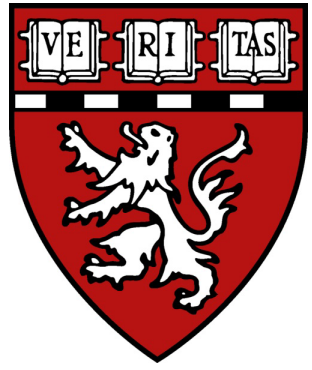
Machine Learning: Day 2

Sherri Rose

Associate Professor
Department of Health Care Policy
Harvard Medical School

drsherrirose.com
@sherrirose

February 28, 2017



Goals: Day 2

1. Understand shortcomings of standard parametric regression-based techniques for the estimation of causal effect quantities.
2. Be introduced to the ideas behind machine learning approaches as tools for confronting the curse of dimensionality.
3. Become familiar with the properties and basic implementation of TMLE for effect estimation.

[Motivation]

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

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

---

4 high impact zones in statistical discovery with big data

FierceBigData

September 22, 2014

SHARE

Email

38 Tweet

23 Share

4 Like

2 +1

Clockwise from top left: Dunson, Rudin, McCormic and Rose

By: Sherri Rose, Harvard University; David Dunson, Duke University; Tyler McCormick, University of Washington; and Cynthia Rudin, MIT

Big data is transforming society with the help of statisticians, who possess in-depth experience and expertise in the art and science surrounding data. The American Statistical Association, or ASA, has recently released a white paper entitled "Discovery with Data: Leveraging Statistics with Computer Science to Transform Science and Society" (pdf) that highlights high-impact areas where statistical science is being applied to transformative big data research questions. Statistics is, by definition, the science of learning from data, and has had a key impact in several of the most prominent fields of discovery, including the biological sciences, health care, business analytics and recommendation systems, and the social sciences. There is a strong need to work in integrated teams comprised of domain experts, statisticians, and computer scientists in order to solve these complicated problems, which require tailored solutions using the influx of big data.

---

variations

Big data and the future

At the beginning of her career **Sherri Rose** discusses big data and stands amazed at its potential.

STATtr@k *A website for new statistics professionals navigating a data-centric*

Home | About Us | ASA Membership | Get Involved | Awards & Scholarships | Career

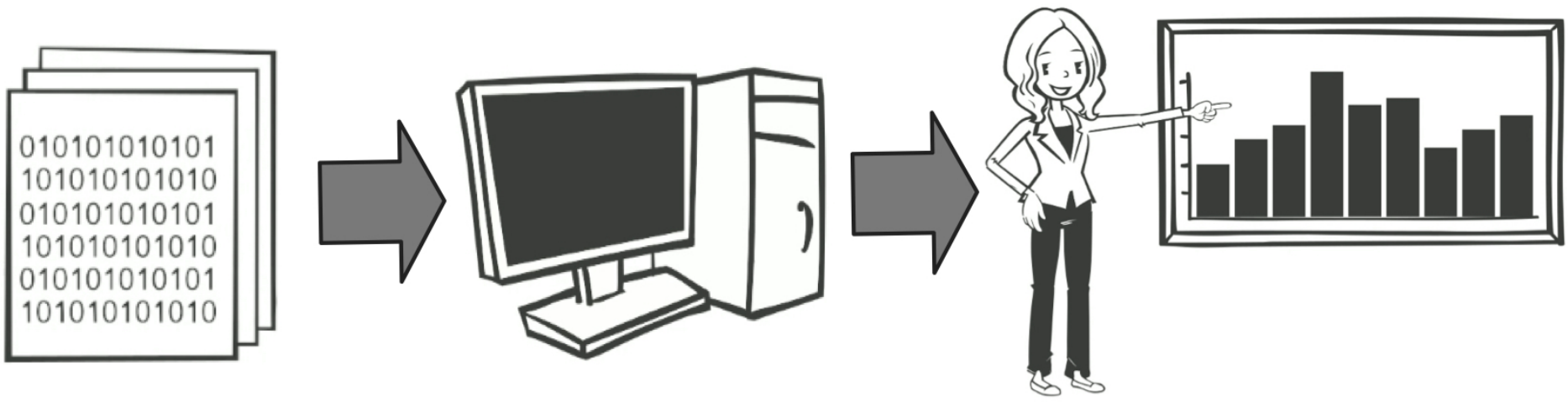
Statisticians' Place in Big Data

– FEBRUARY 1, 2013
POSTED IN: DEVELOPMENT TR@K

Sherri Rose is an NSF mathematical sciences postdoctoral research fellow in the department of biostatistics at the Johns Hopkins Bloomberg School of Public Health.

Big Data has become the new buzz phrase in the world of information collection and analysis. The experiments we conduct and the observational data we collect continue to grow in size, due to rapidly expanding technology.

Large data sets also have drawn the attention of young people, with undergraduate and graduate students choosing computer science, engineering, and statistics for their programs of study. Each of these disciplines brings something unique to the table when discussing the challenges of Big Data, and interdisciplinary collaborations are becoming increasingly common.


010101010101
101010101010
010101010101
101010101010
010101010101
101010101010

Electronic Health Databases

The increasing availability of electronic medical records offers a **new resource to public health researchers**.

General usefulness of this type of data to answer targeted scientific research questions is an open question.

Need **novel statistical methods** that have desirable statistical properties while remaining computationally feasible.

Yesterday Super Learner: Kaiser Permanente Database

Nested case-control sample (n=27,012).

- **Outcome**: death.
- **Covariates**: 184 medical flags, gender & age.

Ensembling method outperformed all other algorithms.

Generally weak signal with $R^2 = 0.11$.

Observed data structure on a subject can be represented as $O = (Y, \Delta, \Delta X)$, where $X = (W, Y)$ is the full data structure, and Δ denotes the indicator of inclusion in the second-stage sample.

How will this electronic database perform in comparison to a cohort study?

van der Laan & Rose (2011)

Yesterday Super Learner: Sonoma Cohort Study

Cohort study of $n = 2,066$ residents of Sonoma, CA aged 54 and over.

- Outcome: death.
- Covariates: gender, age, **self-rated health**, **leisure-time physical activity**, smoking status, cardiac event history, and chronic health condition status.
- $R^2 = 0.201$

Two-fold improvement with less than 10% of the subjects & less than 10% the number of covariates.

What possible conclusions can we draw?

Rose (2013)

High Dimensional 'Big Data' Parametric Regression

- Often dozens, hundreds, or even thousands of potential variables

1513	4.163950	3.854244	3.827496
1555	4.277033	3.373982	489.825226
1597	4.390150	3.795142	221.608444
1639	4.503117	3.640379	26.986557
1681	4.616217	3.336954	104.501778
1723	4.729317	3.561723	8.354190
1765	4.842267	3.576960	146.476227
1807	4.955350	3.858309	58.118893
1849	5.068450	3.514176	3.682388
1891	5.181567	3.794615	32.864357
1933	5.294517	3.311670	1.653655
1975	5.407600	3.931615	72.284065
2017	5.520700	4.319901	15.170299
2059	5.633650	3.938955	2.626603
2101	5.746750	3.924497	16.581503
2143	5.859883	3.771340	33.761124
2185	5.972850	3.797512	9.262811
2227	6.085967	3.795501	126.762199
2269	6.199067	3.759673	108.416565
2311	6.312167	3.373145	10.712665
2353	6.425117	3.464702	56.385990
2395	6.538183	3.640879	30.747551
2437	6.651333	3.702649	5.748046
2479	6.764283	3.941036	58.997993
2521	6.877350	3.393778	24.935211
2563	6.990450	3.213435	6.881421
2605	7.103400	3.635089	12.697396
2647	7.216517	3.749416	4.405899
2689	7.329650	3.450428	6.340690
2731	7.442750	3.287580	231.588028

High Dimensional 'Big Data' Parametric Regression

- Often dozens, hundreds, or even thousands of potential variables
- Impossible challenge to correctly specify the parametric regression

1513	4.163950	3.854244	3.827496
1555	4.277033	3.373982	489.825226
1597	4.390150	3.795142	221.608444
1639	4.503117	3.640379	26.986557
1681	4.616217	3.336954	104.501778
1723	4.729317	3.561723	8.354190
1765	4.842267	3.576960	146.476227
1807	4.955350	3.858309	58.118893
1849	5.068450	3.514176	3.682388
1891	5.181567	3.794615	32.864357
1933	5.294517	3.311670	1.653655
1975	5.407600	3.931615	72.284065
2017	5.520700	4.319901	15.170299
2059	5.633650	3.938955	2.626603
2101	5.746750	3.924497	16.581503
2143	5.859883	3.771340	33.761124
2185	5.972850	3.797512	9.262811
2227	6.085967	3.795501	126.762199
2269	6.199067	3.759673	108.416565
2311	6.312167	3.373145	10.712665
2353	6.425117	3.464702	56.385990
2395	6.538183	3.640879	30.747551
2437	6.651333	3.702649	5.748046
2479	6.764283	3.941036	58.997993
2521	6.877350	3.393778	24.935211
2563	6.990450	3.213435	6.881421
2605	7.103400	3.635089	12.697396
2647	7.216517	3.749416	4.405899
2689	7.329650	3.450428	6.340690
2731	7.442750	3.287580	231.588028

High Dimensional 'Big Data' Parametric Regression

- Often dozens, hundreds, or even thousands of potential variables
- Impossible challenge to correctly specify the parametric regression
- May have more unknown parameters than observations

1513	4.163950	3.854244	3.827496
1555	4.277033	3.373982	489.825226
1597	4.390150	3.795142	221.608444
1639	4.503117	3.640379	26.986557
1681	4.616217	3.336954	104.501778
1723	4.729317	3.561723	8.354190
1765	4.842267	3.576960	146.476227
1807	4.955350	3.858309	58.118893
1849	5.068450	3.514176	3.682388
1891	5.181567	3.794615	32.864357
1933	5.294517	3.311670	1.653655
1975	5.407600	3.931615	72.284065
2017	5.520700	4.319901	15.170299
2059	5.633650	3.938955	2.626603
2101	5.746750	3.924497	16.581503
2143	5.859883	3.771340	33.761124
2185	5.972850	3.797512	9.262811
2227	6.085967	3.795501	126.762199
2269	6.199067	3.759673	108.416565
2311	6.312167	3.373145	10.712665
2353	6.425117	3.464702	56.385990
2395	6.538183	3.640879	30.747551
2437	6.651333	3.702649	5.748046
2479	6.764283	3.941036	58.997993
2521	6.877350	3.393778	24.935211
2563	6.990450	3.213435	6.881421
2605	7.103400	3.635089	12.697396
2647	7.216517	3.749416	4.405899
2689	7.329650	3.450428	6.340690
2731	7.442750	3.287580	231.588028

High Dimensional 'Big Data' Parametric Regression

- Often dozens, hundreds, or even thousands of potential variables
- Impossible challenge to correctly specify the parametric regression
- May have more unknown parameters than observations
- True functional might be described by a complex function not easily approximated by main terms or interaction terms

1513	4.163950	3.854244	3.827496
1555	4.277033	3.373982	489.825226
1597	4.390150	3.795142	221.608444
1639	4.503117	3.640379	26.986557
1681	4.616217	3.336954	104.501778
1723	4.729317	3.561723	8.354190
1765	4.842267	3.576960	146.476227
1807	4.955350	3.858309	58.118893
1849	5.068450	3.514176	3.682388
1891	5.181567	3.794615	32.864357
1933	5.294517	3.311670	1.653655
1975	5.407600	3.931615	72.284065
2017	5.520700	4.319901	15.170299
2059	5.633650	3.938955	2.626603
2101	5.746750	3.924497	16.581503
2143	5.859883	3.771340	33.761124
2185	5.972850	3.797512	9.262811
2227	6.085967	3.795501	126.762199
2269	6.199067	3.759673	108.416565
2311	6.312167	3.373145	10.712665
2353	6.425117	3.464702	56.385990
2395	6.538183	3.640879	30.747551
2437	6.651333	3.702649	5.748046
2479	6.764283	3.941036	58.997993
2521	6.877350	3.393778	24.935211
2563	6.990450	3.213435	6.881421
2605	7.103400	3.635089	12.697396
2647	7.216517	3.749416	4.405899
2689	7.329650	3.450428	6.340690
2731	7.442750	3.287580	231.588028

Complications of Human Art in 'Big Data' Statistics

1. Fit several parametric models; select a favorite one
2. The parametric model is misspecified
3. The target parameter is interpreted as if the parametric model is correct
4. The parametric model is often data-adaptively (or worse!) built, and this part of the estimation procedure is not accounted for in the variance

Estimation is a Science

1. **Data**: realizations of random variables with a probability distribution.
2. **Statistical Model**: actual knowledge about the shape of the data-generating probability distribution.
3. **Statistical Target Parameter**: a feature/function 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.

Roadmap for Effect Estimation

How does one translate the results from studies, how do we take the information in the data, and draw effective conclusions?

- Define the Research Question
  - Specify Data
  - Specify Model
  - Specify the Parameter of Interest
- Estimate the Target Parameter
- Inference
  - Standard Errors / CIs
  - Interpretation

Data

Random variable O, observed n times, could be defined in a simple case as $O = (W, A, Y) \sim P_0$ if we are without common issues such as missingness and censoring.

- W: vector of covariates
- A: exposure or treatment
- Y: outcome

This data structure makes for effective examples, but data structures found in practice are frequently more complicated.

Data: Censoring & Missingness

Define $O = (W, A, \tilde{T}, \Delta) \sim P_0$.

- T: time to event Y
- C: censoring time
- $\tilde{T} = \min(T, C)$: represents the T or C that was observed first
- $\Delta = I(T \leq \tilde{T}) = I(C \geq T)$: indicator that T was observed at or before C

Define $O = (W, A, \Delta, \Delta Y) \sim P_0$.

- Δ: Indicator of missingness

Model

General case: Observe n i.i.d. copies of random variable O with probability distribution P_0.

The data-generating distribution P_0 is also known to be an element of a statistical model $\mathcal{M}$: $P_0 \in \mathcal{M}$.

A **statistical model** $\mathcal{M}$ is the set of possible probability distributions for P_0; it is a collection of probability distributions.

If all we know is that we have n i.i.d. copies of O, this can be our statistical model, which we call a nonparametric statistical model

Model

A statistical model can be augmented with additional (nontestable causal) assumptions, allowing one to enrich the interpretation of $\Psi(P_0)$.

This does not change the statistical model.

Target Parameters

Define the parameter of the probability distribution P as function of $P : \Psi(P)$.

$$\begin{aligned} \psi_{RD} = \Psi_{RD}(P) &= E_W[E(Y \mid A = 1, W) - E(Y \mid A = 0, W)] \\ &= E(Y_1) - E(Y_0) \\ &= P(Y_1 = 1) - P(Y_0 = 1) \end{aligned}$$

$$\psi_{RR} = \frac{P(Y_1 = 1)}{P(Y_0 = 1)}$$

and

$$\psi_{OR} = \frac{P(Y_1 = 1)P(Y_0 = 0)}{P(Y_1 = 0)P(Y_0 = 1)}.$$

Y is the outcome, A the exposure, and W baseline covariates.

Effect Estimation vs. Prediction

Both **effect** and **prediction** research questions are inherently *estimation* questions, but they are distinct in their goals.

Effect Estimation vs. Prediction

Both **effect** and **prediction** research questions are inherently *estimation* questions, but they are distinct in their goals.

Effect: Interested in estimating the effect of exposure on outcome adjusted for covariates.

Effect Estimation vs. Prediction

Both **effect** and **prediction** research questions are inherently *estimation* questions, but they are distinct in their goals.

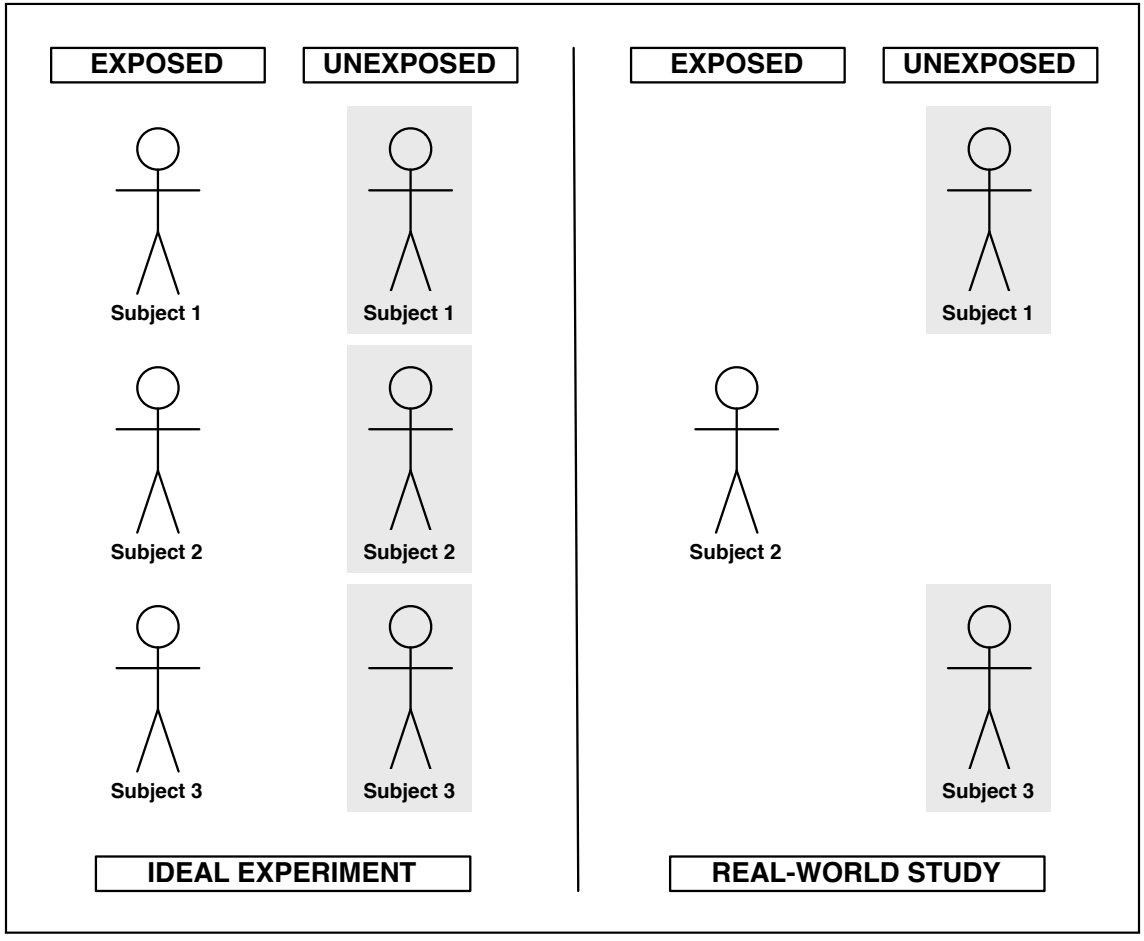
Effect: Interested in estimating the effect of exposure on outcome adjusted for covariates.

Prediction: Interested in generating a function to input covariates and predict a value for the outcome.

[(Causal) Effect Estimation]

Learning from Data

Just what type of studies are we conducting? The often quoted "ideal experiment" is one that cannot be conducted in real life.



Causal Model

Assume a structural causal model (SCM) (Pearl 2009), comprised of endogenous variables $X = (X_j : j)$ and exogenous variables $U = (U_{X_j} : j)$.

- Each X_j is a deterministic function of other endogenous variables and an exogenous error U_j.
- The errors U are never observed.
- For each X_j we characterize its parents from among X with $Pa(X_j)$.

Causal Model

$$X_j = f_{X_j}(Pa(X_j), U_{X_j}),\ j = 1 \ldots, J,$$

The functional form of f_{X_j} is often unspecified.

An SCM can be fully parametric, but we do not do that here as our background knowledge does not support the assumptions involved.

Causal Model

We could specify the following SCM:

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

Recall that we assume for the full data:

1. for each X_j, $X_j = f_j(Pa(X_j), U_{X_j})$ depends on the other endogenous variables only through the parents $Pa(X_j)$,
2. the exogenous variables have a particular joint distribution P_U; $U_A \perp U_Y \mid W$.

In our simple study, $X = (W, A, Y)$, and $Pa(A) = W$. We know this due to the time ordering of the variables.

Causal Graph

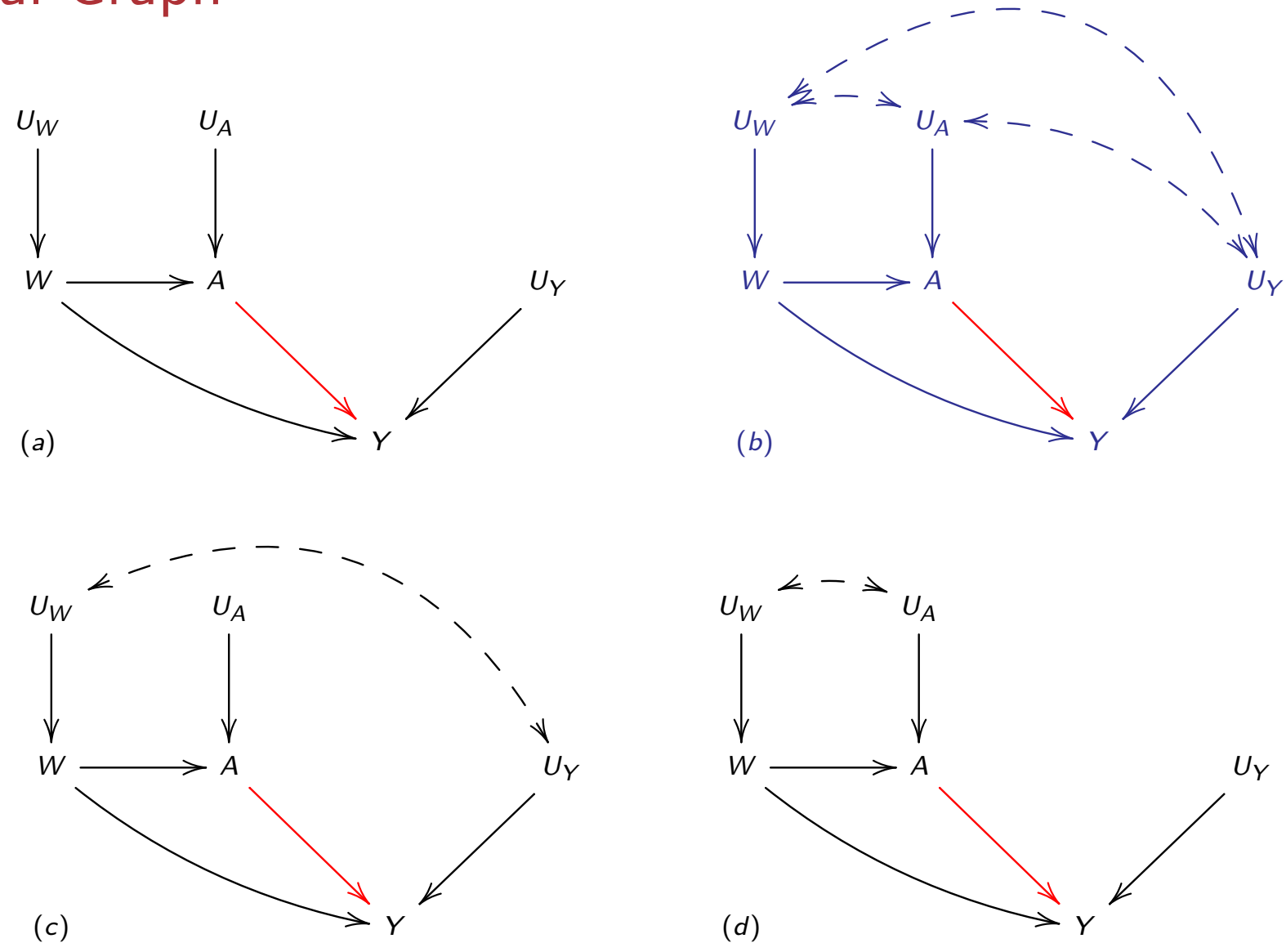


Figure: Causal graphs with various assumptions about the distribution of P_U

A Note on Causal Assumptions

We could alternatively use the Neyman–Rubin Causal Model and assume

- randomization ($A \perp Y_a \mid W$) and
- stable unit treatment value assumption (SUTVA; no interference between subjects and consistency assumption).

Positivity Assumption

We need that each possible exposure level occurs with some positive probability within each stratum of W.

For our data structure (W, A, Y) we are assuming:

$$P_0(A = 1 \mid W = w) > 0 \text{ and } P_0(A = 0 \mid W = w) > 0,$$

for each possible w.

Landscape: Effect Estimators

An **estimator** is an algorithm that can be applied to any empirical distribution to provide a mapping from the empirical distribution to the parameter space.

- **Maximum-Likelihood-Based Estimators**
- **Estimating-Equation-Based Methods**

The target parameters we discussed depend on P_0 through the conditional mean $\bar{Q}_0(A, W) = E_0(Y \mid A, W)$, and the marginal distribution $Q_{W,0}$ of W. Thus we can also write $\Psi(Q_0)$, where $Q_0 = (\bar{Q}_0, Q_{W,0})$.

Landscape: Effect Estimators

- **Maximum-Likelihood-Based Estimators** will be of the type

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

  where this estimate is obtained by plugging in $Q_n = (\bar{Q}_n, Q_{W,n})$ into the mapping Ψ. $\bar{Q}_n(A = a, W_i) = E_n(Y \mid A = a, W_i)$.

- **Estimating-Equation-Based Methods** An estimating function is a function of the data O and the parameter of interest. If $D(\psi)(O)$ is an estimating function, then we can define a corresponding estimating equation: $0 = \sum_{i=1}^{n} D(\psi)(O_i)$, and solution ψ_n satisfying $\sum_{i=1}^{n} D(\psi_n)(O_i) = 0$.

Maximum-Likelihood-Based Methods

MLE using regression. Outcome regression estimated with parametric methods and plugged into

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

Maximum-Likelihood-Based Methods

MLE using regression. Outcome regression estimated with parametric methods and plugged into

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

STOP! When does this differ from traditional regression?

Maximum-Likelihood-Based Methods

MLE using regression: Continuous outcome example.
True effect is -0.35
W_1 = gender
W_2 = medication use
A = high ozone exposure
Y = continuous measure of lung function

Model 1: $E(Y \mid A) = \alpha_0 + \alpha_1 A$
Both Effects: -0.23

Model 2: $E(Y \mid A, W) = \alpha_0 + \alpha_1 A + \alpha_2 W_1 + \alpha_3 W_2$
Both Effects: -0.36

Model 3: $E(Y \mid A, W) = \alpha_0 + \alpha_1 A + \alpha_2 W_1 + \alpha_3 A \cdot W_2$
Regression Effect: -0.49
MLE Effect: -0.34

Maximum-Likelihood-Based Methods

MLE using regression: Binary outcomes.

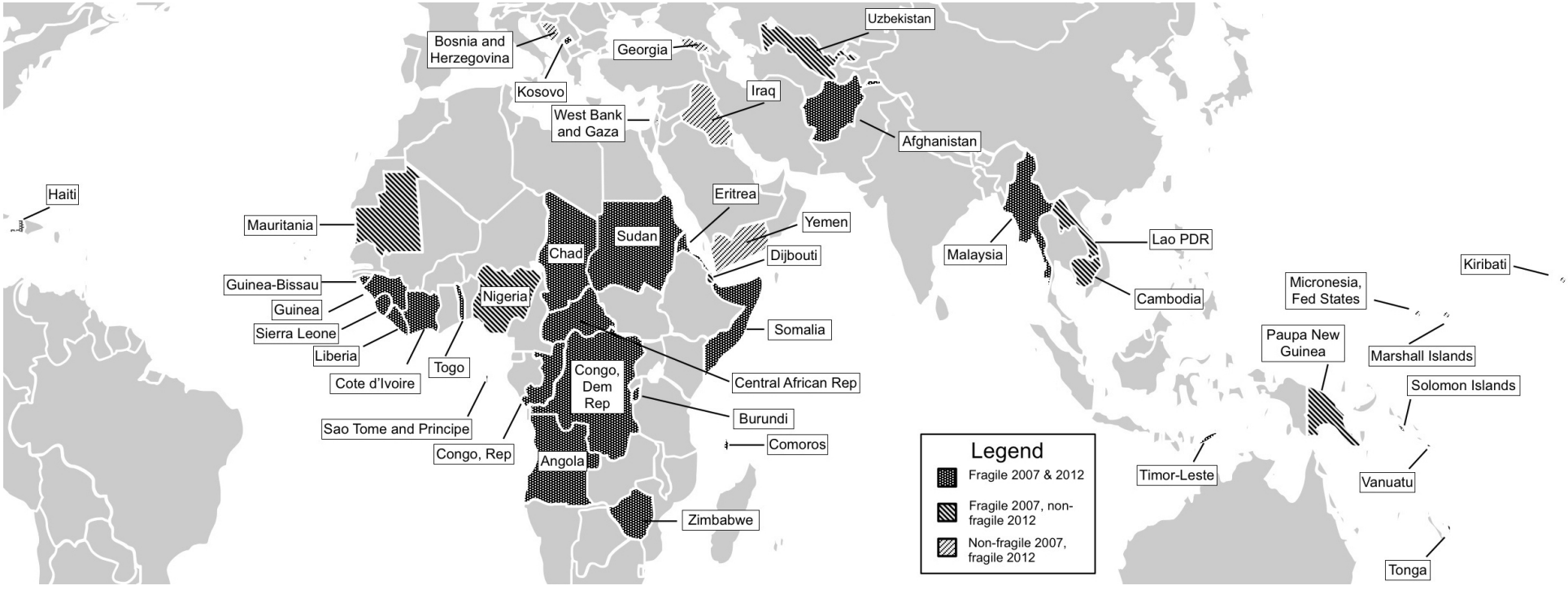
$$P(Y = 1 \mid A, W) = \frac{1}{1 + e^{-\beta_0 + \beta_1 A + \beta_2 W}}$$

$$EY_a = P(Y_a = 1) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{1 + e^{-\beta_0 + \beta_1 A_i + \beta_2 W_i}}$$

$$\frac{EY_1/(1 - EY_1)}{EY_0/(1 - EY_0)} \neq e^{\beta_1}$$

Medical Schools in Fragile States: Delivery of Care

We found that fragile states lack the infrastructure to train sufficient numbers of medical professionals to meet their population health needs.



Fragile states were 1.76 (95%CI 1.07-2.45) to 2.37 (95%CI 1.44-3.30) times more likely to have < 2 medical schools than non-fragile states.

Mateen, McKenzie, Rose (2017)

Maximum-Likelihood-Based Methods

MLE using machine learning. Outcome regression estimated with *machine learning* and plugged into

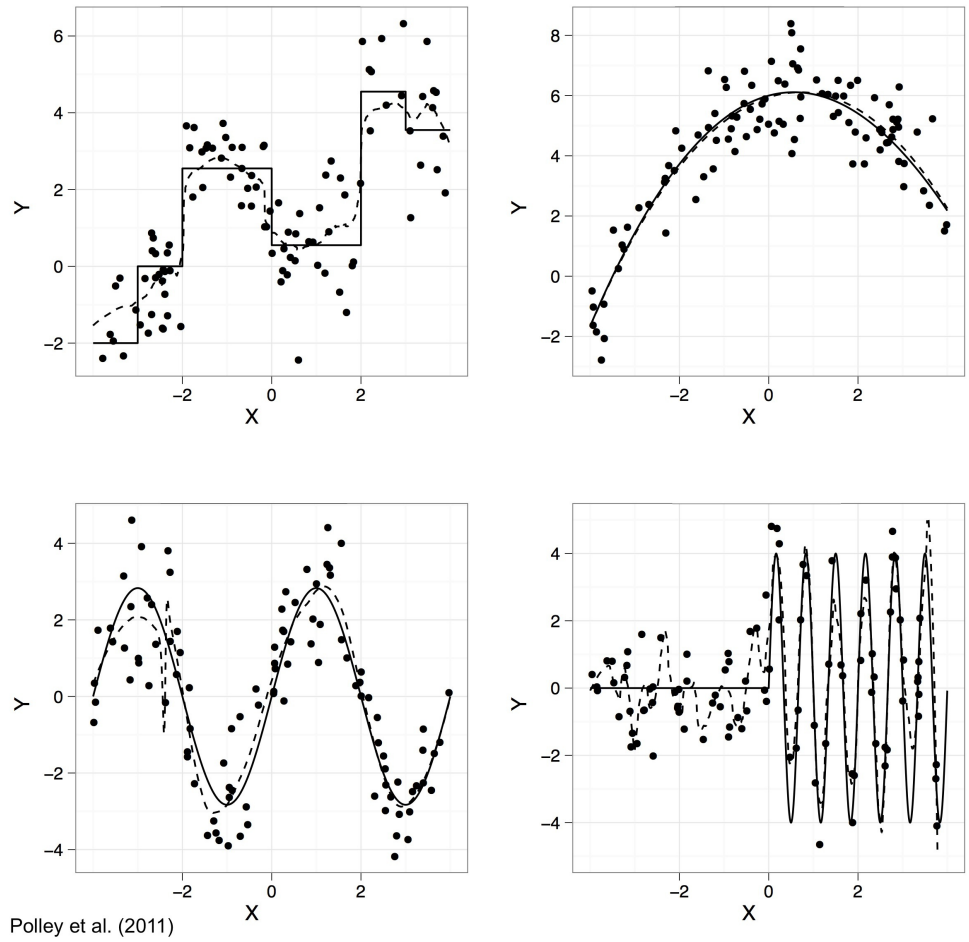
$$\psi_n = \frac{1}{n} \sum_{i=1}^{n} \{\bar{Q}_n(1, W_i) - \bar{Q}_n(0, W_i)\}.$$

Machine Learning Estimation of $\bar{Q}(A, W) = E(Y \mid A, W)$

Machine Learning Big Picture

Machine learning aims to

- "smooth" over the data
- make fewer assumptions

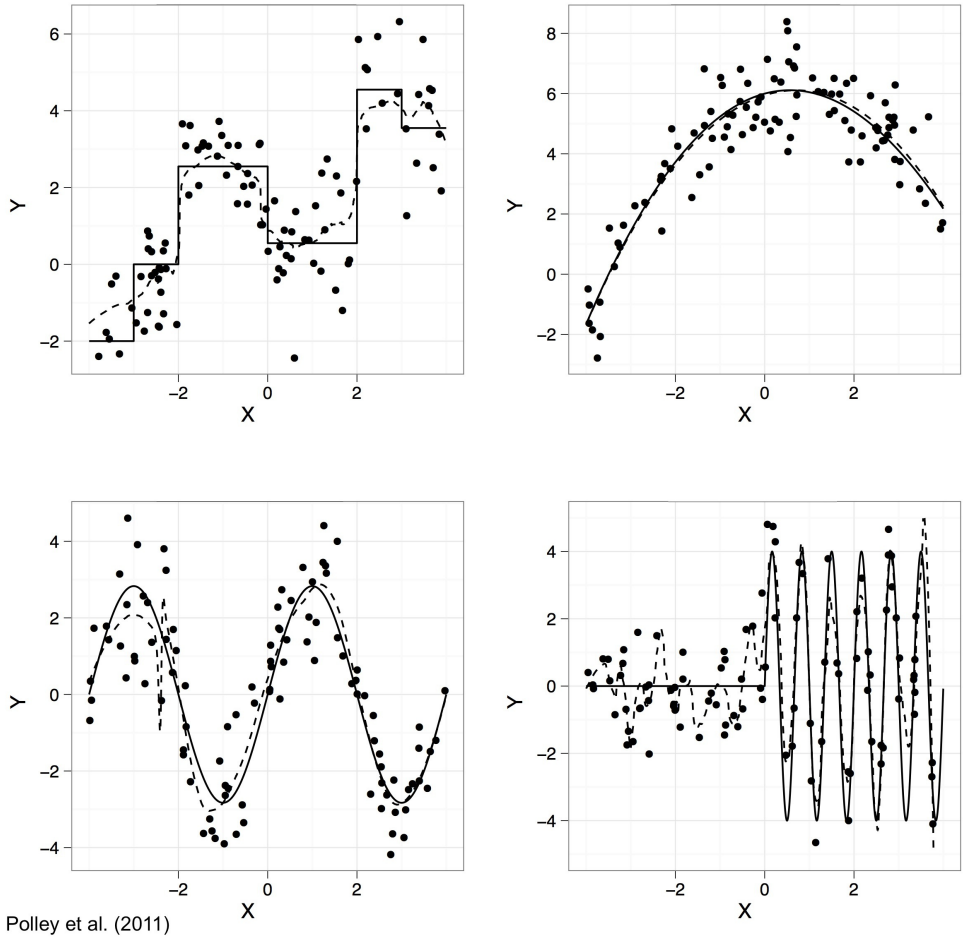


Polley et al. (2011)

Machine Learning Big Picture

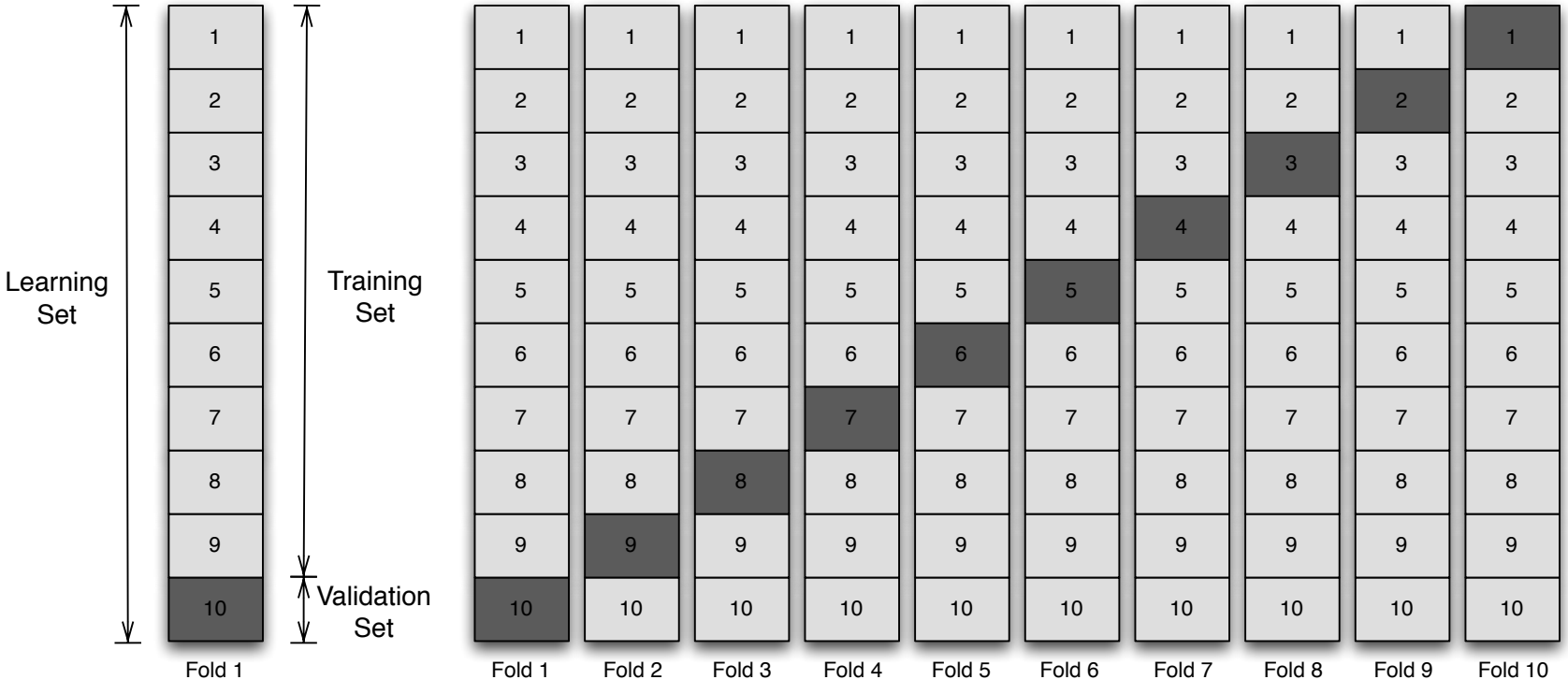
Purely nonparametric model with high dimensional data?

- $p > n$!
- data sparsity



Machine Learning Big Picture: Ensembling

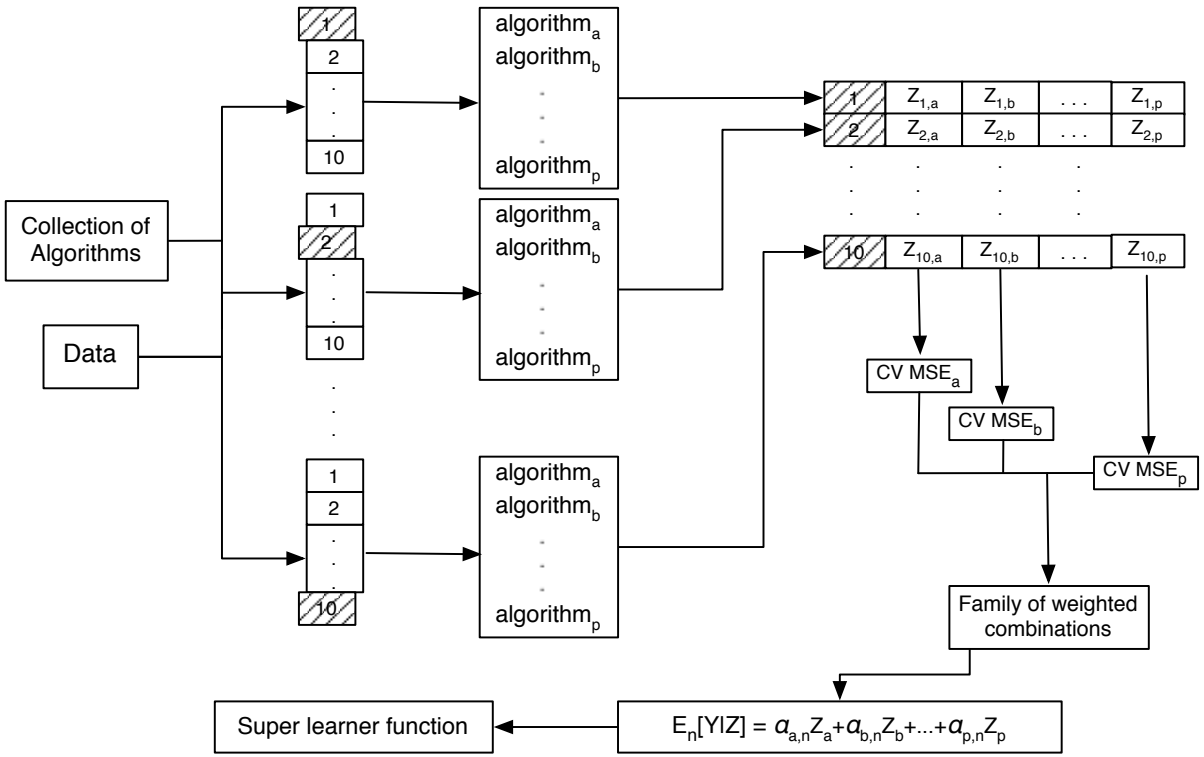
- Ensembling methods allow implementation of multiple algorithms.
- Do not need to decide beforehand which single technique to use; can use several by incorporating **cross-validation**.



Machine Learning Big Picture: Ensembling

Build a collection of algorithms consisting of all weighted averages of the algorithms.

One of these weighted averages might perform better than one of the algorithms alone.



Noncommunicable Disease and Poverty

Studied relative risk of death from noncommunicable disease on three poverty measures in Matlab, Bangladesh.

Parametric regression standardization	
Asset quintile	1.19
Self-rated condition	1.16
Landholding	1.14
Machine-learning (super learner) estimation	
Asset quintile	1.15
Self-rated condition	1.13
Landholding	1.11

Implemented parametric and machine learning substitution estimators.

Mirelman et al. (2016)

Estimating Equation Methods

IPW. Estimate causal risk difference with

$$\psi_n = \frac{1}{n}\sum_{i=1}^{n}\{I(A_i = 1) - I(A_i = 0)\}\frac{Y_i}{g_n(A_i, W_i)}.$$

This estimator is a solution of an IPW estimating equation that relies on an estimate of the treatment mechanism, playing the role of a nuisance parameter of the IPW estimating function.

A-IPW. One estimates $\Psi(P_0)$ with

$$\begin{aligned}\psi_n &= \frac{1}{n}\sum_{i=1}^{n}\frac{\{I(A_i = 1) - I(A_i = 0)\}}{g_n(A_i, W_i)}(Y_i - \bar{Q}_n(A_i, W_i)) \\ &+ \frac{1}{n}\sum_{i=1}^{n}\{\bar{Q}_n(1, W_i) - \bar{Q}_n(0, W_i)\}.\end{aligned}$$

Targeted Learning in Nonparametric Models

- Parametric MLE not targeted for effect parameters
- Need a subsequent targeted bias-reduction step

Targeted Learning

- Avoid reliance on human art and unrealistic parametric models
- Define interesting parameters
- Target the parameter of interest
- Incorporate machine learning
- Statistical inference

TMLE for Causal Effects

TMLE

Produces a well-defined, unbiased, efficient substitution estimator of target parameters of a data-generating distribution.

It is an iterative procedure that updates an initial (super learner) estimate of the relevant part Q_0 of the data generating distribution P_0, possibly using an estimate of a nuisance parameter g_0.

TMLE for Causal Effects

Super Learner

Allows researchers to use multiple algorithms to outperform a single algorithm in nonparametric statistical models.

Builds weighted combination of estimators where weights are optimized based on loss-function specific cross-validation to guarantee best overall fit.

Targeted Maximum Likelihood Estimation

With an initial estimate of the outcome regression, the second stage of TMLE updates this initial fit in a step targeted toward making an optimal bias-variance tradeoff for the parameter of interest.

TMLE for Causal Effects

TMLE: Double Robust

- Removes asymptotic residual bias of initial estimator for the target parameter, if it uses a consistent estimator of censoring/treatment mechanism g_0.
- If initial estimator was consistent for the target parameter, the additional fitting of the data in the targeting step may remove finite sample bias, and preserves consistency property of the initial estimator.

TMLE: Efficiency

- If the initial estimator and the estimator of g_0 are both consistent, then it is also asymptotically efficient according to semi-parametric statistical model efficiency theory.

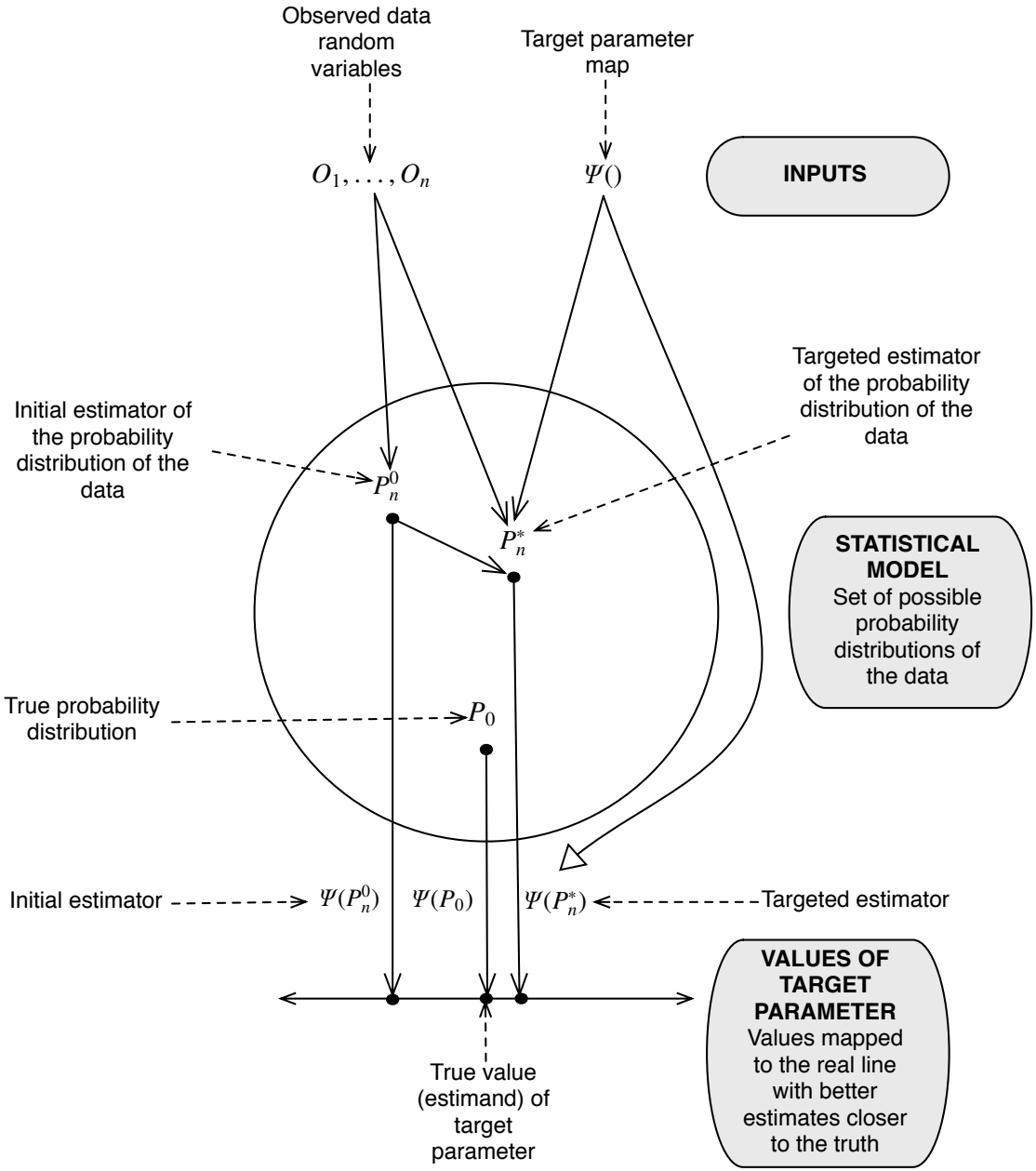
TMLE for Causal Effects

TMLE: In Practice

Allows the incorporation of machine learning methods for the estimation of both Q_0 and g_0 so that we do not make assumptions about the probability distribution P_0 we do not believe.

Thus, every effort is made to achieve minimal bias and the asymptotic semi-parametric efficiency bound for the variance.

Targeted Learning in Nonparametric Models



Example: TMLE for the Risk Difference

Note that ϵ_n is obtained by performing a regression of Y on $H_n^*(A, W)$, where $\bar{Q}_n^0(A, W)$ is used as an offset, and extracting the coefficient for $H_n^*(A, W)$.

We then update $\bar{Q}_n^0$ with $\text{logit}\bar{Q}_n^1(A, W) = \text{logit}\bar{Q}_n^0(A, W) + \epsilon_n^1 H_n^*(A, W)$. This updating process converges in one step in this example, so that the TMLE is given by $Q_n^* = Q_n^1$.

Example: Sonoma Cohort Study

Cohort study of $n = 2,066$ residents of Sonoma, CA aged 54 and over.

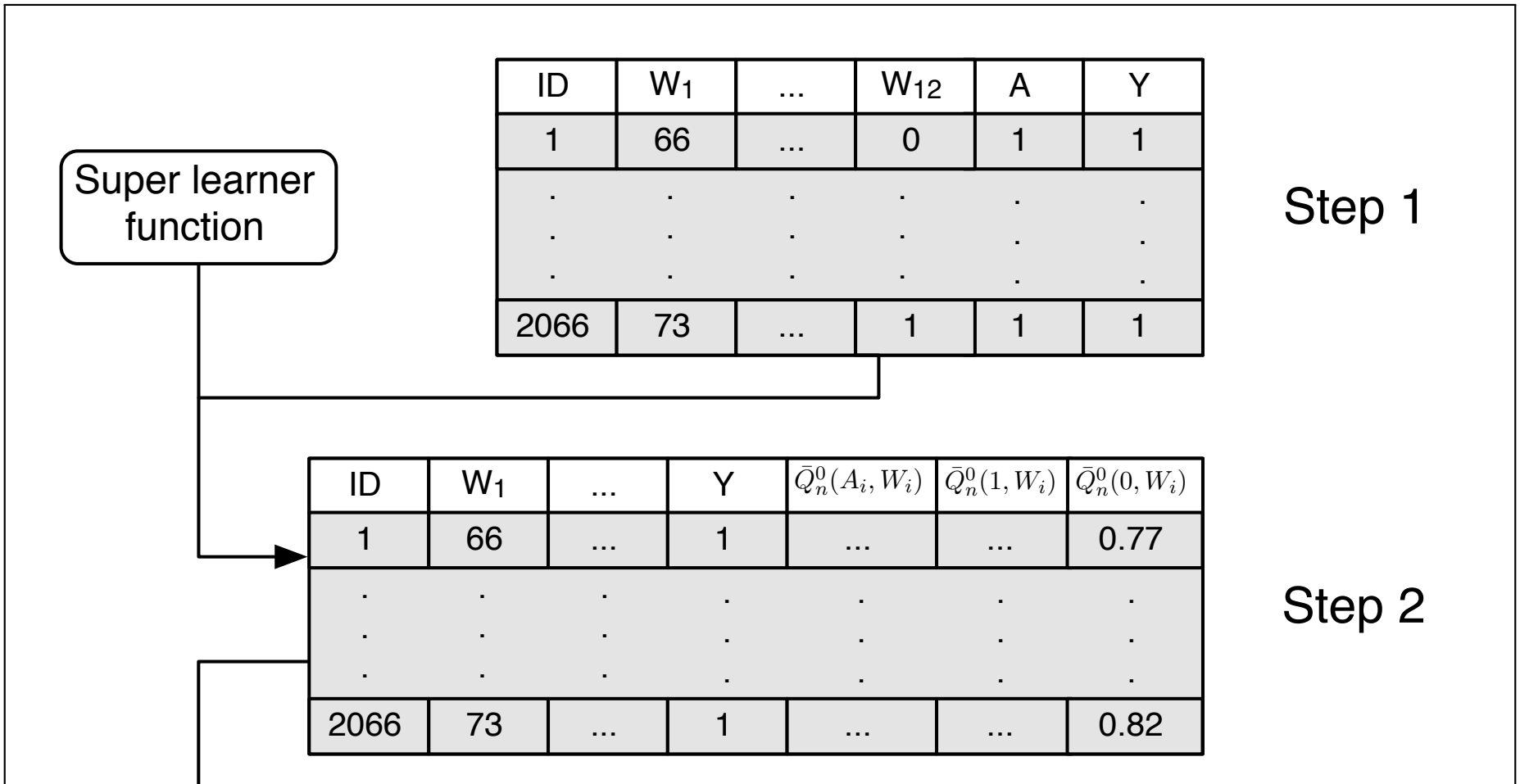
- Outcome was death.
- Covariates were gender, age, **self-rated health**, **leisure-time physical activity**, smoking status, cardiac event history, and chronic health condition status.
- The data structure is $O = (W, A, Y)$, where $Y = I(T \leq 5 \text{ years})$, T is time to the event death
- No right censoring in this cohort.

Sonoma Study

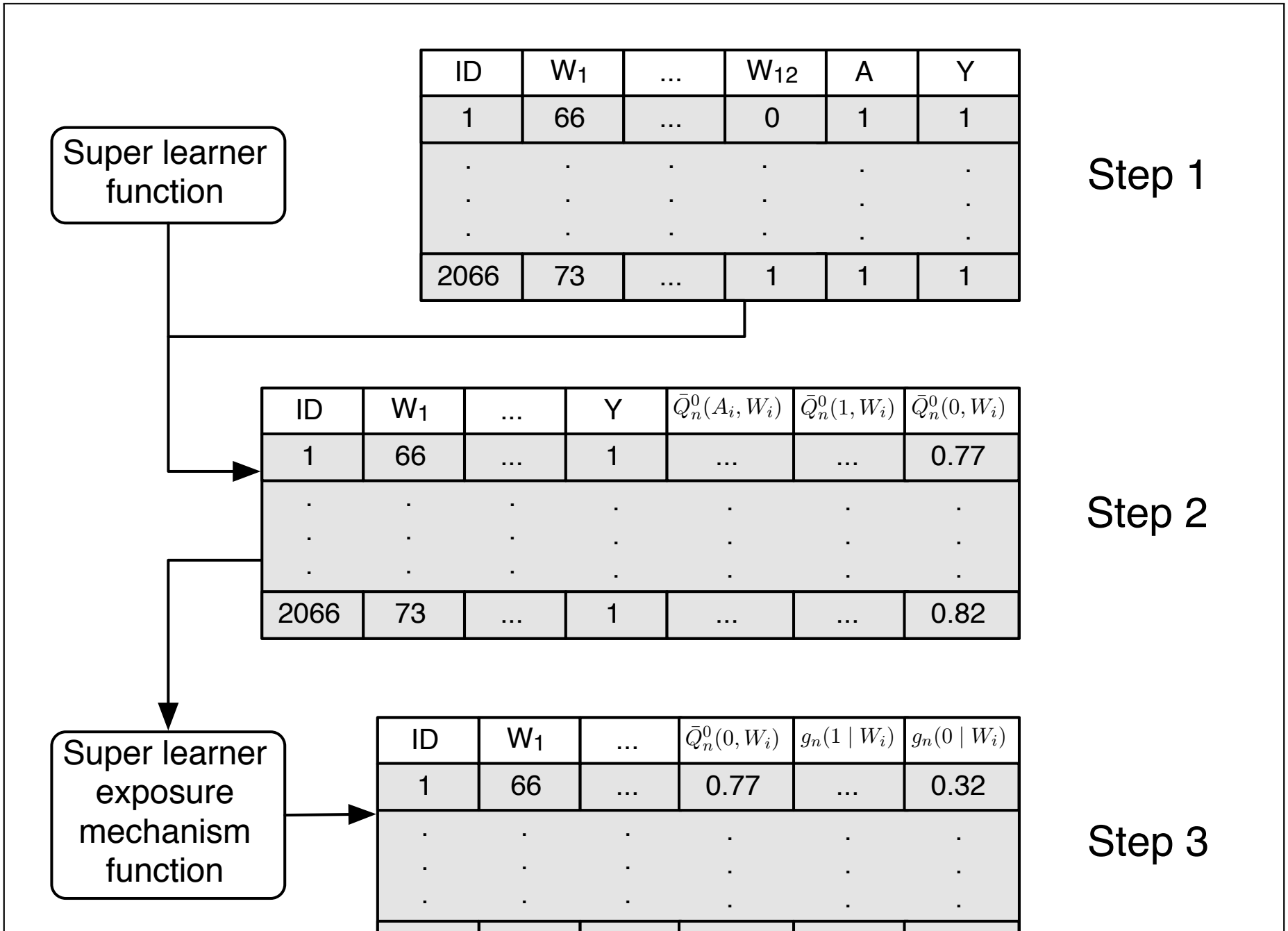
Variable	Description
Y	Death occurring within 5 years of baseline
A	LTPA score $\geq$ 22.5 METs at baseline[‡]
W_1	Health self-rated as "excellent"
W_2	Health self-rated as "fair"
W_3	Health self-rated as "poor"
W_4	Current smoker
W_5	Former smoker
W_6	Cardiac event prior to baseline
W_7	Chronic health condition at baseline
W_8	$x \leq 60$ years old
W_9	$60 < x \leq 70$ years old
W_{10}	$80 < x \leq 90$ years old
W_{11}	$x > 90$ years old
W_{12}	Female

[‡] LTPA is calculated from a detailed questionnaire where prior performed vigorous physical activities are assigned standardized intensity values in metabolic equivalents (METs). The recommended level of energy expenditure for the elderly is 22.5 METs.

Sonoma Study



Sonoma Study: **Estimating** $\bar{Q}_0$



Sonoma Study: **Estimating** $\bar{Q}_0$

At this stage we could plug our estimates $\bar{Q}_n^0(1, W_i)$ and $\bar{Q}_n^0(0, W_i)$ for each subject into our substitution estimator of the risk difference:

$$\psi_{MLE,n} = \Psi(Q_n) = \frac{1}{n} \sum_{i=1}^{n} \{\bar{Q}_n^0(1, W_i) - \bar{Q}_n^0(0, W_i)\}.$$

Sonoma Study: **Estimating** g_0

Our targeting step required an estimate of the conditional distribution of LTPA given covariates W.

This estimate of $P_0(A \mid W) \equiv g_0$ is denoted g_n.

We estimated predicted values using a super learner prediction function, adding two more columns to our data matrix: $g_n(1 \mid W_i)$ and $g_n(0 \mid W_i)$.

(Step 3.)

ID	W_1	...	Y	$\bar{Q}_n^0(A_i, W_i)$	$\bar{Q}_n^0(1, W_i)$	$\bar{Q}_n^0(0, W_i)$
1	66	...	1	...	...	0.77
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
2066	73	...	1	...	...	0.82

Step 2

Super learner exposure mechanism function

ID	W_1	...	$\bar{Q}_n^0(0, W_i)$	$g_n(1 \mid W_i)$	$g_n(0 \mid W_i)$
1	66	...	0.77	...	0.32
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
2066	73	...	0.82	...	0.45

Step 3

ID	W_1	...	$g_n(0 \mid W_i)$	$H_n^*(A_i, W_i)$	$H_n^*(1, W_i)$	$H_n^*(0, W_i)$
1	66	...	0.32	...	...	-3.13
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
2066	73	...	0.45	...	...	-2.22

Step 4

ID	W_1	...	$H^*(0, W_i)$	$\bar{Q}^1(1, W_i)$	$\bar{Q}^1(0, W_i)$

Sonoma Study: **Determining a Submodel**

The targeting step used the estimate g_n in a clever covariate to define a parametric working model coding fluctuations of the initial estimator. This clever covariate $H_n^*(A, W)$ is given by

$$H_n^*(A, W) \equiv \left(\frac{I(A = 1)}{g_n(1 \mid W)} - \frac{I(A = 0)}{g_n(0 \mid W)} \right).$$

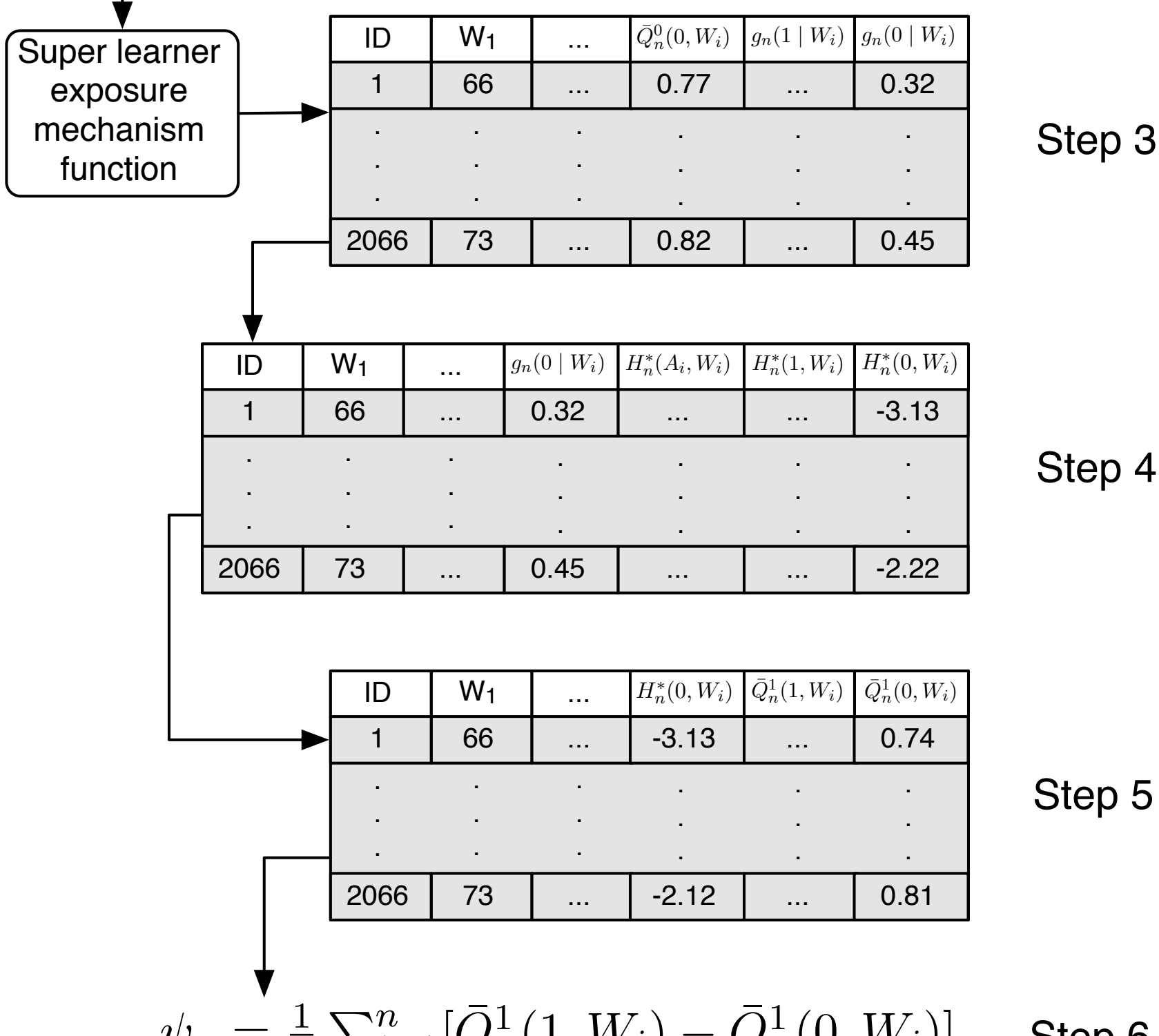
Sonoma Study: **Determining a Submodel**

Thus, for each subject with $A_i = 1$ in the observed data, we calculated the clever covariate as $H_n^*(1, W_i) = 1/g_n(1 \mid W_i)$.

Similarly, for each subject with $A_i = 0$ in the observed data, we calculated the clever covariate as $H_n^*(0, W_i) = -1/g_n(0 \mid W_i)$.

We combined these values to form a single column $H_n^*(A_i, W_i)$ in the data matrix. We also added two columns $H_n^*(1, W_i)$ and $H_n^*(0, W_i)$. The values for these columns were generated by setting $a = 0$ and $a = 1$.

(Step 4.)


Super learner exposure mechanism function
ID W1 ... $\bar{Q}_n^0(0, W_i)$ $g_n(1 \mid W_i)$ $g_n(0 \mid W_i)$
1 66 ... 0.77 ... 0.32
2066 73 ... 0.82 ... 0.45
Step 3
ID W1 ... $g_n(0 \mid W_i)$ $H_n^*(A_i, W_i)$ $H_n^*(1, W_i)$ $H_n^*(0, W_i)$
1 66 ... 0.32 ... ... -3.13
2066 73 ... 0.45 ... ... -2.22
Step 4
ID W1 ... $H_n^*(0, W_i)$ $\bar{Q}_n^1(1, W_i)$ $\bar{Q}_n^1(0, W_i)$
1 66 ... -3.13 ... 0.74
2066 73 ... -2.12 ... 0.81
Step 5
$\psi_n = \frac{1}{n} \sum_{i=1}^{n} [\bar{Q}^1(1, W_i) - \bar{Q}^1(0, W_i)]$
Step 6

Sonoma Study: **Updating** $\bar{Q}_n^0$

We then ran a logistic regression of our outcome Y on the clever covariate using as intercept the offset $\text{logit}\bar{Q}_n^0(A, W)$ to obtain the estimate ϵ_n, where ϵ_n is the resulting coefficient in front of the clever covariate $H_n^*(A, W)$.

We next wanted to update the estimate $\bar{Q}_n^0$ into a new estimate $\bar{Q}_n^1$ of the true regression function $\bar{Q}_0$:

$$\text{logit}\, \bar{Q}_n^1(A, W) = \text{logit}\, \bar{Q}_n^0(A, W) + \epsilon_n H_n^*(A, W).$$

This parametric working model incorporated information from g_n, through $H_n^*(A, W)$, into an updated regression.

Sonoma Study: **Updating** $\bar{Q}_n^0$

The TMLE of Q_0 was given by $Q_n^* = (\bar{Q}_n^1, Q_{W,n}^0)$. With ϵ_n, we were ready to update our prediction function at $a = 1$ and $a = 0$ according to the logistic regression working model. We calculated

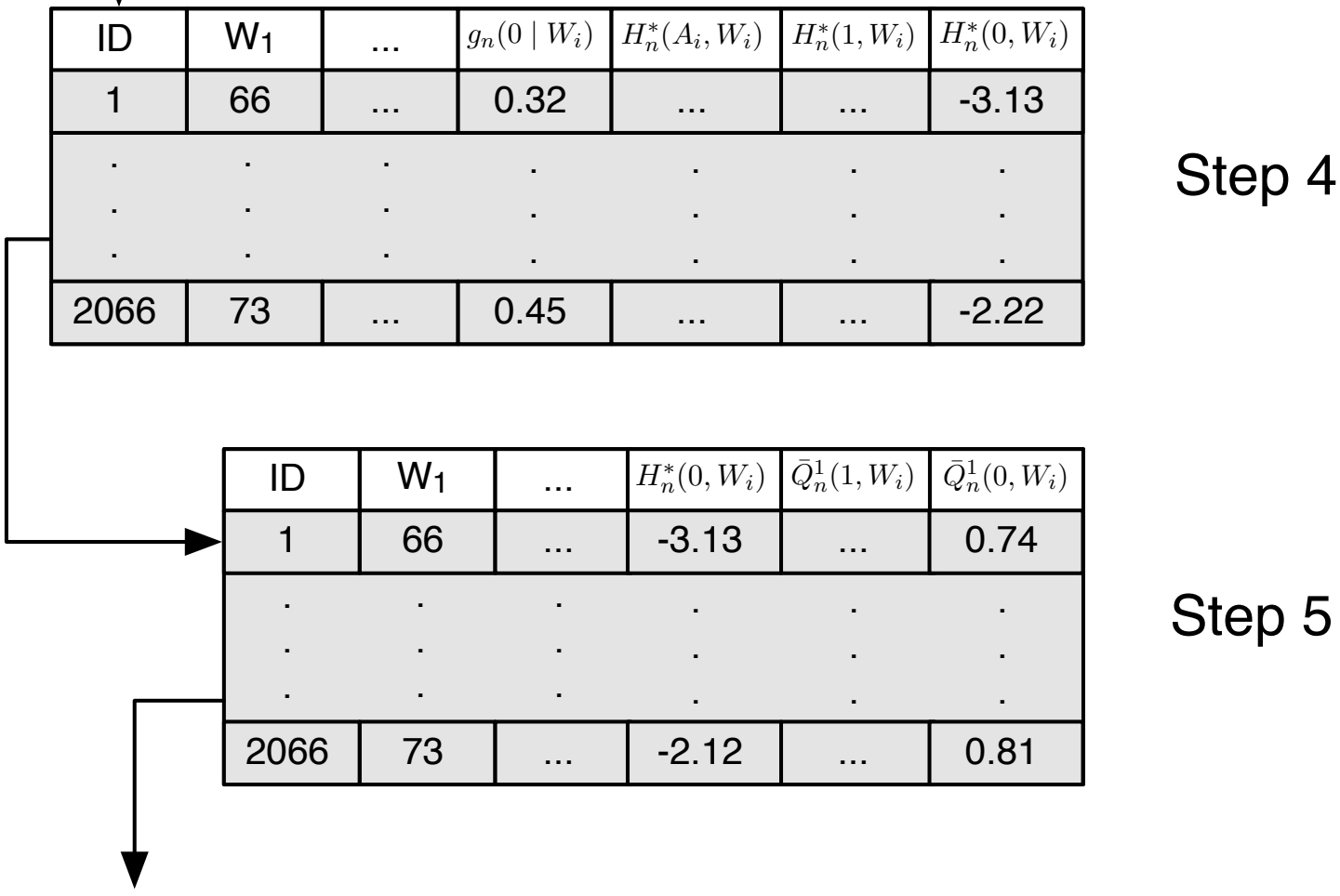
$$\text{logit}\, \bar{Q}_n^1(1, W) = \text{logit}\bar{Q}_n^0(1, W) + \epsilon_n H_n^*(1, W),$$

for all subjects, and then

$$\text{logit}\, \bar{Q}_n^1(0, W) = \text{logit}\bar{Q}_n^0(0, W) + \epsilon_n H_n^*(0, W)$$

for all subjects and added a column for $\bar{Q}_n^1(1, W_i)$ and $\bar{Q}_n^1(0, W_i)$ to the data matrix.

Updating $\bar{Q}_n^0$ is also illustrated in Step 5.


ID
W1
...
$g_n(0 \mid W_i)$
$H_n^*(A_i, W_i)$
$H_n^*(1, W_i)$
$H_n^*(0, W_i)$
1
66
...
0.32
...
...
-3.13
2066
73
...
0.45
...
...
-2.22
Step 4
ID
W1
...
$H_n^*(0, W_i)$
$\bar{Q}_n^1(1, W_i)$
$\bar{Q}_n^1(0, W_i)$
1
66
...
-3.13
...
0.74
2066
73
...
-2.12
...
0.81
Step 5
$\psi_n = \frac{1}{n} \sum_{i=1}^{n} [\bar{Q}_n^1(1, W_i) - \bar{Q}_n^1(0, W_i)]$
Step 6

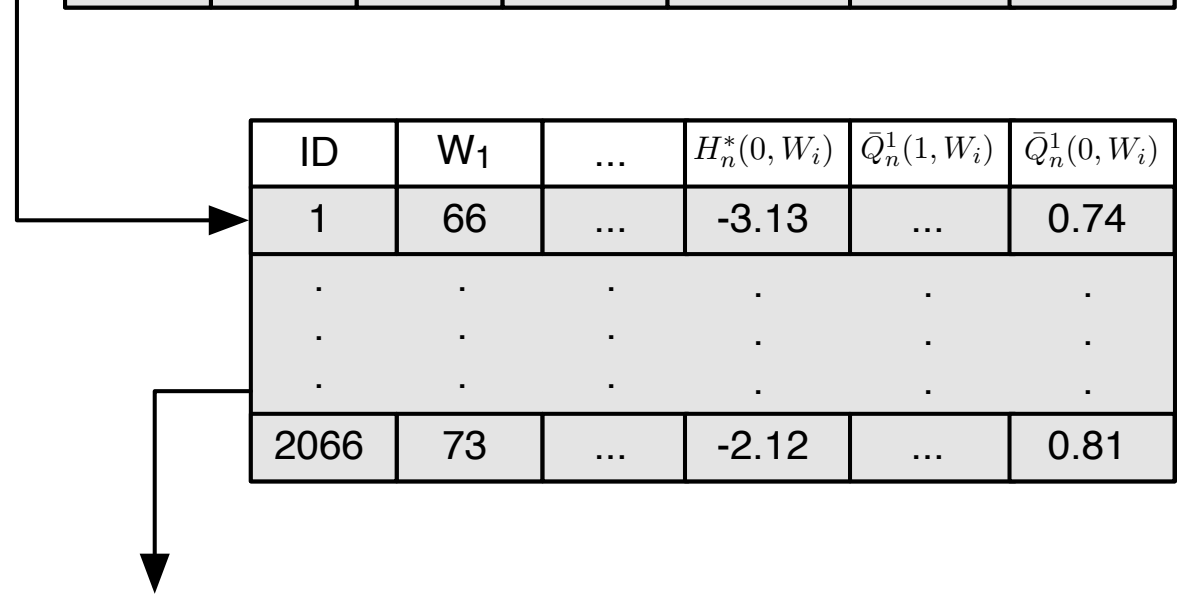
Sonoma Study: **Targeted Substitution Estimator**

Our formula from the first step becomes

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

This mapping was accomplished by evaluating $\bar{Q}_n^1(1, W_i)$ and $\bar{Q}_n^1(0, W_i)$ for each observation i, and plugging these values into the above equation.

Our estimate of the causal risk difference for the mortality study was $\psi_{TMLE,n} = -0.055$.



ID	W_1	...	$H_n^*(0, W_i)$	$\bar{Q}_n^1(1, W_i)$	$\bar{Q}_n^1(0, W_i)$
1	66	...	-3.13	...	0.74
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
2066	73	...	-2.12	...	0.81

Step 5

$$\psi_n = \frac{1}{n} \sum_{i=1}^{n} [\bar{Q}_n^1(1, W_i) - \bar{Q}_n^1(0, W_i)]$$

Step 6

Sonoma Study: **Inference (Standard errors)**

We then needed to calculate the influence curve for our estimator in order to obtain standard errors:

$$IC_n(O_i) = \left(\frac{I(A_i = 1)}{g_n(1 \mid W_i)} - \frac{I(A_i = 0)}{g_n(0 \mid W_i)} \right) (Y - \bar{Q}_n^1(A_i, W_i)) + \bar{Q}_n^1(1, W_i) - \bar{Q}_n^1(0, W_i) - \psi_{TMLE,n},$$

where I is an indicator function: it equals 1 when the logical statement it evaluates, e.g., $A_i = 1$, is true.

Sonoma Study: **Inference (Standard errors)**

Note that this influence curve is evaluated for each of the n observations O_i.

With the influence curve of an estimator one can now proceed with statistical inference as if the estimator minus its estimand equals the empirical mean of the influence curve.

Sonoma Study: **Inference (Standard errors)**

Next, we calculated the sample mean of these estimated influence curve values: $\bar{IC}_n = \frac{1}{n}\sum_{i=1}^{n} IC_n(o_i)$. For the TMLE we have $\bar{IC}_n = 0$. Using this mean, we calculated the sample variance of the estimated influence curve values:

$$S^2(IC_n) = \frac{1}{n}\sum_{i=1}^{n}\left(IC_n(o_i) - \bar{IC}_n\right)^2.$$

Lastly, we used our sample variance to estimate the standard error of our estimator:

$$\sigma_n = \sqrt{\frac{S^2(IC_n)}{n}}.$$

This estimate of the standard error in the mortality study was $\sigma_n = 0.012$.

Sonoma Study: **Inference (CIs)**

$$\psi_{TMLE,n} \pm z_{0.975} \frac{\sigma_n}{\sqrt{n}},$$

where z_α denotes the α-quantile of the standard normal density $N(0,1)$.

Sonoma Study: **Inference (p-values)**

A p-value for $\psi_{TMLE,n}$ can be calculated as:

$$2\left[1-\Phi\left(\left|\frac{\psi_{TMLE,n}}{\sigma_n/\sqrt{n}}\right|\right)\right],$$

where Φ denotes the standard normal cumulative distribution function.

The p-value was < 0.001 and the confidence interval was $[-0.078, -0.033]$.

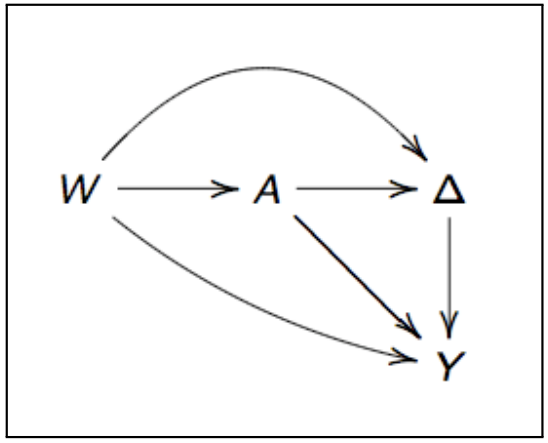
Sonoma Study: **Interpretation**

The interpretation of our estimate $\psi_{TMLE,n} = -0.055$, under causal assumptions, is that meeting or exceeding recommended levels of LTPA decreases 5-year mortality in an elderly population by 5.5 percentage points.

This result was significant, with a p-value of < 0.001 and a confidence interval of $[-0.078, -0.033]$.

Example: TMLE with Missingness

SCM 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.

The additive causal effect $EY_1 - EY_0$ equals:
$\Psi(P) = E[E(Y \mid A = 1, \Delta = 1, W) - E(Y \mid A = 0, \Delta = 1, W)$

Example: TMLE with Missingness

Our first step is to generate an initial estimator of P_n^0 of P; we estimate $E(Y \mid A, \Delta = 1, W)$, possible 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

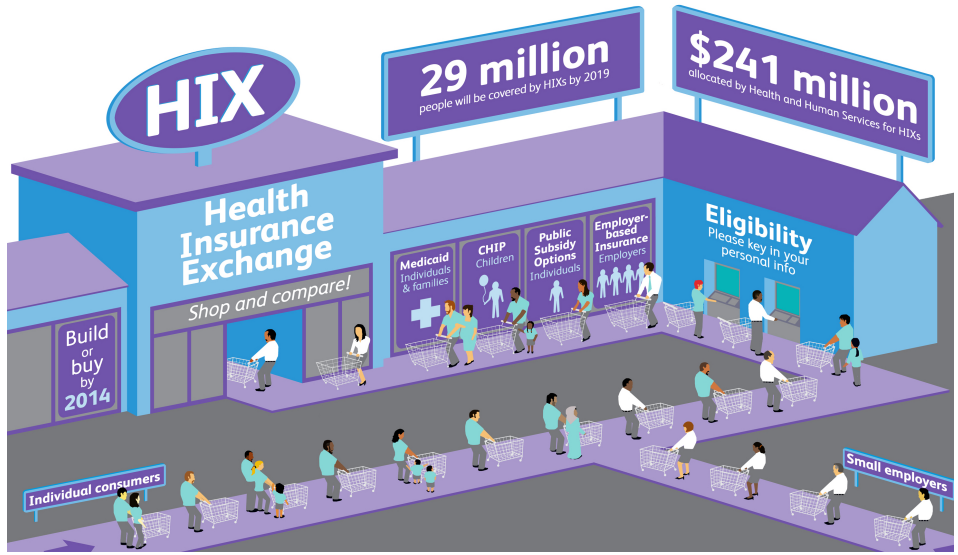
$$P_n^* = P_n^0(\epsilon_n).$$

The TMLE is given by $\Psi(P_n^*)$.

Plan Payment Risk Adjustment

Over 50 million people in the United States currently enrolled in an insurance program that uses risk adjustment.

- Redistributes funds based on health
- Encourages competition based on efficiency/quality



xerox.com

Results

- Machine learning finds novel insights
- Potential to impact policy, including diagnostic upcoding and fraud

Health Insurance Marketplace

Rose (2016)

Plan Payment Risk Adjustment: Key Results

1. **Super Learner had best performance.**
2. **Top 5 algorithms with reduced set of variables retained 92% of the relative efficiency of their full versions (86 variables).**
   - age category 21-34
   - all five inpatient diagnoses categories
     - heart disease
     - cancer
     - diabetes
     - mental health
     - other inpatient diagnoses
   - metastatic cancer
   - stem cell transplantation/complication
   - multiple sclerosis
   - end stage renal disease

But what if we care about the individual impact of medical condition categories on health spending?

TMLE Example: Impact of Medical Conditions

Evaluate how much more enrollees with each medical condition cost after controlling for demographic information and other medical conditions.

TMLE Example: Impact of Medical Conditions

Evaluate how much more enrollees with each medical condition cost after controlling for demographic information and other medical conditions.

HEALTH TRACKING

TRENDS

National Health Spending By Medical Condition, 1996–2005

Mental disorders and heart conditions were found to be the most costly.

by Charles Roehrig, George Miller, Craig Lake, and Jenny Bryant

ABSTRACT: This study responds to recent calls for information about how personal health expenditures from the National Health Expenditure Accounts are distributed across medical conditions. It provides annual estimates from 1996 through 2005 for thirty-two conditions mapped into thirteen all-inclusive diagnostic categories. Circulatory system spending was highest among the diagnostic categories, accounting for 17 percent of spending in 2005. The most costly conditions were mental disorders and heart conditions. Spending growth rates were lowest for lung cancer, chronic obstructive pulmonary disease, pneumonia, coronary heart disease, and stroke, perhaps reflecting benefits of preventive care. [*Health Affairs* 28, no. 2 (2009): w358–w367 (published online 24 February 2009; 10.1377/hlthaff.28.2.358)]

TMLE Example: Impact of Medical Conditions

Evaluate how much more enrollees with each medical condition cost after controlling for demographic information and other medical conditions.

HEALTH TRACKING

TRENDS

National Health Spending By Medical Condition, 1996–2005

Mental disorders and heart conditions were found to be the most costly.

by Charles Roehrig, George Miller, Craig Lake, and Jenny Bryant

ABSTRACT: This study responds to recent calls for information about how personal health expenditures from the National Health Expenditure Accounts are distributed across medical conditions. It provides annual estimates from 1996 through 2005 for thirty-two conditions mapped into thirteen all-inclusive diagnostic categories. Circulatory system spending was highest among the diagnostic categories, accounting for 17 percent of spending in 2005. The most costly conditions were mental disorders and heart conditions. Spending growth rates were lowest for lung cancer, chronic obstructive pulmonary disease, pneumonia, coronary heart disease, and stroke, perhaps reflecting benefits of preventive care. [*Health Affairs* 28, no. 2 (2009): w358–w367 (published online 24 February 2009; 10.1377/hlthaff.28.2.358)]

HEALTH SPENDING

Which Medical Conditions Account For The Rise In Health Care Spending?

The fifteen most costly medical conditions accounted for half of the overall growth in health care spending between 1987 and 2000.

by Kenneth E. Thorpe, Curtis S. Florence, and Peter Joski

ABSTRACT: We calculate the level and growth in health care spending attributable to the fifteen most expensive medical conditions in 1987 and 2000. Growth in spending by medical condition is decomposed into changes attributable to rising cost per treated case, treated prevalence, and population growth. We find that a small number of conditions account for most of the growth in health care spending—the top five medical conditions accounted for 31 percent. For four of the conditions, a rise in treated prevalence, rather than rising treatment costs per case or population growth, accounted for most of the spending growth.

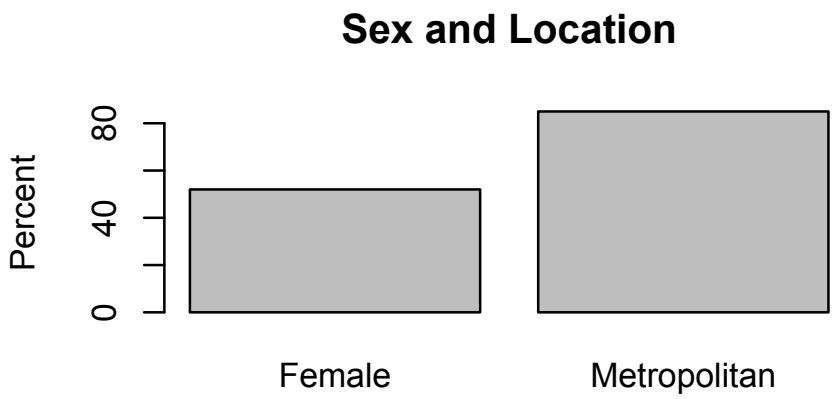
TMLE Example: Impact of Medical Conditions

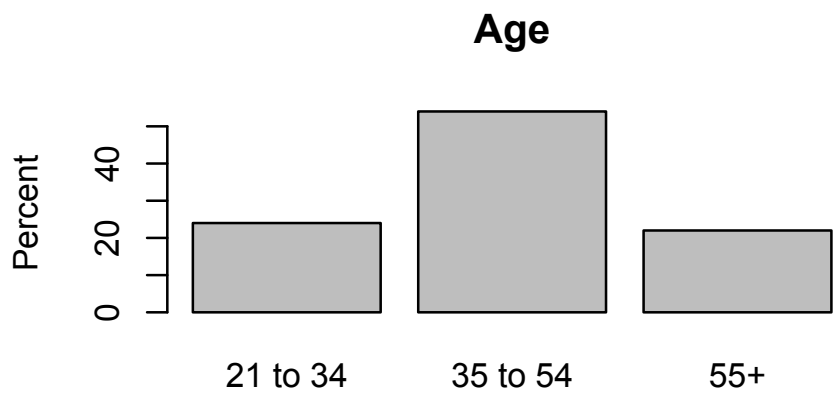
- **Truven MarketScan** database, those with continuous coverage in 2011-2012; 10.9 million people. Variables: age, sex, region, procedures, expenditures, etc.

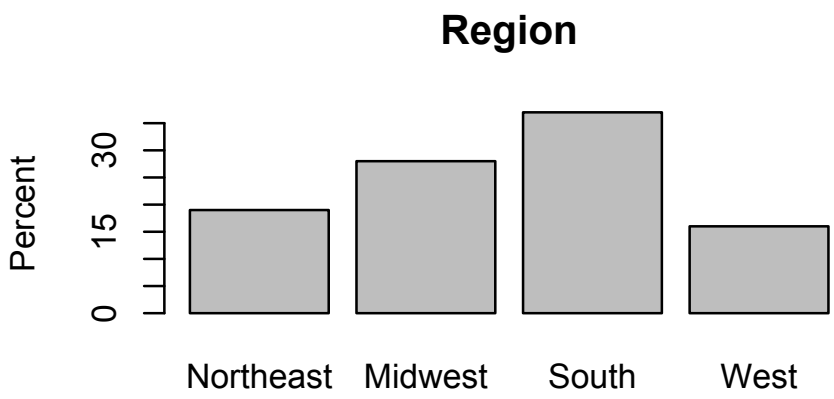
- Enrollment and claims from private health plans and employers.
- Extracted random sample of 1,000,000 people.
- Enrollees were eligible for insurance throughout this entire 24 month period and thus there is no drop-out due to death.

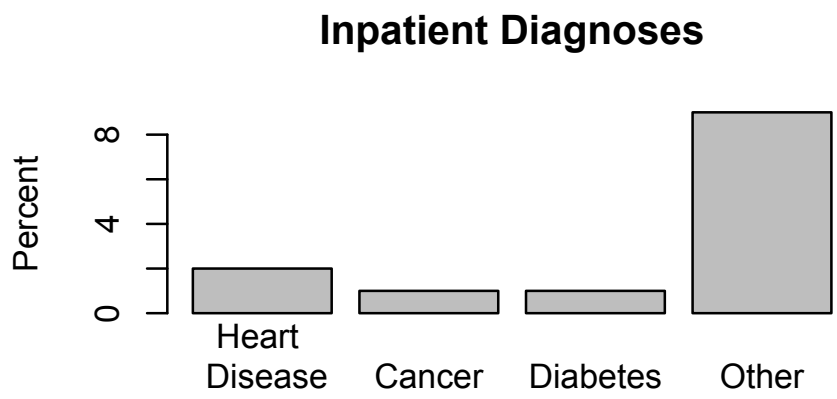
MARKETSCAN® RESEARCH

TMLE Example: Impact of Medical Conditions



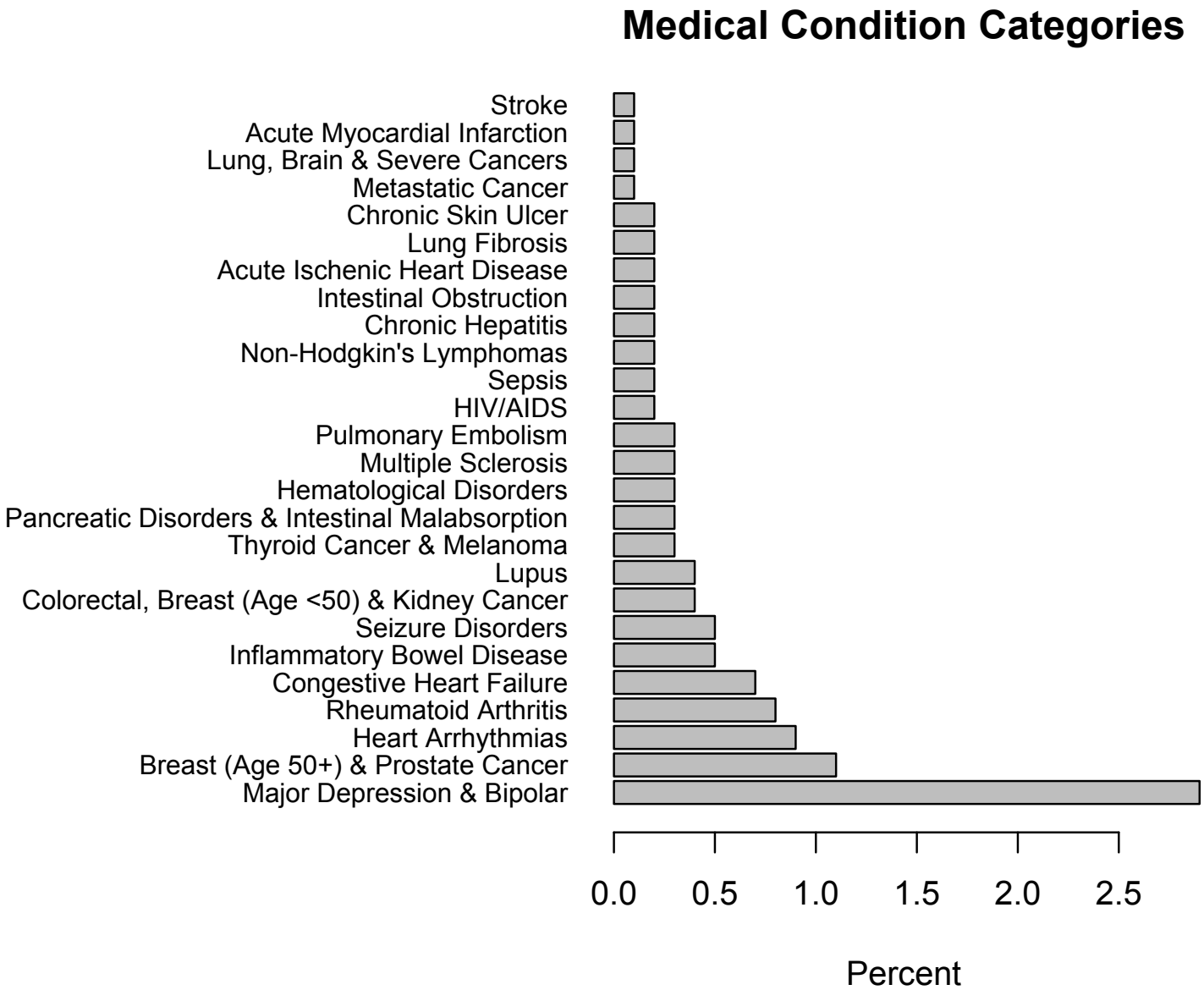






n=1,000,000

TMLE Example: Impact of Medical Conditions



n=1,000,000

TMLE Example: Impact of Medical Conditions

$$\psi = E_{W,M^-}[E(Y \mid A = 1, W, M^-) - E(Y \mid A = 0, W, M^-)],$$

represents the effect of $A = 1$ versus $A = 0$ after adjusting for all other medical conditions M^- and baseline variables W.

Interpretation

The difference in total annual expenditures when enrollees have the medical condition under consideration (i.e., $A = 1$).

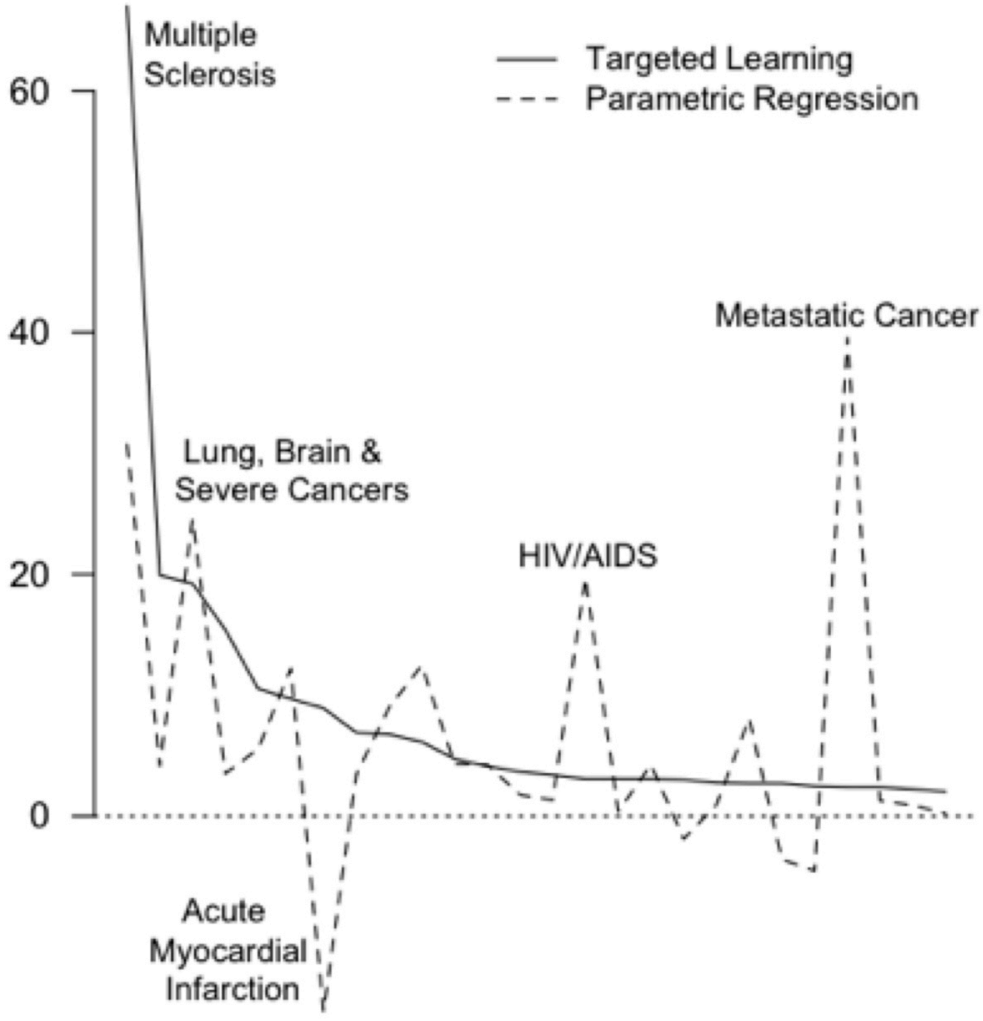
Y=total annual expenditures, A=medical condition category of interest

TMLE Example: Impact of Medical Conditions

Leverage

- available big data
- novel machine learning tools

to improve conclusions and policy insights



Rose (2017)

TMLE Example: Impact of Medical Conditions

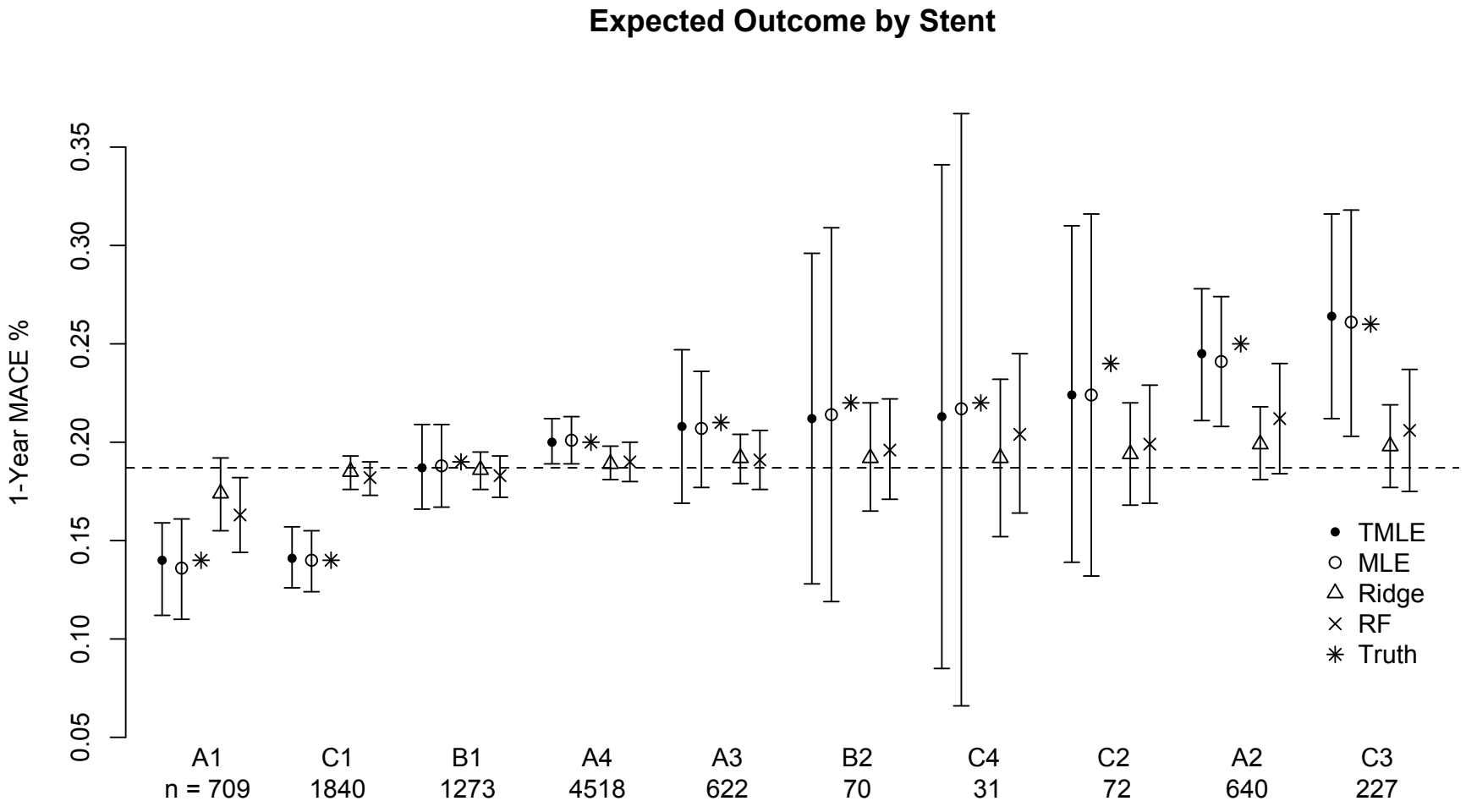
First investigation of the impact of medical conditions on health spending as a variable importance question using double robust estimators.

Five most expensive medical conditions were

1. multiple sclerosis
2. congestive heart failure
3. lung, brain, and other severe cancers
4. **major depression and bipolar disorders**
5. **chronic hepatitis.**

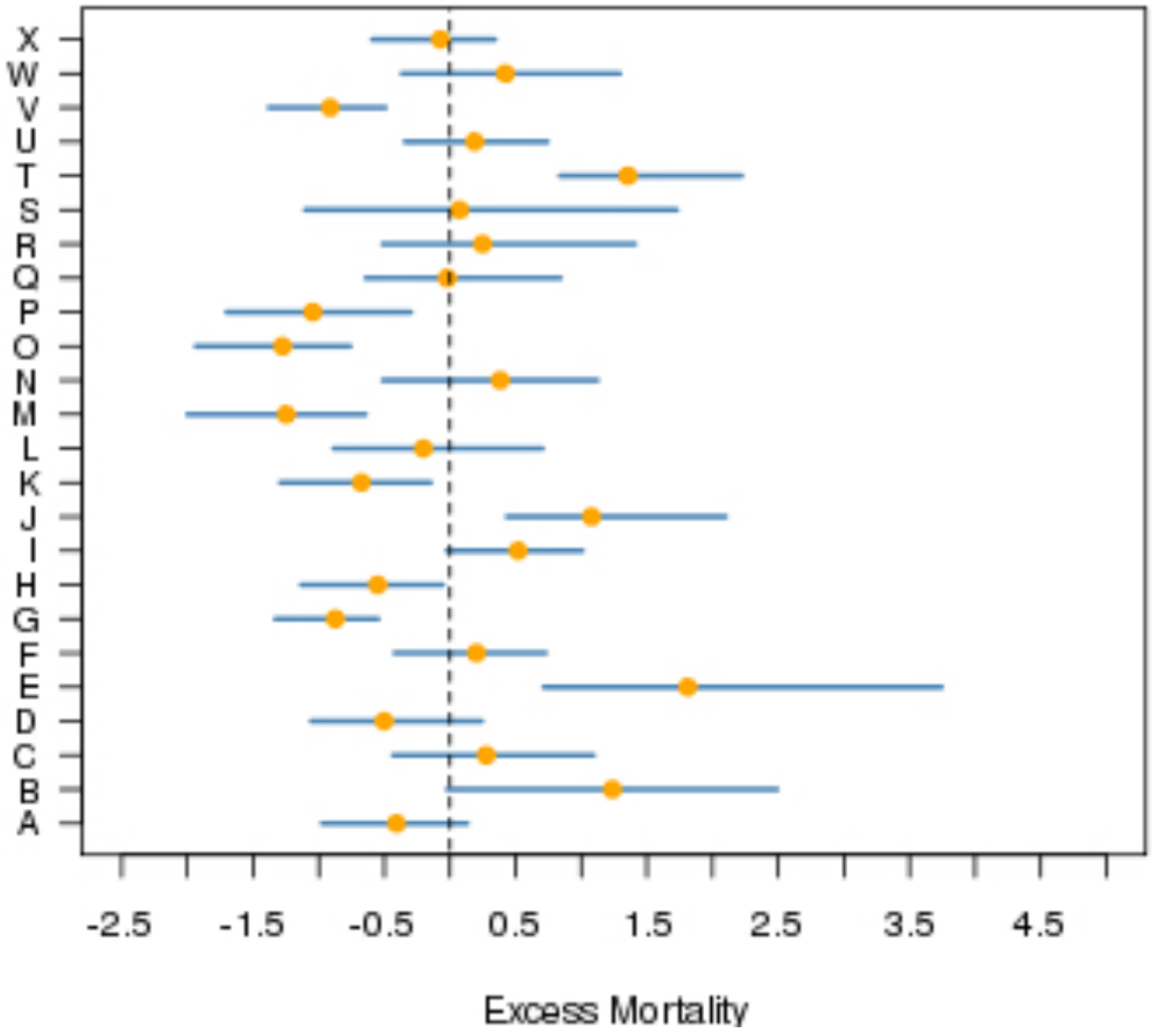
- Differing results compared to parametric regression.
- What does this mean for incentives for prevention and care?

Effect of Drug-Eluting Stents



Rose and Normand (2017)

Hospital Profiling



Spertus et al. (2016)

Effect Estimation Literature

- Maximum-Likelihood-Based Estimators: g-formula, Robins 1986
- Estimating equations: Robins and Rotnitzky 1992, Robins 1999, Hernan et al. 2000, Robins et al. 2000, Robins 2000, Robins and Rotnitzky 2001.
- Additional bibliographic history found in Chapter 1 of van der Laan and Robins 2003.
- For even more references, see Chapter 4 of *Targeted Learning*.

[TMLE Example Code]

TMLE Packages

- `tmle` (Gruber): Main point-treatment TMLE package
- `ltmle` (Schwab): Main longitudinal TMLE package
- SAS code (Brooks): Github
- Julia code (Lendle): Github

More: `targetedlearningbook.com/software`

[TMLE Example Code]

TMLE Sample Code

```
##Code lightly adapted from Schuler & Rose, 2017, AJE##
library(tmle)
set.seed(1)
N <- 1000
```

TMLE Sample Code

```
##Generate simulated data##

#X1=Gender; X2=Therapy; X3=Antidepressant use
X1 <- rbinom(N, 1, prob=.55)
X2 <- rbinom(N, 1, prob=.30)
X3 <- rbinom(N, 1, prob=.25)
W <- cbind(X1,X2,X3)

#Exposure=regular physical exercise
A <- rbinom(N, 1, plogis(-0.5 + 0.75*X1 + 1*X2 + 1.5*X3))

#Outcome=CES-D score
Y <- 24-3*A+3*X1-4*X2-6*X3-1.5*A*X3+rnorm(N,mean=0,sd=4.5)
```

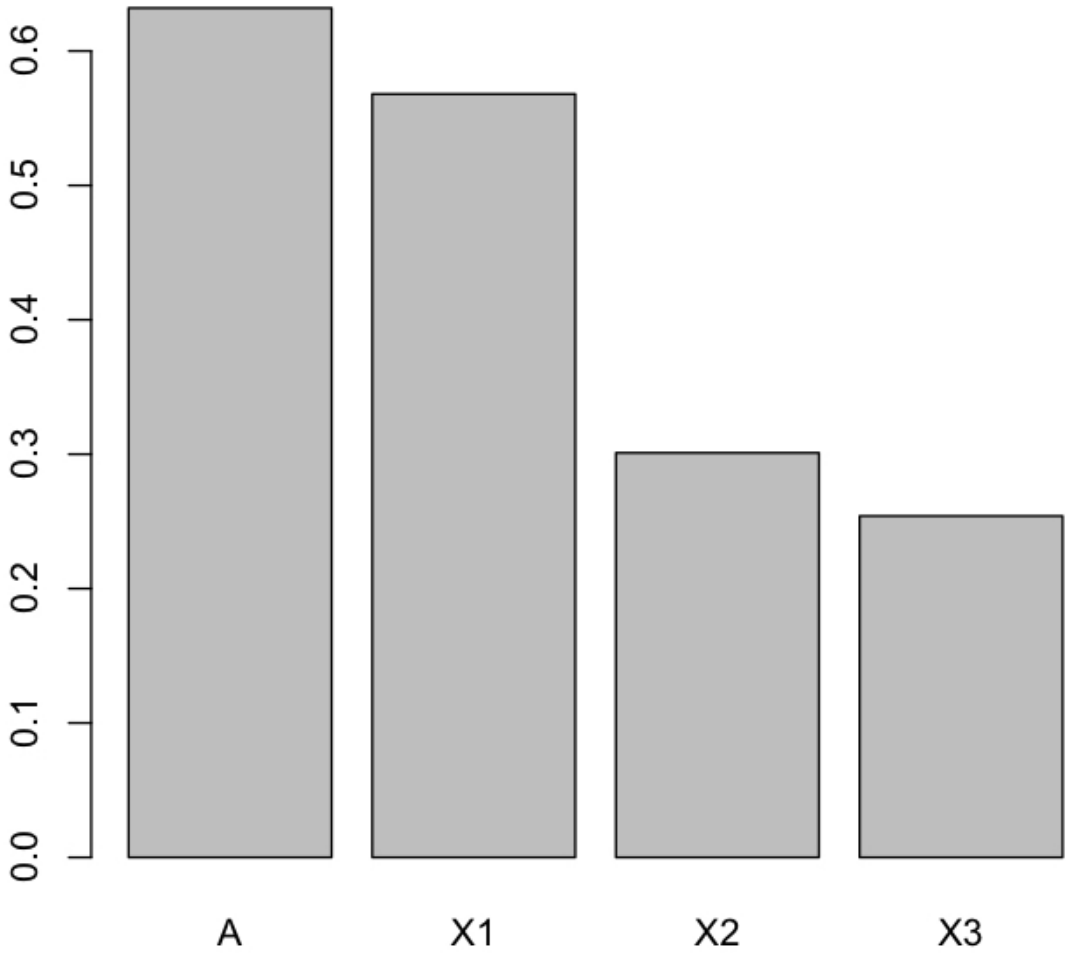
TMLE Sample Code

```
##Examine simulated data##

data <- data.frame(cbind(A,X1,X2,X3,Y))
summary(data)
barplot(colMeans(data[,1:4]))
```

TMLE Sample Code

```
> summary(data)
       A                X1                X2
 Min.   :0.000   Min.   :0.000   Min.   :0.000
 1st Qu.:0.000   1st Qu.:0.000   1st Qu.:0.000
 Median :1.000   Median :1.000   Median :0.000
 Mean   :0.632   Mean   :0.568   Mean   :0.301
 3rd Qu.:1.000   3rd Qu.:1.000   3rd Qu.:1.000
 Max.   :1.000   Max.   :1.000   Max.   :1.000
       X3                Y
 Min.   :0.000   Min.   : 0.9629
 1st Qu.:0.000   1st Qu.:16.4549
 Median :0.000   Median :21.2744
 Mean   :0.254   Mean   :20.8369
 3rd Qu.:1.000   3rd Qu.:25.2316
 Max.   :1.000   Max.   :39.8796
```

TMLE Sample Code



TMLE Sample Code

```
##Specify a library of algorithms##

SL.library <- c("SL.glm","SL.step.interaction","SL.glmnet",
     "SL.randomForest","SL.gam","SL.rpart" )
```

TMLE Sample Code

Could use various forms of "screening" to consider differing variable sets

```
SL.library <- list(c("SL.glm","screen.randomForest", "All"),
       c("SL.mean", "screen.randomForest", "All"),
       c("SL.randomForest", "screen.randomForest", "All"),
       c("SL.glmnet", "screen.randomForest","All"))
```

Or the same algorithm with different tuning parameters

```
SL.glmnet.alpha0 <- function(..., alpha=0){
       SL.glmnet(..., glmnet.alpha=alpha)}
SL.glmnet.alpha50 <- function(..., alpha=.50){
       SL.glmnet(..., glmnet.alpha=alpha)}

SL.library <- c("SL.glm","SL.glmnet", "SL.glmnet.alpha50",
       "SL.glmnet.alpha0","SL.randomForest")
```

TMLE Sample Code

```
##Specify a library of algorithms##

SL.library <- c("SL.glm","SL.step.interaction","SL.glmnet",
     "SL.randomForest","SL.gam","SL.rpart" )
```

TMLE Sample Code

```
##TMLE approach: Super Learning##

tmleSL1 <- tmle(Y, A, W,
       Q.SL.library = SL.library, g.SL.library = SL.library)
tmleSL1
```

TMLE Sample Code

```
> tmleSL1 <- tmle(Y, A, W, Q.SL.library = SL.library,
g.SL.library = SL.library)
Loading required package: gam
Loading required package: splines
Loading required package: foreach
foreach: simple, scalable parallel programming from
Revolution Analytics
Use Revolution R for scalability, fault tolerance and more.
http://www.revolutionanalytics.com
Loaded gam 1.14

Loading required package: glmnet
Loading required package: Matrix
Loaded glmnet 2.0-2

Loading required package: randomForest
randomForest 4.6-12
Type rfNews() to see new features/changes/bug fixes.
Loading required package: rpart
```

TMLE Sample Code

```
> tmleSL1
 Additive Effect
    Parameter Estimate:  -3.4074
    Estimated Variance:  0.11084
               p-value:  <2e-16
     95% Conf Interval: (-4.0599, -2.7549)
```

True value is -3.38

TMLE Sample Code

```
##TMLE approach: GLM, MT misspecification of outcome##
#Misspecified outcome regression:  Y ~ A + X1 + X2 + X3#

tmleGLM1 <- tmle(Y, A, W, Qform=Y~A+X1+X2+X3, gform=A~X1+X2+X3)
tmleGLM1
```

TMLE Sample Code

```
> tmleGLM1 <- tmle(Y, A, W, Qform=Y~A+X1+X2+X3,
gform=A~X1+X2+X3)
> tmleGLM1
 Additive Effect
    Parameter Estimate:  -3.416
    Estimated Variance:  0.1132
                p-value:  <2e-16
     95% Conf Interval: (-4.0754, -2.7565)
```

True value is -3.38

TMLE Sample Code

```
##TMLE approach: GLM, OV misspecification of outcome##
#Misspecified outcome regression:  Y ~ A + X1 + X2#

tmleGLM2 <- tmle(Y, A, W, Qform=Y~A+X1+X2, gform=A~X1+X2+X3)
tmleGLM2
```

TMLE Sample Code

```
> tmleGLM2 <- tmle(Y, A, W, Qform=Y~A+X1+X2,
gform=A~X1+X2+X3)
> tmleGLM2
 Additive Effect
    Parameter Estimate:  -3.3976
    Estimated Variance:  0.1416
                p-value:  <2e-16
     95% Conf Interval: (-4.1351, -2.6601)
```

True value is -3.38

TMLE Sample Code

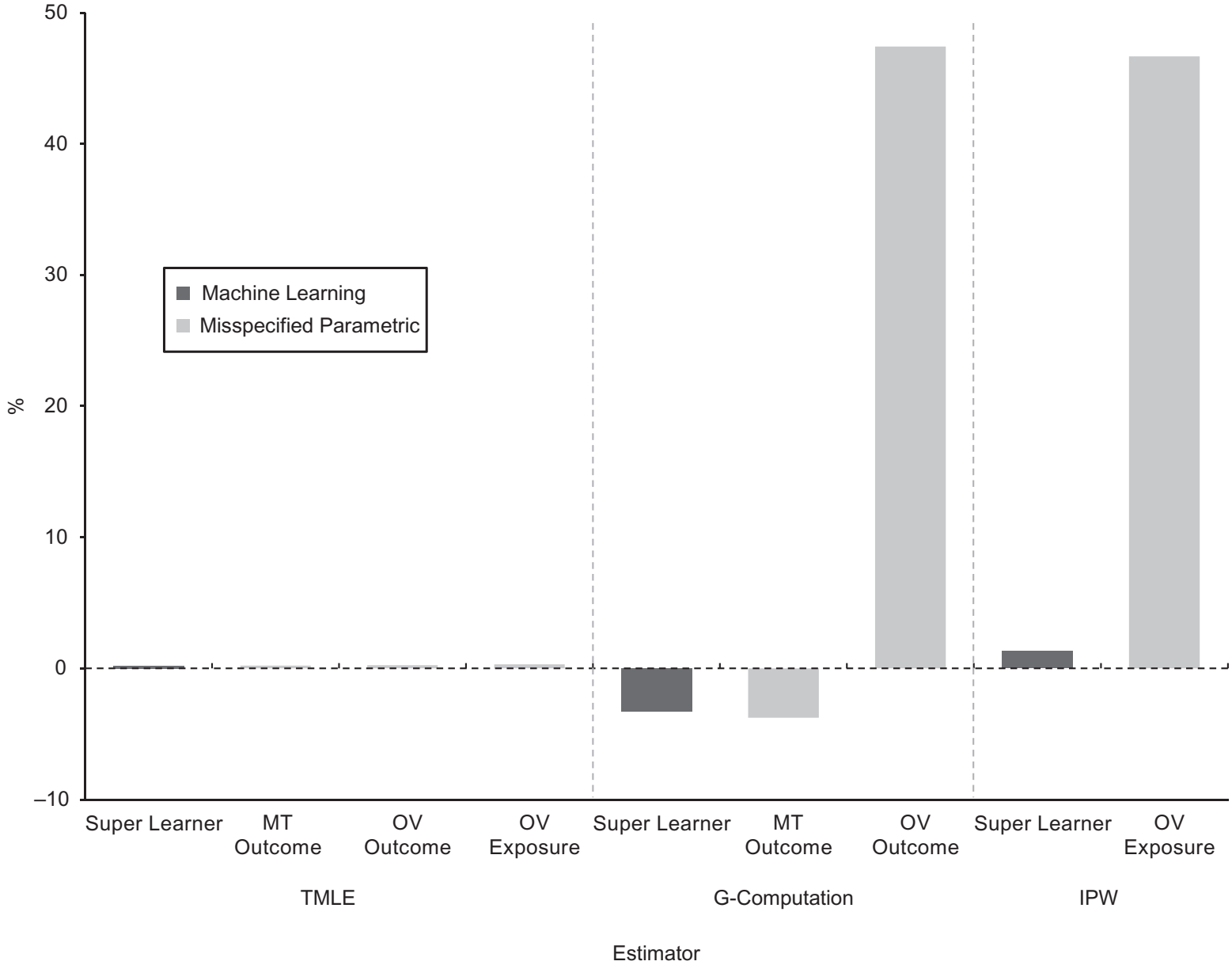
```
##TMLE approach: GLM, OV misspecification of exposure##
#Misspecified exposure regression:  A ~ X1 + X2#

tmleGLM3 <- tmle(Y, A, W, Qform=Y~A+X1+X2+X3+A:X3, gform=A~X1+X2)
tmleGLM3
```

TMLE Sample Code

```
> tmleGLM3 <- tmle(Y, A, W, Qform=Y~A+X1+X2+X3+A:X3,
gform=A~X1+X2)
> tmleGLM3
 Additive Effect
    Parameter Estimate:  -3.4277
    Estimated Variance:  0.10156
                p-value:  <2e-16
     95% Conf Interval: (-4.0524, -2.8031)
```

True value is -3.38

TMLE Sample Code



Schuler and Rose (2017)

TMLE Sample Code

	G-Computation		
Super learner			
Outcome variables: A, X_1, X_2, X_3	−3.27 (0.35)	0.11	−3.98, −2.56
Misspecified parametric regression			
Main-terms misspecification			
Outcome variables: A, X_1, X_2, X_3	−3.25 (0.33)	0.13	−3.91, −2.59
Omitted-variable misspecification			
Outcome variables: A, X_1, X_2	−4.98 (0.37)	−1.60	−5.69, −4.24[b]
	Inverse Probability Weighting		
Super learner			
Exposure variables: X_1, X_2, X_3	−3.43 (0.37)	−0.05	−4.17, −2.63
Misspecified parametric regression			
Omitted-variable misspecification			
Exposure variables: X_1, X_2	−4.96 (0.37)	−1.58	−5.67, −4.21[b]

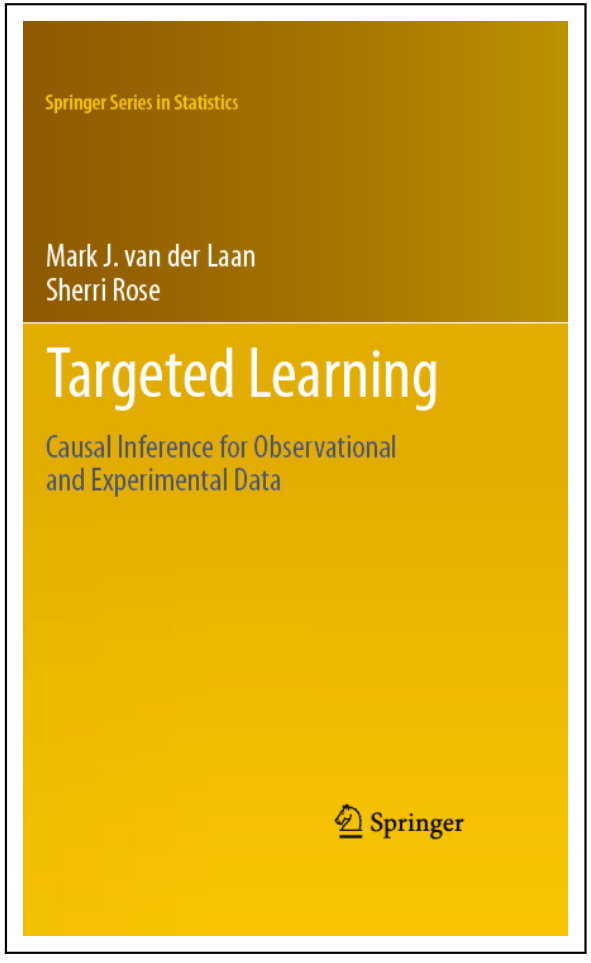
Schuler and Rose (2017)

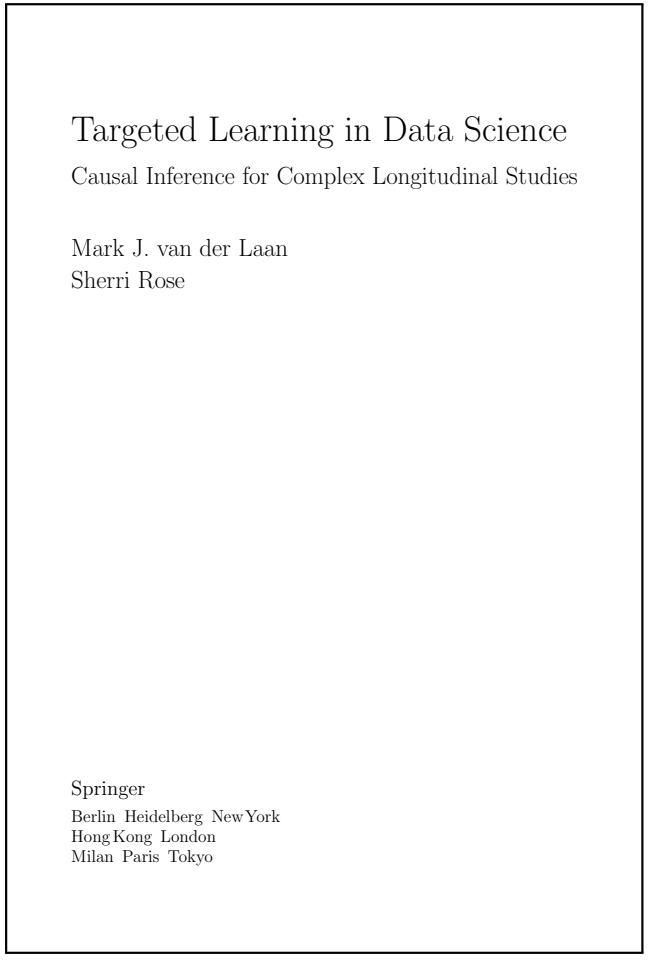
TMLE Packages

- `tmle` (Gruber): Main point-treatment TMLE package
- `ltmle` (Schwab): Main longitudinal TMLE package
- SAS code (Brooks): Github
- Julia code (Lendle): Github

More: `targetedlearningbook.com/software`

Targeted Learning (targetedlearningbook.com)





van der Laan & Rose, *Targeted Learning: Causal Inference for Observational and Experimental Data*. New York: Springer, 2011.

[Q & A]