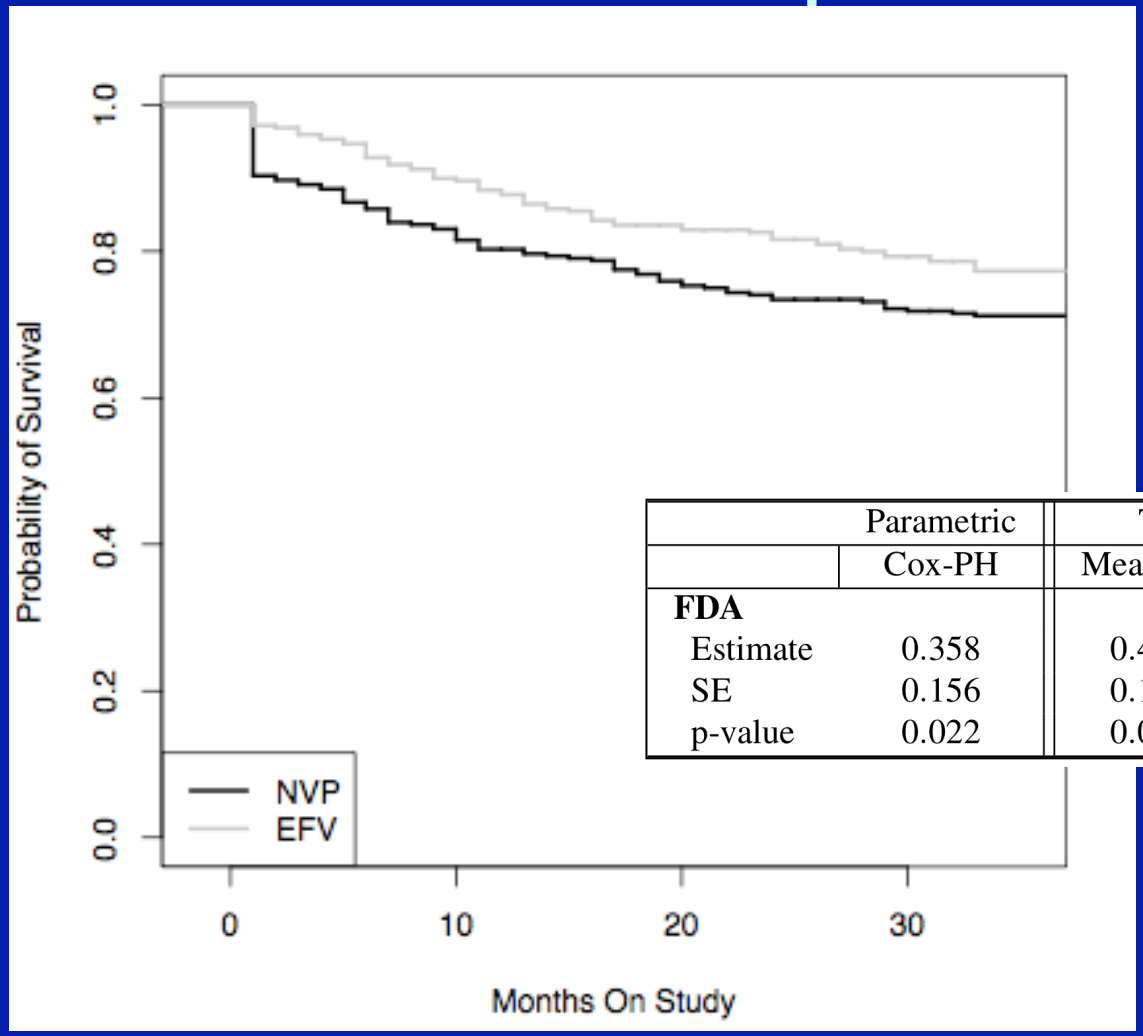
$$Q_{L(t,j,l)} = P(L(t,j,l) \mid Pa(L(t,j,l)))$$

$$Pa(L(t,j,l)) = Pa(L(t,j)), L(t,j,1), \ldots, L(t,j,l-1)$$

- Finally, the entire contribution of Q to the likelihood is:

$$Q = Q_{L(0)} \prod_{t=1}^{K} \prod_{j=1}^{n(t)} \prod_{l=1}^{n(t,j)-1} Q_{L(t,j,l)}$$

Causal Effect of NNRTI: Death, VF, Drop-out



	Parametric	Targeted Maximum Likelihood		
	Cox-PH	Mean RH	Mean RD	RD at t = 34
FDA				
Estimate	0.358	0.451	−0.072	−0.060
SE	0.156	0.165	0.025	0.034
p-value	0.022	0.006	0.003	0.072

Effect of Treatment on Viral Failure or Death

- Mean Risk Difference

	TD TMLE	TD DR-EE	TD IPW	BASE TMLE	BASE DR-EE	BASE IPW
Est	-0.011	-0.009	-0.028	-0.014	-0.014	-0.010
SE	0.021	0.020	0.079	0.019	0.019	0.080
p	0.613	0.651	0.725	0.453	0.454	0.896

- Risk Difference @ 36 Months

	TD TMLE	TD DR-EE	TD IPW	BASE TMLE	BASE DR-EE	BASE IPW
Est	-0.001	0.001	0.001	-0.006	-0.006	0.001
SE	0.033	0.031	0.083	0.029	0.029	0.083
p	0.966	0.964	0.994	0.843	0.843	0.994

Gender Effect Modification on Viral Failure, Death

- Mean Risk Difference

	TD TMLE	TD DR-EE	TD IPW	BASE TMLE	BASE DR-EE	BASE IPW
Est	0.076	0.078	0.073	0.068	0.068	0.074
SE	0.029	0.028	0.112	0.027	0.027	0.112
p	0.010	0.005	0.519	0.012	0.012	0.511

- Risk Difference @ 36 Months

	TD TMLE	TD DR-EE	TD IPW	BASE TMLE	BASE DR-EE	BASE IPW
Est	0.152	0.156	0.139	0.140	0.140	0.139
SE	0.045	0.042	0.118	0.042	0.042	0.118
p	0.001	0.000	0.238	0.001	0.001	0.238

Effect of Treatment on Death, Viral Failure, Drop-out

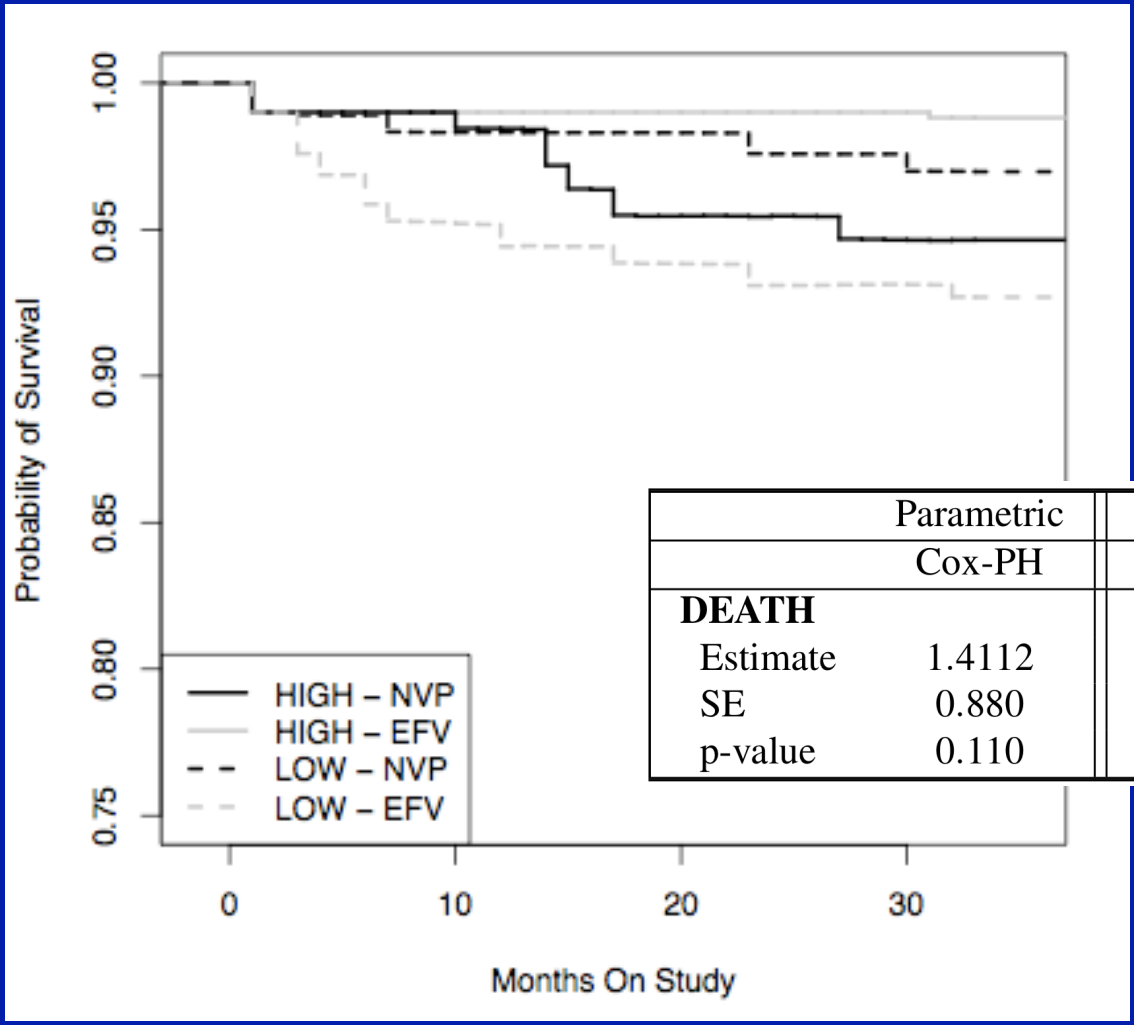
- Mean Risk Difference

	TD TMLE	TD DR-EE	TD IPW	BASE TMLE	BASE DR-EE	BASE IPW
Est	-0.077	-0.076	-0.077	-0.079	-0.079	-0.078
SE	0.028	0.027	0.070	0.027	0.027	0.070
p	0.006	0.006	0.276	0.003	0.003	0.265

- Risk Difference @ 36 Months

	TD TMLE	TD DR-EE	TD IPW	BASE TMLE	BASE DR-EE	BASE IPW
Est	-0.066	-0.065	-0.071	-0.070	-0.070	-0.071
SE	0.036	0.036	0.071	0.035	0.035	0.071
p	0.069	0.068	0.316	0.044	0.043	0.316

Causal Effect Modification By CD4 Level: Death



	Parametric	Targeted Maximum Likelihood		
	Cox-PH	Mean RH	Mean RD	RD at t = 34
DEATH				
Estimate	1.4112	2.182	−0.050	−0.084
SE	0.880	0.826	0.021	0.032
p-value	0.110	0.008	0.017	0.009

The Need for Targeted Learning in Semi-Parametric Models

1. MLE/machine learning are not targeted for effect parameters.
2. For that, we need a subsequent targeted bias-reduction step.

Targeted MLE