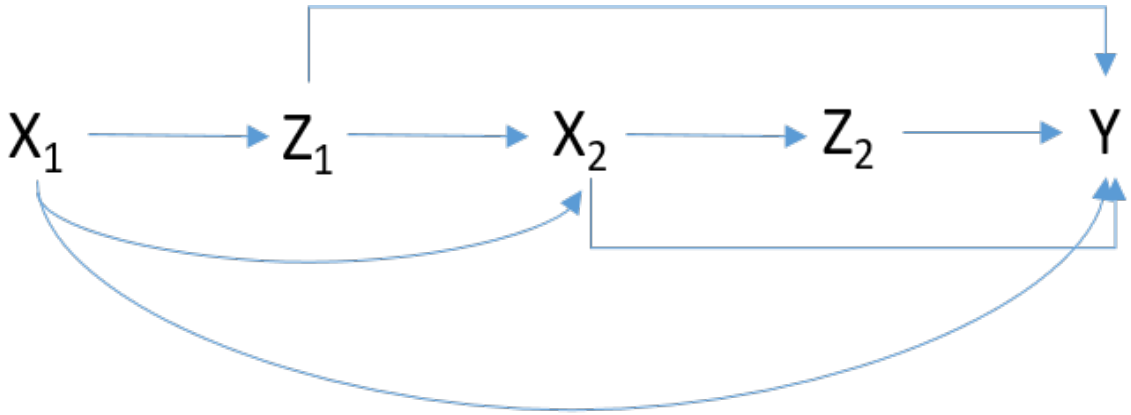
Robust Method for Causal Inference: Penalized Spline of Propensity Methods for Treatment Comparison (PENCOMP)

Tingting Zhou [1]

[1]University of Michigan

April 2, 2018

Objectives



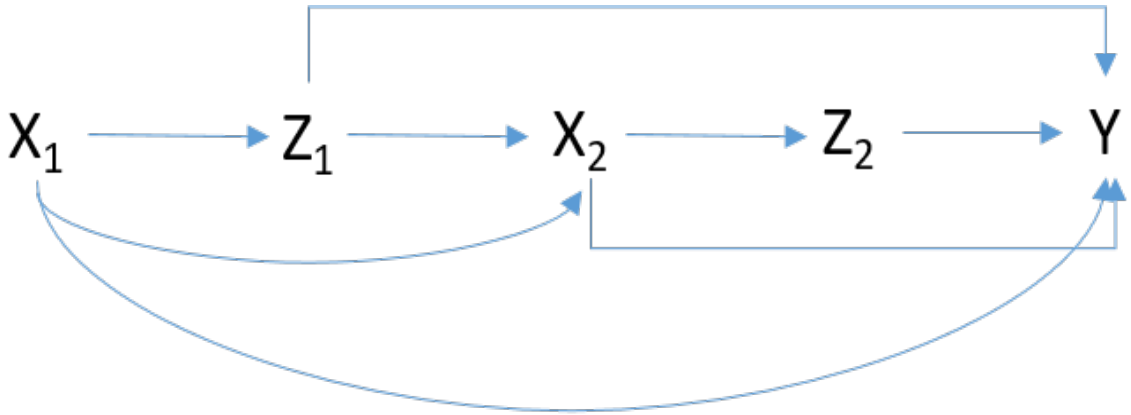
Confounding by indication in two-time point treatment

- Standard regression methods fail when there are time dependent confounders
  - Mediators of earlier treatments and confounders of future treatments.

Outline

- Introduction
  - Setup, Notations and Assumptions
  - Existing approaches: IPTW, AIPTW, G-Computation
- Our proposed method (PENCOMP) to deal with time dependent confounders
  - Relation to missing data
  - Estimation and Inference
- Simulation Studies
- Application to an AIDS dataset
- Discussion and Extensions

Introduction



Confounding by indication in two-time point treatment

- Treatments are applied over time and treatment decisions driven by intermediate outcomes
- Intermediate outcomes are simultaneously mediators and confounders

Setup and Notations

- Assume n subjects and $T+1$ discrete time points by $t = 1, \cdots, T+1$
- $\bar{X}_t$ observed covariate history up to and including time point t
- $\bar{Z}_t$ observed treatment history up to and including time point t
- Y final outcome of interest measured at time point $T+1$
- $Y^{\bar{Z}_T}$ and $X_t^{\bar{Z}_{t-1}}$ denote the potential final and intermediate outcomes
  - Observed if assigned to treatment history $\bar{Z}_T$ and $\bar{Z}_{t-1}$ respectively
- Focus for illustration on two-time point treatments
  - Assume $Z_t \in (0, 1)$ at each t
  - $(Y^{00}, Y^{01}, Y^{10}, Y^{11})$ and (X_2^1, X_2^0)

Assumptions

- The stable unit treatment value assumption (SUTVA) (Angrist, Imbens and Rubin 1996)
  - a) Consistency
  - b) No interference between subjects
- The positivity assumption: $0 < \Pr(Z_t = z_t | \bar{X}_{t-1}, \bar{Z}_{t-1}) < 1$
- Sequential ignorable treatment assumption states that

$$(Y^{\bar{Z}_T}, X_{t+1}^{\bar{Z}_t}) \perp\!\!\!\perp Z_t | (\bar{Z}_{t-1}, \bar{X}_t)$$

  - At each time t, treatment assignment Z_t is as if randomized conditional on all the past history

IPTW and AIPTW

- Inverse Probability of Treatment Weighting (IPTW)
  - Weights subjects by the inverse of the probability of receiving the observed treatment sequence
  - Consistent if the propensity model is correct
- Augmented Inverse Probability of Treatment Weighting (AIPTW)
  - Doubly robust
    - Consistent if the propensity models are correct
    - Consistent if all the conditional distributions relating outcomes to covariates are correctly specified (Scharfstein, Rotnitzky, and Robins, 1999)

G Computation

- Fix treatment regime $\bar{z}_T = (z_1, \cdots, z_T)$

$$E(Y^{\bar{z}_T}) = \sum_{x_1, \cdots, x_T} E(Y|\bar{X}_T, \bar{Z}_T = \bar{z}_T) \times P(X_1) \times P(X_2|X_1, Z_1 = z_1)$$
$$\cdots \times P(X_T|\bar{X}_{T-1}, \bar{Z}_{T-1} = \bar{z}_{T-1})$$

- Simulate potential outcomes under each treatment sequence based on conditional distribution of covariates and outcomes estimated from the data (Robins, 1987)
- Consistent if all conditional distributions are correctly specified

PENCOMP

- An alternative approach: a multiple-imputation based approach called Penalized Spline of Propensity Methods for Treatment Comparison (PENCOMP)
  - Only one potential outcome is observed for each subject
  - The idea behind PENCOMP: impute the missing potential outcomes
  - Use the propensity as a predictor
  - Conceptually simple: relies on regression models for prediction of potential outcomes

PENCOMP: Two-Time Point Treatment

Observed and missing intermediate and final outcomes for treatment at two-time point treatment

Subjects	X_1	Z_1	X_2^0	X_2^1	Z_2	Y^{00}	Y^{01}	Y^{10}	Y^{11}
1		0		?	0		?	?	?
$\cdots$		0		?	0		?	?	?
n_{00}		0		?	0		?	?	?
$n_{00}+1$		0		?	1	?		?	?
$\cdots$		0		?	1	?		?	?
$n_0 = n_{00}+n_{01}$		0		?	1	?		?	?
n_0+1		1	?		0	?	?		?
$\cdots$		1	?		0	?	?		?
n_0+n_{10}		1	?		0	?	?		?
$n_0+n_{10}+1$		1	?		1	?	?	?	
$\cdots$		1	?		1	?	?	?	
$n = n_0+n_{10}+n_{11}$		1	?		1	?	?	?	

Step 1: Imputing Missing Potential Outcomes X_2

How to build a robust imputation model for intermediate outcome X_2?

Result 1

$$(X_2^1, X_2^0) \perp\!\!\!\perp Z_1 | X_1 \Longrightarrow (X_2^1, X_2^0) \perp\!\!\!\perp Z_1 | P_{z_1}(X_1)$$

- where $P_{z_1}(X_1) = Pr(Z_1 = z_1 | X_1)$ is the probability of being assigned z_1 (Rosenbaum and Rubin 1983)

$$\begin{aligned} E(X_2^1 - X_2^0) &= E[E(X_2 | X_1, Z_1 = 1)] - E[E(X_2 | X_1, Z_1 = 0)] \\ &= E[E(X_2 | P_{z_1}(X_1), Z_1 = 1)] - E[E(X_2 | P_{z_1}(X_1), Z_1 = 0)] \end{aligned}$$

- Conditioning on the covariates X_1 or on the propensity scores

Imputation Model for X_2

- PENCOMP imputes the missing potential outcomes $\hat{X}_2^{z_1}$ from the mean model

$$E(X_2^{z_1}|X_1, Z_1 = z_1, \theta_{z_1}, \beta_{z_1}) = s(\hat{P}^*_{z_1}|\theta_{z_1}) + g_{z_1}(X_1; \beta_{z_1})$$

  - $\hat{P}^*_{z_1}$ =log$[\hat{P}_{z_1}(X_1)/(1 - \hat{P}_{z_1}(X_1))]$
  - $g_{z_1}()$ denotes parametric function; improves efficiency by including predictors of outcome

Penalized Spline

- $s(\hat{P}^*_{z_1}|\theta_{z_1}) = \theta_{1_{z_1}}\hat{P}^*_{z_1} + \sum_{k=1}^{K}\theta_{1k_{z_1}}(\hat{P}^*_{z_1} - K_k)_+$, with truncated linear basis
  - $K_1, \cdots, K_K$ are fixed knots
  - $(\hat{P}^*_{z_1} - K_k)_+ = (\hat{P}^*_{z_1} - K_k)$ if $\hat{P}^*_{z_1} > K_k$; and $= 0$ if $\hat{P}^*_{z_1} \leq K_k$.

Reformulation as a Mixed Model

$$E(X_2^{z_1}|X_1, Z_1 = z_1, \theta_{z_1}, \beta_{z_1}) = s(\hat{P}^*_{z_1}|\theta_{z_1}) + g_{z_1}(X_1; \beta_{z_1}) \quad (1)$$

- Reformulate (1) as a mixed model (Wand, 2003)

$$E(X_2^{z_1}|X_1, Z_1 = z_1, \theta_{z_1}, \beta_{z_1}) = C_1\beta_{z_1} + C_2\theta_{z_1}$$
$$\theta_{z_1} \sim N(0, \sigma^2_{\theta_{z_1}} I)$$

- $C_1 = [1, X_1, \hat{P}^*_{z_1}]$, $C_2 = [(\hat{P}^*_{z_1} - K_1)_+, \cdots, (\hat{P}^*_{z_1} - K_K)_+]$
- Fixed effects: $\beta_{z_1} = (\beta_{0_{z_1}}, \beta_{1_{z_1}}, \theta_{1_{z_1}})$
- Random basis coefficients: $\theta_{z_1} = (\theta_{11_{z_1}}, \cdots, \theta_{1K_{z_1}})$

Estimation

- Parameters estimated using REML
- Easily fitted in statistical software, such as PROC MIXED in SAS or lme in R
- The fitted values of X_2 are $\hat{X}_2 = C(C^T C + \hat{\lambda} D)^{-1} C^T X_2$, where $\hat{\lambda} = \hat{\sigma}^2_\epsilon / \hat{\sigma}^2_\theta$ is the REML estimator of λ and, $C = [C_1, C_2]$

$$D = \begin{pmatrix} 0_{3\times 3} & 0 \\ 0 & I_{K\times K} \end{pmatrix}$$

Consistency

- The marginal mean of the imputed values $\hat{X}_2^{z_1}$ from our imputation model is consistent (Little and An, 2004; Zhang and Little, 2009) if:
  - a) $g_{z_1}(X_1; \beta_{z_1})$ correctly specified
  - b) $P_{z_1}(X_1)$ and $s(\hat{P}^*_{z_1}|\theta_{z_1})$ correctly specified. The latter assumption is relatively weak, since the spline does not impose strong parametric assumptions

PENCOMP: Two-Time Point Treatment

After imputing intermediate outcomes X_2^0 and X_2^1 with draws

Subjects	X_1	Z_1	X_2^0	X_2^1	Z_2	Y^{00}	Y^{01}	Y^{10}	Y^{11}
1		0			0		?	?	?
$\cdots$		0			0		?	?	?
n_{00}		0			0		?	?	?
$n_{00}+1$		0			1	?		?	?
$\cdots$		0			1	?		?	?
$n_0 = n_{00}+n_{01}$		0			1	?		?	?
n_0+1		1			0	?	?		?
$\cdots$		1			0	?	?		?
n_0+n_{10}		1			0	?	?		?
$n_0+n_{10}+1$		1			1	?	?	?	
$\cdots$		1			1	?	?	?	
$n = n_0+n_{10}+n_{11}$		1			1	?	?	?	

Step 2: Imputing Missing Potential Outcomes Y

How to build a robust imputation model for final outcome Y?

Result 2

$$Y^{\bar{z}_2} \perp\!\!\!\perp I(\bar{Z}_2 = \bar{z}_2) \Big| P_{\bar{z}_2}$$

- where $P_{\bar{z}_2} = \prod_{k=1}^{2} P(Z_k = z_k | \bar{Z}_{k-1} = \bar{z}_{k-1}, \bar{X}_k)$ is the propensity of being assigned treatment regime $\bar{z}_2$
- Treatment regime $\bar{Z}_2$ is independent of potential outcomes $Y^{\bar{z}_2}$ given the propensity of receiving that treatment regime $\bar{Z}_2$, for all $t = 1, 2$

Result 2

$$
\begin{aligned}
E\left(Y^{11}\right) &= E\left[E(Y^{11}|\bar{X}_2)\right] \\
&= E\left[E(Y|\bar{X}_2, Z_1 = 1, Z_2 = 1)\right] \\
&= E\left[E(Y|P_{\bar{z}_2=(11)}, Z_1 = 1, Z_2 = 1)\right]
\end{aligned}
$$

- Conditioning on covariates or the propensity scores
- Similar derivations hold by Y^{10}, Y^{01}, and Y^{00}

Imputation Model for $Y^{\bar{Z}_2}$

$$E(Y^{\bar{z}_2}|\bar{X}_2, Z_1 = z_1, Z_2 = z_2, \theta_{\bar{z}_2}, \beta_{\bar{z}_2})$$
$$= s(\hat{P}^*_{\bar{z}_2}|\theta_{\bar{z}_2}) + g_{\bar{z}_2}(\bar{X}_2, \bar{Z}_2; \beta_{\bar{z}_2})$$

- $\hat{P}^*_{\bar{z}_2}$ =log$[\hat{P}_{\bar{z}_2}/(1 - \hat{P}_{\bar{z}_2})]$
- $s(\hat{P}^*_{\bar{z}_2}|\theta_{\bar{z}_2})$ is a penalized spline with fixed knots as described before
- $g_{\bar{z}_2}()$ improves efficiency by including predictors of the outcome

Consistency

- The marginal mean from the imputation model for Y is consistent if:
  - 1) $g_{\bar{z}_t}$: All the prediction models for the intermediate and final outcomes are correctly specified
  - OR 2) The propensity models and the relationship between X_{t+1} and $\hat{P}^*_{\bar{z}_t}$ are correctly specified at each time point. Again this assumption can be weakened by assuming only a smooth functional form, such as a penalized spline as in PENCOMP

PENCOMP: Two-Time Point Treatment

After imputing all the missing potential outcomes with draws

Subjects	X_1	Z_1	X_2^0	X_2^1	Z_2	Y^{00}	Y^{01}	Y^{10}	Y^{11}
1		0			0				
$\cdots$		0			0				
n_{00}		0			0				
$n_{00}+1$		0			1				
$\cdots$		0			1				
$n_0 = n_{00}+n_{01}$		0			1				
n_0+1		1			0				
$\cdots$		1			0				
n_0+n_{10}		1			0				
$n_0+n_{10}+1$		1			1				
$\cdots$		1			1				
$n = n_0+n_{10}+n_{11}$		1			1				

PENCOMP: Inference

- Combine the imputed and observed outcomes using Rubin's Combining Rules:
- Estimate of the causal effect $\Delta_{jk} = E(Y^{jk} - Y^{00})$ is then $\bar{\Delta}_{jkD} = \sum_{d=1}^{D} \hat{\Delta}_{jk}^{(d)}$ over D complete data sets.
- Estimate of the variance of $\bar{\Delta}_{jkD}$ is $T_{\mathrm{D}} = \bar{W}_{jkD} + (1 + 1/D)B_{jkD}$
  - Between imputation variance: $\bar{W}_{jkD} = \sum_{d=1}^{D} W_{jk}^{(d)}/D$
  - Within imputation variance: $B_{jkD} = \sum_{d=1}^{D} \left(\hat{\Delta}_{jk}^{(d)} - \bar{\Delta}_{jkD}\right)^2/(D-1)$ (Rubin 1987)

Simulation Setup

- Each simulated data set contains X_{1a}, X_{1b}, Z_1, X_{2a}, X_{2b}, Z_2, and Y
- Two baseline covariates: $X_{1a} \sim N(0.2, 1)$ and $X_{1b} \sim N(0.2, 1)$
- The intermediate outcomes X_{2a} and X_{2b}, and final outcome Y are normally distributed with residual variance of 1
- Interested in inference about Δ_{11}, Δ_{10}, and Δ_{01}, where $\Delta_{\bar{z}_2} = E(Y^{\bar{Z}_2} - Y^{00})$
- The linear outcome model $(\Delta_{11}, \Delta_{10}, \Delta_{01}) = (22.35, 11.17, 10.45)$
- The nonlinear outcome model $(\Delta_{11}, \Delta_{10}, \Delta_{01}) = (25.31, 12.69, 10.57)$

Simulation Setup 2

	Intercept	X_{1a}	X_{1b}	X_{2a}	X_{2b}	$X_{2a}X_{2b}$
Treatment Assignment at $t = 1$						
	-0.01	γ_{11}	-0.3			
Treatment Assignment at $t = 2$						
$Z_1 = 0$	-0.01	$-\gamma_{21}$	0.1	γ_{21}	-0.1	
$Z_1 = 1$	-0.01	$-\gamma_{21} - \gamma_{22}$	0.1$-\gamma_{24}$	$\gamma_{21} + \gamma_{22}$	-0.1$+\gamma_{24}$	
$X_{2a}^{z_1=0}$		1	0.5			
$X_{2a}^{z_1=1}$	0.5	1.5	0.5			
$X_{2b}^{z_1=0}$		1	0.3			
$X_{2b}^{z_1=1}$		1	0.4			
Y_{11}	25	2	1.5	2	1.5	1.6*
Y_{10}	15	2	1.5	1	1	1*
Y_{01}	15	1	1	2	1.5	0.8*
Y_{00}	15	1	1	1	1	0.7*

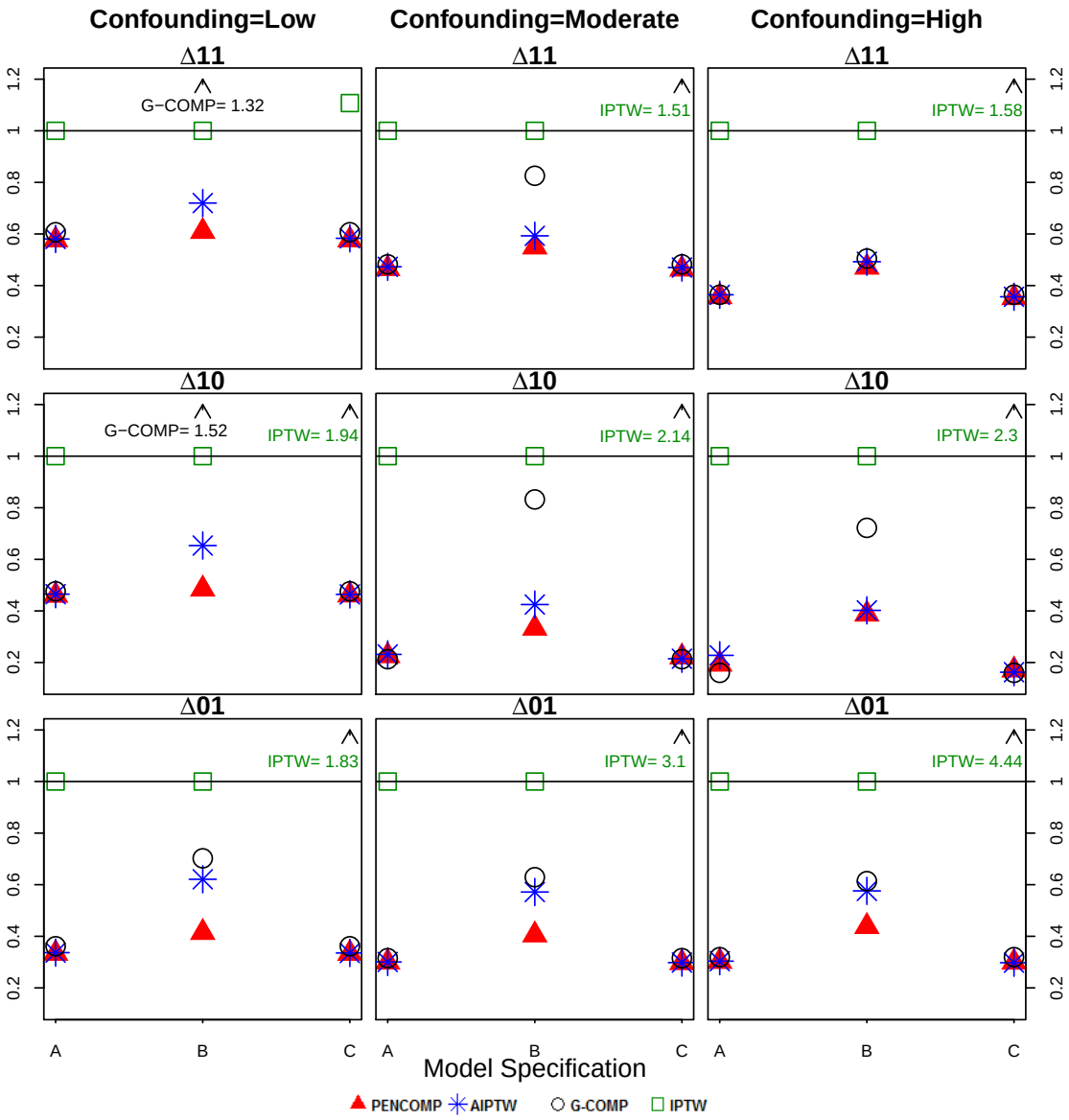
$(\gamma_{11}, \gamma_{21}, \gamma_{22}, \gamma_{24})$= $(-0.5, -0.1, 0.2, 0.2)$ for high, $(-0.8, -0.1, 0.6, 0.6)$ for moderate, and $(-0.8, -0.5, 1.1, 1.1)$ for low confounding.

Nonlinear outcome model (*) includes interactions $X_{2a}X_{2b}$

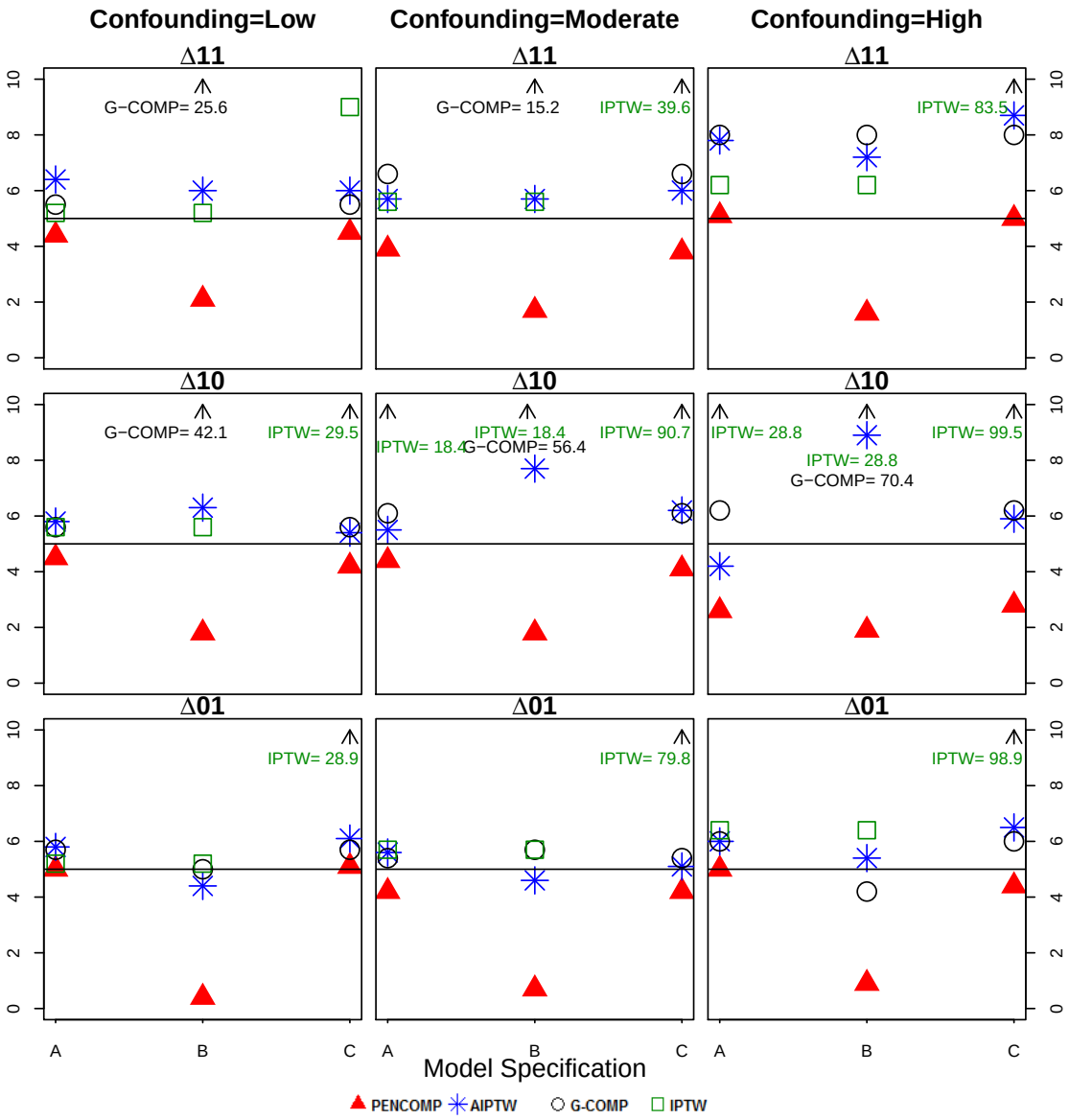
Simulation Setup 3

- Three Cases considered:
  - A) Both propensity score and prediction models correctly specified
  - B) The prediction model misspecified
  - C) The propensity score model misspecified
- The propensity score model was misspecified as $logit(P(Z_2 = 1|\bar{X}_2, Z_1, \lambda)) = \lambda_0 + \lambda_1 X_{1a} + \lambda_2 X_{2a} + \lambda_3 X_{1b}$
- The prediction model was misspecified by dropping terms involving X_{2a} and X_{2b}
- Logistic regression models were assumed to model treatment assignment
- Assess the finite sample performance of PENCOMP, compared with g-computation, IPTW and AIPTW
- Here show results for sample size of 500 based on 1,000 simulated data sets

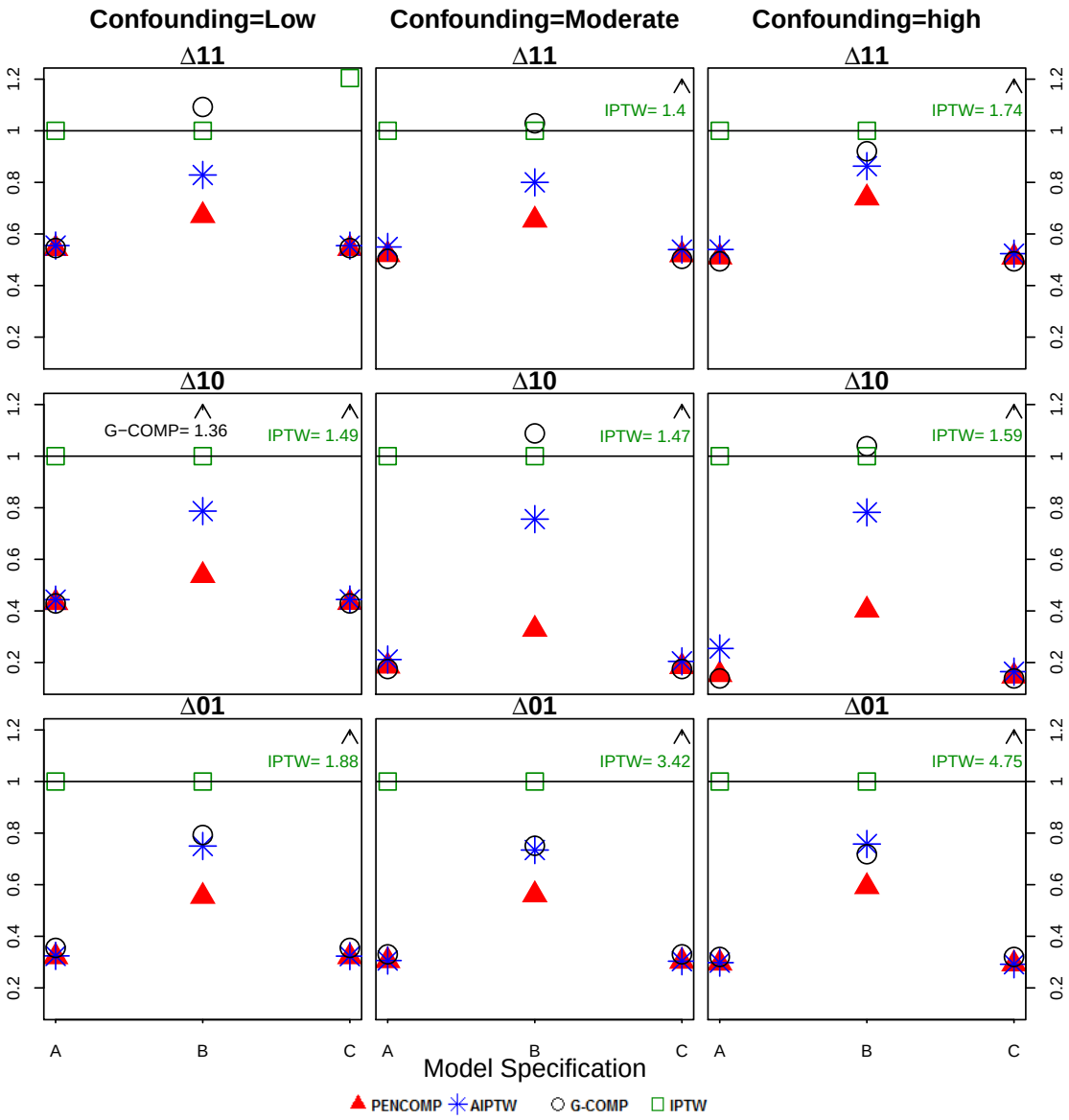
Relative RMSE in Linear Outcome



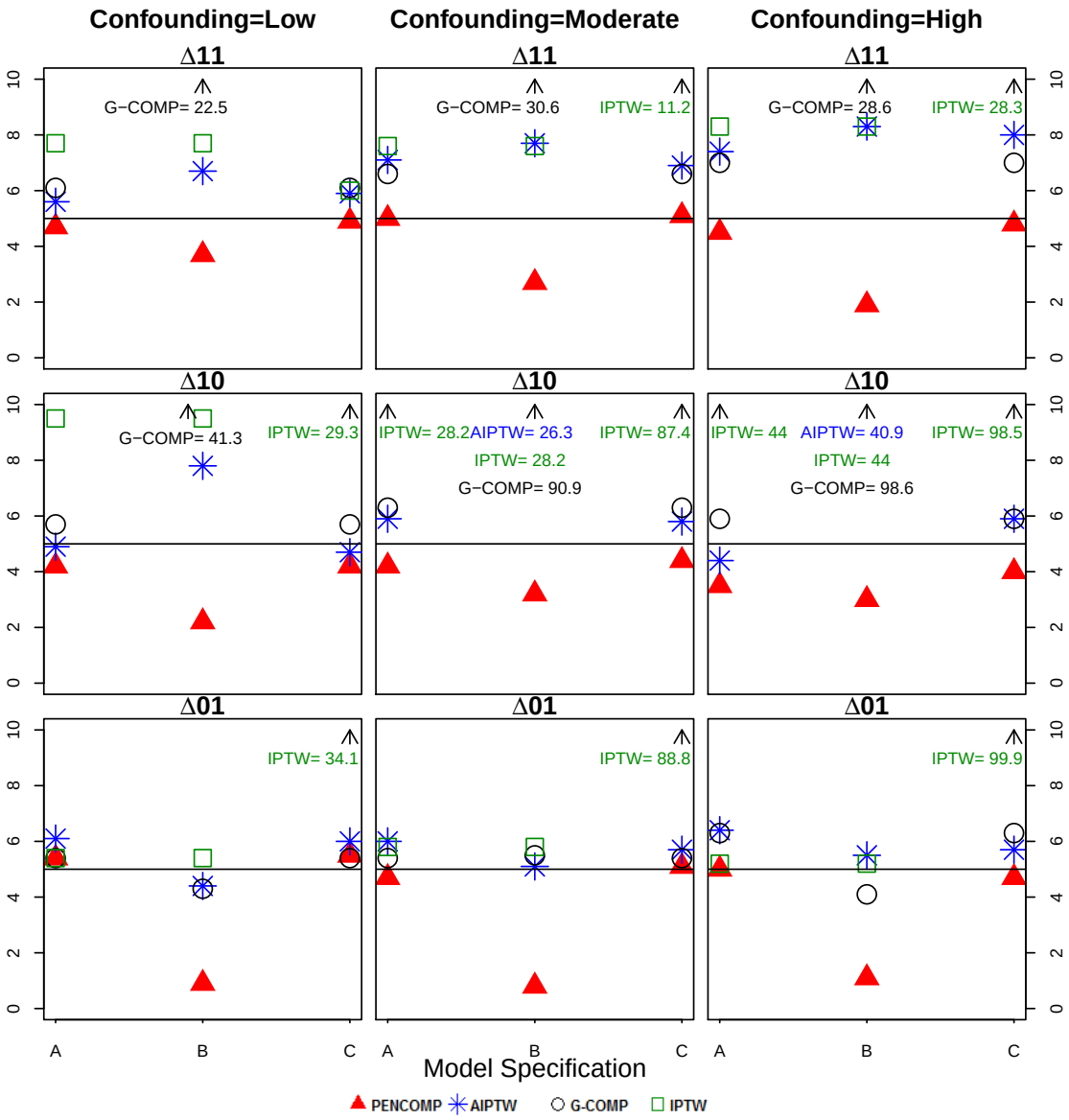
95% Noncoverage Rate in Linear Outcome



Relative RMSE in Nonlinear Outcome



95% Noncoverage Rate in Nonlinear Outcome



Application: AIDS Cohort Study

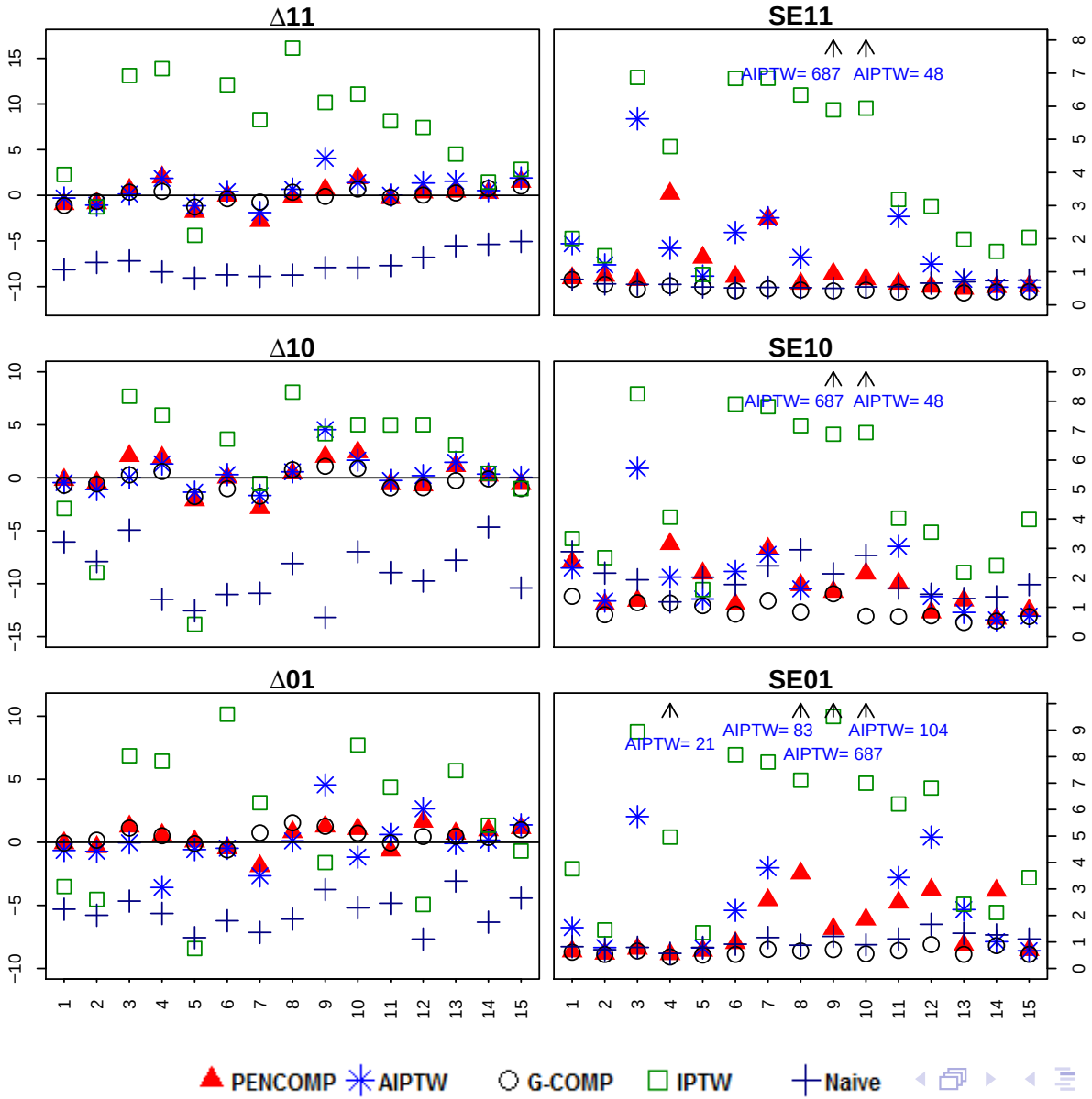
- We applied our method to the Multicenter AIDS Cohort study (MACS) to analyze the effect of antiretroviral treatment on CD4 counts for HIV+
- CD4 count is an intermediate outcome of past treatments and confounds the next treatment
- Restrict our analyses to the period between visit 7 and 21, when zidovudine was approved and available for use and before the advent of highly active antiretroviral therapy (HAART)
- We estimate the short-term (1 year) effects of using antiretroviral treatment for HIV+ subjects during this period
- For each of the three-visit moving windows $1, \cdots, 15$

Covariate Balance Checking

Balance of covariates between subjects with observed treatment sequence $(1, 1)$ and everybody else before and after adjusting for propensity scores. $**$ significant at 0.005 level. d is standardized mean difference.

	Before Adjusting		After Adjusting	
Covariate	d	T Stats	d	T stats
RBC	1.83	25.23**	0.016	0.22
CD4	1.11	15.28**	0.0048	0.067
WBC	0.59	8.11**	0.028	0.39
CD8	0.0012	0.017	0.032	0.44
PLATE	0.10	1.37	0.044	0.61
CD4 at $t = 2$	1.12	15.28**	0.017	0.23

Results


Δ11
SE11
AIPTW= 687
AIPTW= 48
Δ10
SE10
AIPTW= 687
AIPTW= 48
Δ01
SE01
AIPTW= 21
AIPTW= 83
AIPTW= 687
AIPTW= 104
PENCOMP
AIPTW
G-COMP
IPTW
Naive

Discussion

- The performance of PENCOMP is similar to that of AIPTW estimator when the confounding is low
- But PENCOMP tends to outperform AIPTW in RMSE, coverage probability and efficiency, when the confounding is moderate or high
- Our simulation studies suggest that our new straightforward method-PENCOMP is a viable alternative to IPTW and AIPTW estimators

Extensions

- A fully Bayesian version of PENCOMP: has attractive small-sample properties
- Address lack or limited overlap in the covariate distributions between treatment groups
  - Strategies for diagnosing and identifying common support region
  - Defining alternative estimands
- Model selection
  - Variables to include in the propensity and prediction models
  - Accounting for model uncertainty

Thank You!

References

- Eilers, P. H. C. and Marx, B. D. (1996). Flexible Smoothing with b-splines and Penalties. Statistical Science 11, 89-121.
- Little, R. J. A. and An, H. (2004). Robust Likelihood-Based Analysis of Multivariate Data with Missing Values. Statistica Sinica 14, 949-968.
- Ngo, L. and Wand, M. P. (2004). Smoothing with Mixed Model Software. Journal of Statistical Software 9, 1-54.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika 70, 41-55.

References

- Rubin, D. B. (1976). Inference and Missing Data. Biometrika 63, 581-592.
- Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: Wiley.
- Scharfstein, D., Rotnitzky, A., and Robins, J. M. (1999). Adjusting for Non-ignorable Drop-Out Using Semi-parametric Nonresponse Models. Journal of the American Statistical Association, 1096-1120.
- Yu, Z. and van der Laan, M. J. (2006). "Double Robust Estimation in Longitudinal Marginal Structural Models." Journal of Statistical Planning and Inference 136, 1061-1089.
- Zhang, G. and Little, R. J. A. (2009). Extensions of the Penalized Spline of Propensity Prediction Method of Imputation. Biometrics 65, 911-918.