School of Public Health
UNIVERSITY OF CALIFORNIA, BERKELEY

Estimating the causal effects of longitudinal exposures

Novel statistical methods and software

Maya Petersen and Joshua Schwab

Joint work with

Mark van der Laan

Div. of Biostatistics, School of Public Health,
University of California, Berkeley

IeDEA Network, CROI; Atlanta, GA, March 4th, 2013

Motivation

- A causal inference framework allows us to define many important questions in HIV
  - Effects of longitudinal exposures
  - Censoring dependent on time varying variables
  - Effects of "dynamic" treatment strategies
- Answering these questions requires moving beyond standard multivariable regression
  - Inverse Probability Weighted Estimators (IPTW)
  - Targeted Maximum Likelihood Estimators (TMLE)
- New R software implementing these estimators is now available

Outline

1. What types of questions can these methods help us answer?
2. Why are more complex statistical methods needed?
   - Challenge of time-varying confounders
3. Overview of available estimators
   - Inverse Probability of Treatment Weighted (IPTW)
   - Targeted Maximum Likelihood (TMLE)
4. Introduction to R software: Joshua Schwab
   - Sample data, code, and simulation results
5. Conclusion

Outline

1. What types of questions can these methods help us answer?
2. Why are more complex statistical methods needed?
   - Challenge of time-varying confounders
3. Overview of available estimators
   - Inverse Probability of Treatment Weighted (IPTW)
   - Targeted Maximum Likelihood (TMLE)
4. Introduction to R software: Joshua Schwab
   - Sample data, code, and simulation results
5. Conclusion

Overview: What types of questions these methods can let us answer?

- Effects of longitudinal treatments
  - Cumulative effect of multiple treatment decisions or exposures over time
    - Ex: Cumulative exposure to a specific ARV (Abacavir)
    - Ex: Time to starting ART after HIV diagnosis
- Effects of dynamic treatments
  - Strategies for assigning treatment in response to patient characteristics
    - Ex: Different CD4 thresholds for ART start
    - Ex: Different CD4, clinical or viral load-based strategies for defining 1st line failure and initiating switch

Overview: What types of questions these methods can let us answer?

- Exposure effects when censoring can depend on time-varying variables
  - Ex: Patients who get sicker over course of study more likely lost to follow up
- Effects on many types of outcomes
  - At a single time point
    - Ex: Survival at 12 months
  - Repeated measures
    - Ex: CD4 over time
  - Time to event
    - Ex: Survival time

Example: When to switch to second line ART following virological failure?

- HIV can develop resistance to 1st line NNRTI-based therapy quickly, resulting in HIV RNA rebound (virological failure)
- In Africa, delayed switch to second line ART following virological failure is common
  - Routine plasma viral load monitoring often unavailable
  - CD4 and clinical criteria detect rebound poorly
- Clinical effect of delayed switch in resource-limited settings not adequately quantified

Example 1: Point treatment

- Treatment: A
  - A single treatment decision or an exposure at a single time point
  - Ex: Indicator if switched immediately (first visit failure detected)
- Outcome: Y
  - Ex. Death within 1 quarter (3 months) after failure
- Counterfactual outcome: Outcome that would have been observed under a specific treatment
  - Ex: An individual's vital status had she switched immediately (in reality she may not have)

Example 1: Point treatment

- Example target quantity:
  - Proportion of patients that would have died within 1 quarter if all had been switched immediately
- Can use this to define various effects
  - Causal Risk Difference/Additive Treatment Effect
    - Ex: Difference in the proportion who would have died if all had been switched immediately versus if none had been switched immediately
  - Causal Relative Risk
  - Causal Odds Ratio
  - Etc…

Example 2: Longitudinal treatment

- Time point: t=1,…,K (end of study)
  - Ex: quarterly intervals at which patients seen and treatment decisions made
- Treatment: A(t), t=1,…,K
  - Treatment decisions at multiple time points
  - Ex: A(t)= Indicator if switched by time *t*
    - A(1)=0, A(2)=0, A(3)=1, A(4)=1 : switch at time 3
- Outcome: Y(t), t=1,…,K
  - Ex: Vital status at end of each interval

Example 2: Longitudinal treatment

- Example counterfactual outcome:
  - An individual's vital status at time *t* under a specific switch time
- Example target quantity:
  - The proportion of patients that would have died within *t* time points had none switched at any point
- Can again use this to define various effects
  - Ex: Difference in the proportion that would died within *t* time points had all switched immediately vs. had none switched at any point

Example 3: Marginal Structural Models

- Model how the expected counterfactual outcome varies as a function of treatment
  - Can model survival or hazard for a single or multiple time points
  - For point or longitudinal treatments
- Ex: How counterfactual probability of death within 3 time points varies as a function of switch time
  - Hypothetical randomized trial: Randomly assign subjects with virological failure a switch time and measure their survival
  - Ex MSM: $\text{logit}(E(Y(3)_{\text{switch}}))=\beta_0 + \beta_1 \text{switch time}$

Outline

1. What types of questions can these methods help us answer?
2. Why are more complex statistical methods needed?
   - Challenge of time-varying confounders
3. Overview of available estimators
   - Inverse Probability of Treatment Weighted (IPTW)
   - Targeted Maximum Likelihood (TMLE)
4. Introduction to R software: Joshua Schwab
   - Sample data, code, and simulation results
5. Conclusion

The methodological challenge

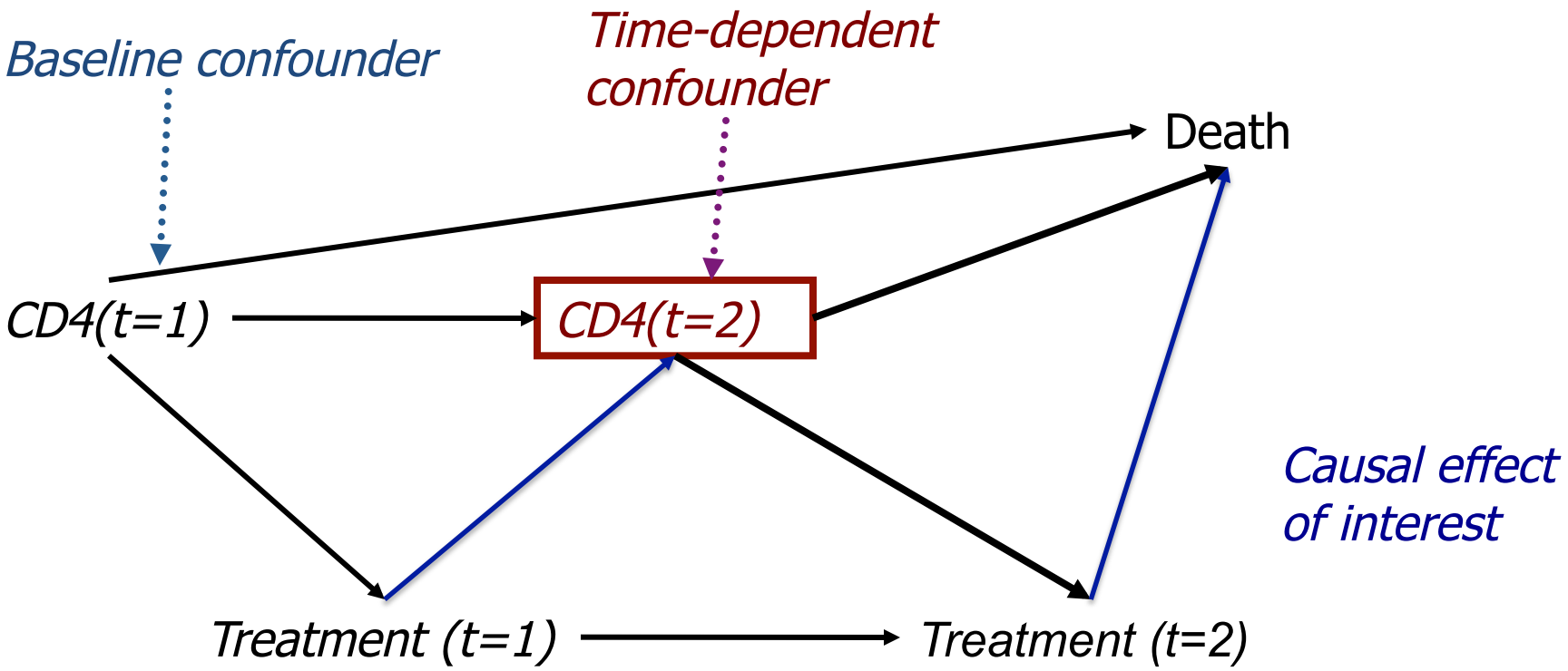
- For results to have causal interpretation, all methods (including these) rely on untestable assumptions
- We want

1. A clear understanding of those assumptions
   - Study designs to optimize their plausibility
2. Statistical methods that give the best possible answers given what we measure
   - Standard parametric regression not sufficient

Time-dependent confounding

- Time varying variables can
  1. Confound treatment-outcome relationships
  2. Be affected by prior treatment
     - Part of the causal pathway of interest
- How to analyze?
  - If we don't adjust -> Bias
  - If we adjust using stratification or standard multivariable regression -> Bias
- Similar issue when censoring depends on time varying variables

Ex: Time-dependent confounding



- Can't control for *CD4(t=2)* in standard analyses: on causal pathway!

Outline

1. What types of questions can these methods help us answer?
2. Why are more complex statistical methods needed?
   - Challenge of time-varying confounders
3. Overview of available estimators
   - Inverse Probability of Treatment Weighted (IPTW)
   - Targeted Maximum Likelihood (TMLE)
4. Introduction to R software: Joshua Schwab
   - Sample data, code, and simulation results
5. Conclusion

Overview: IPTW

- Likelihood of the Observed Data

P(Treatment & Censoring | Past) P(Other covariates | Past)

"Treatment Mechanism"

- IPTW

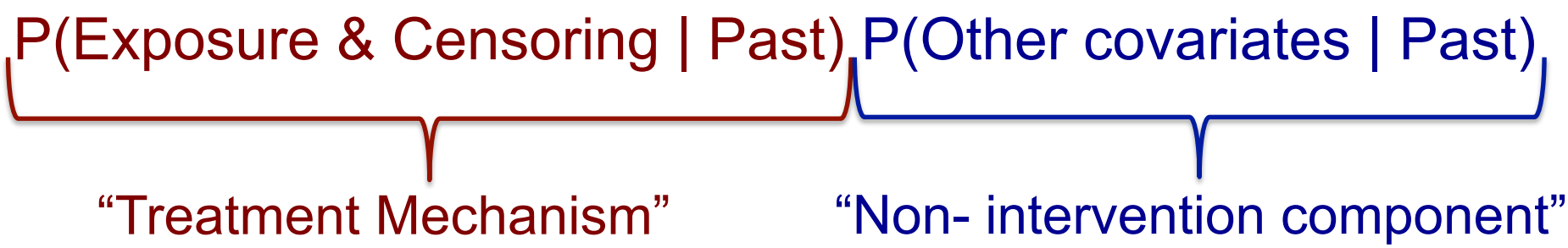
1. Estimate treatment mechanism
   - How treatment (and/or censoring) depend on the observed past
2. Use this estimate to reweight data

IPTW Properties

- Relies on doing a good job estimating the weights
  - If treatment mechanism estimated using a misspecified model -> Bias
- Subject to bias and high variance with moderate to strong confounding
  - When certain treatment or exposure levels of interest are rare/absent for some patient histories

Overview: TMLE

- Likelihood of the Observed Data



- TMLE

1. Estimate non-intervention component
2. Update this estimate to remove bias for quantity you care about
   - Update step uses estimate of treatment mechanism

TMLE Properties

- Minimize bias due to model misspecification
  - Double Robust - Two chances to get it "right"
    - Consistent if you estimate either component correctly
- Maximize precision of effect estimates
  - Efficient (minimal asymptotic variance in semi-parametric model) if you get both right
- In practice (finite samples)
  - May reduce bias and variance compared to IPTW

Outline

1. What types of questions can these methods help us answer?
2. Why are more complex statistical methods needed?
   - Challenge of time-varying confounders
3. Overview of available estimators
   - Inverse Probability of Treatment Weighted (IPTW)
   - Targeted Maximum Likelihood (TMLE)
4. Introduction to R software: Joshua Schwab
   - Sample data, code, and simulation results
5. Conclusion

Example 1: Point treatment

Simulated Data:

id	male	age	CD4_1	A1	Y1
1	1	27	345	0	0
2	1	33	78	1	1
3	0	25	264	0	0
4	0	18	212	1	0
5	0	34	363	0	0
6	1	31	414	0	0

…

CD4_1: CD4 count at first visit failed

A1: Indicator switched to second line ART at first visit failed

Y1: Death by the end of first quarter after failure

Example 1: Point treatment

What proportion of patients who failed would have died within one quarter if all had been switched immediately?

Call:

```
ltmle(data, Anodes="A1”, Ynodes="Y1", abar=1)
```

Output:

```
TMLE Estimate:  0.0294344
```

Example 1: Point treatment

- Additional outputs: IPTW and TMLE Estimates (95% CI) for proportion died under immediate switch ("est1") and proportion died under no immediate switch ("est0")

```
> est1 <- ltmle(data, Anodes="A1", Ynodes="Y1", abar=1)
> est0 <- ltmle(data, Anodes="A1", Ynodes="Y1", abar=0)

> summary(est1, estimator="tmle")
    Parameter Estimate:  0.029434
    Estimated Variance:  0.000170
    95% Conf Interval:  (0.003832, 0.055036)

> summary(est1, estimator="iptw")
    Parameter Estimate:  0.023879
    Estimated Variance:  0.000147
    95% Conf Interval:  (5.0134e-05, 0.047708)
```

Example 1: Point treatment

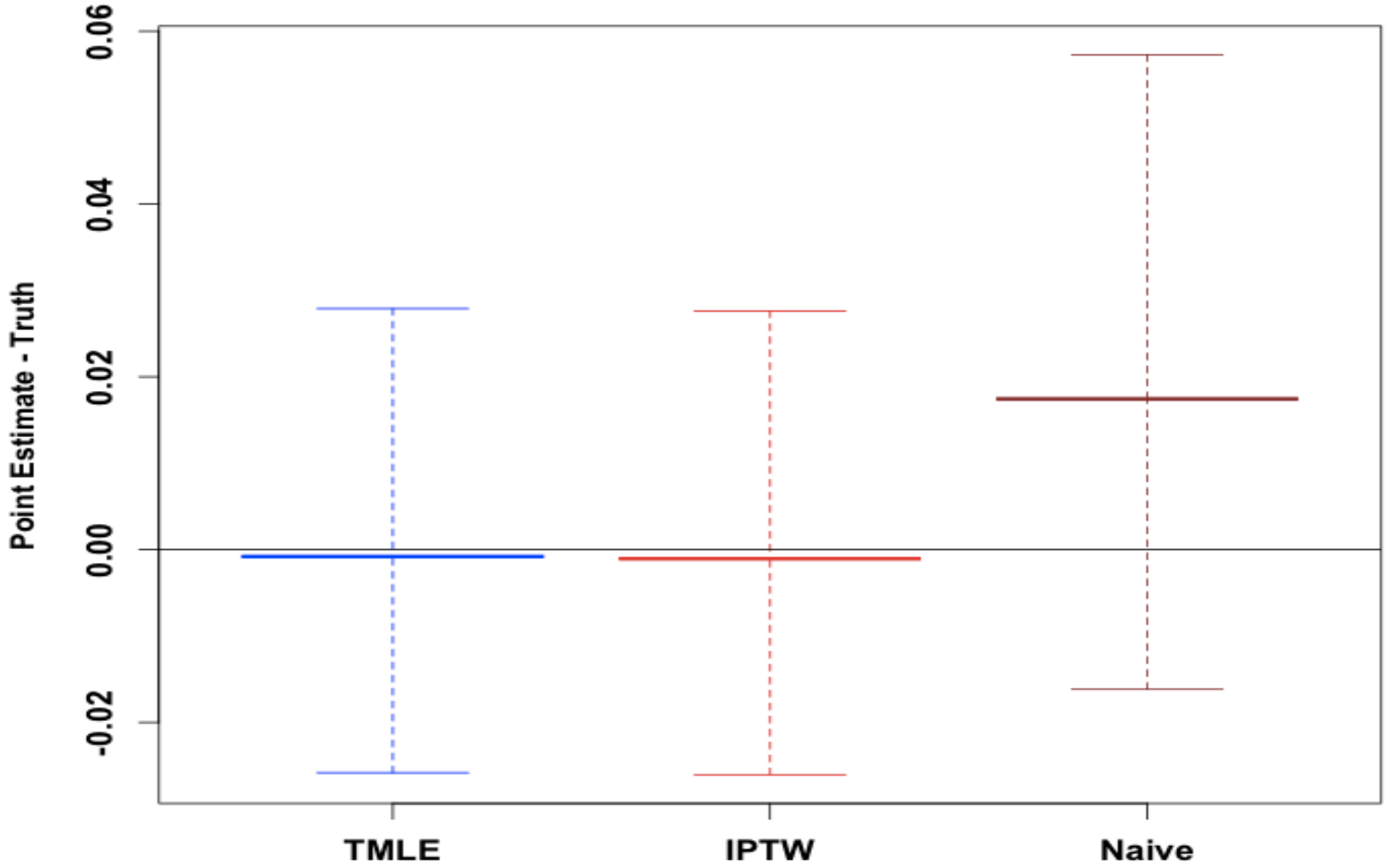
- Additional outputs: Estimates and 95% CI for risk difference (additive effect)
  - Difference in the proportion who would have died if all had been switched immediately versus if none had been switched immediately

```
> summary(est1, est0,  estimator="tmle")
   Additive Effect:
   Parameter Estimate:  -0.11425
   Estimated Variance:  0.0017416
               p-value:  0.0061865
    95% Conf Interval: (-0.19605, -0.032457)
```

- Also gives: relative risk, and odds ratio

Example 1: Point treatment

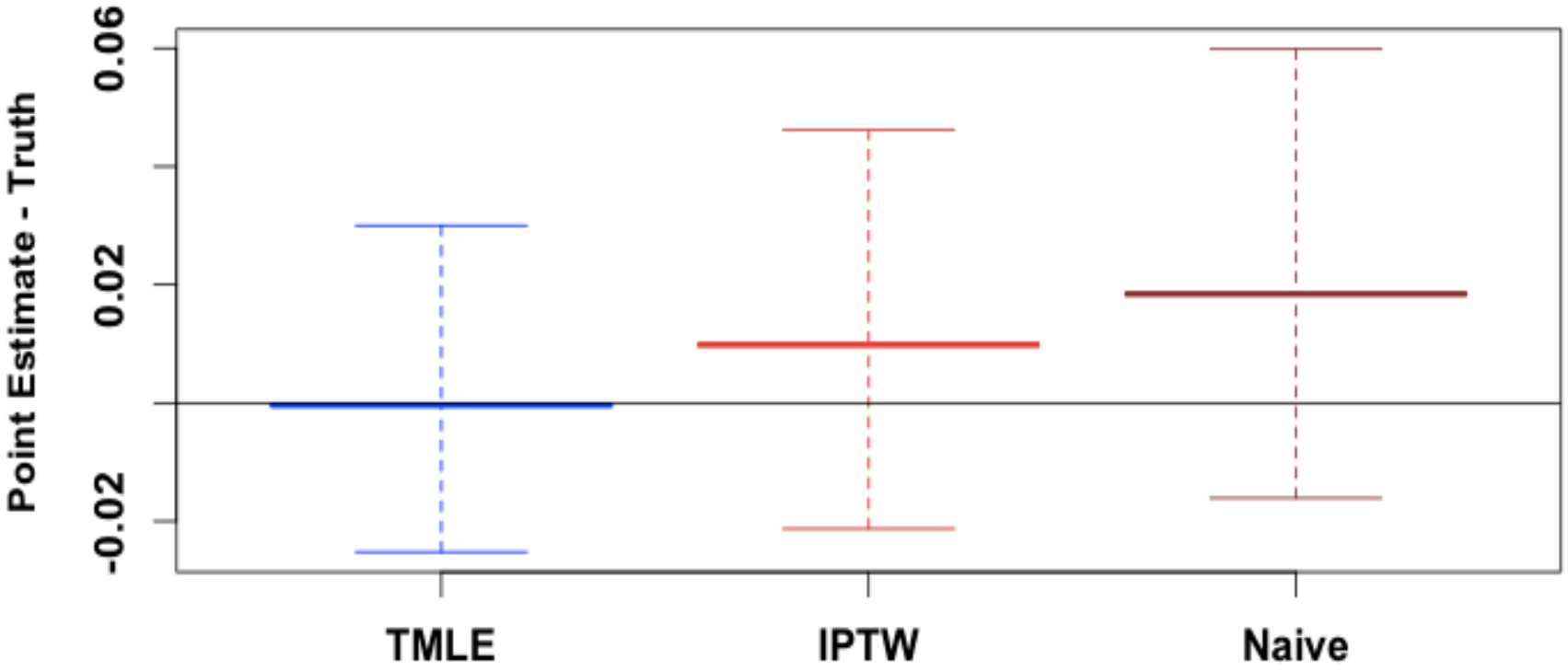
Results from simulated data:



Example 1: Point treatment

What if the treatment mechanism is estimated using a misspecified model?

```
ltmle(data, Anodes="A1", Ynodes="Y1", abar=1,
    gform="A1 ~ male + age")
```



Example 2: Longitudinal treatment

Simulated Data:

id	male	age	CD4_1	A1	Y1	CD4_2	A2	Y2	CD4_3	A3	Y3
1	0	37	101	1	0	152	1	0	228	1	0
2	0	31	301	0	0	297	0	0	302	0	0
3	1	31	323	0	0	351	1	0	395	1	0
4	0	26	29	1	0	81	1	0	124	1	1
5	0	23	280	0	0	315	0	0	321	0	0
6	1	43	237	1	1	NA	NA	1	NA	NA	1
...											

CD4_t: CD4 count at start of quarter *t*

At: switched to second line at or before start of quarter *t*

Yt: death by end of quarter *t*

Example 2: Longitudinal treatment

What proportion of patients would have died within 3 quarters had none switched?

Call:

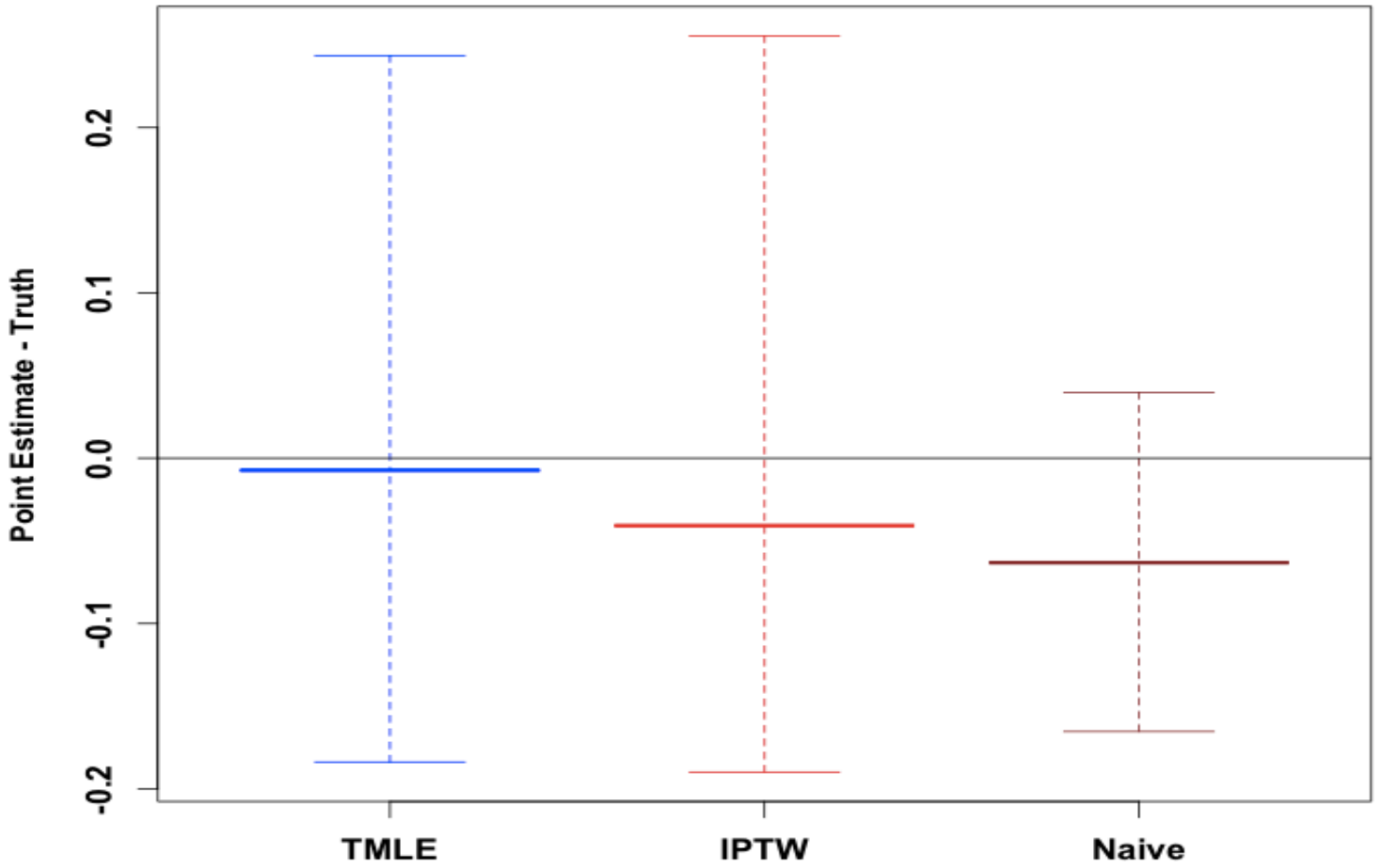
```
ltmle(data, Anodes=c("A1", "A2", "A3”),
   Ynodes=c("Y1", "Y2", "Y3"),
   abar=c(0, 0, 0))
```

Output:

```
TMLE Estimate:  0.1546657
```

Example 2: Longitudinal treatment

Results from simulated data:



Example 3: Marginal Structural Model

- How does counterfactual probability of dying within 3 quarters vary as function of switch time?
  - MSM: logit(E($Y3_{switch}$))=β_0 + $\beta_1$switch time

Call:

```
ltmleMSM(data, Anodes=c("A1", "A2","A3"),
   Ynodes=c("Y1", "Y2", "Y3"),
   working.msm="Y ~ switch.time",
   regimens=regimens,
   summary.measures=summary.measures)
```

Output:

```
(Intercept) switch.time
  -2.733044    0.418045
```

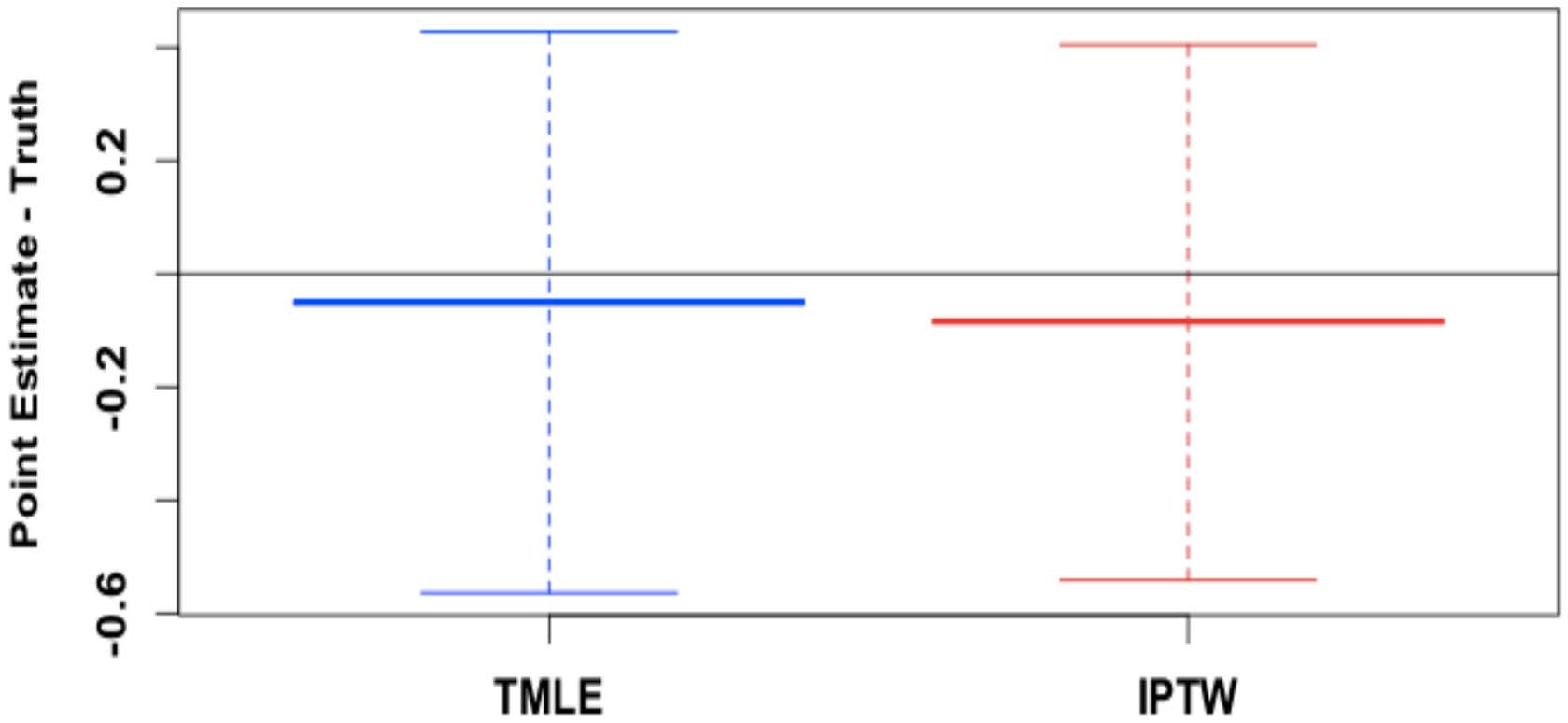
β_0 estimate

β_1 estimate:

OR per additional quarter until switch = exp(0.418) = 1.52

Example 3: Marginal Structural Model

Results from simulated data:



Take home points

- Applying these methods to answer real questions with real data

1. Can give you better answers
2. Is feasible
   - Software is available
3. Mastery not possible in 45 minutes
   - We are showing simple calls for simplified data
     - Using default for many options
   - Software has substantial additional functionality

- Look for upcoming training workshop!

Select References and Resources

- Marginal structural models, time dependent confounding, IPTW:
  - Robins, Hernan, Brumback. Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology* 2000; 11(5) 550-560
- Longitudinal TMLE
  - van der Laan, Gruber. Targeted Minimum Loss Based Estimation of Causal Effects of Multiple Time Point Interventions" *The International Journal of Biostatistics* 2012; 8.1
- Websites
  - Mark's: www.stat.berkeley.edu/~laan
    - R code at: www.stat.berkeley.edu/~laan/Software
  - Maya's: works.bepress.com/maya_petersen

Thanks to our funders!

- This work was supported by NIH
  - Grant # U01 AI069924 (NIAID, NICHD, NCI)
    - PIs: Egger and Davies
  - Grant # R01 AI074345-06 (NIAID)
    - PI: van der Laan
- Maya Petersen is supported by Doris Duke Charitable Foundation
  - Grant #:2011042

R package: ltmle

- Casual effect estimation with multiple intervention nodes
  - Intervention-specific mean under longitudinal static and dynamic interventions
  - Static and dynamic marginal structural models
- General longitudinal data structures
  - Repeated measures outcomes
  - Right censoring
- Estimators
  - IPTW
  - Non-targeted MLE
  - TMLE (two algorithms for MSM)
- Options include nuisance parameter estimation via glm regression formulas or calling SuperLearner()
- Available on CRAN April 2013