

Methods: Mind the Gap
Webinar Series

Robust Inference for Stepped Wedge Designs



Presented by:
Jim Hughes, Ph.D.
University of Washington



Robust Inference for Stepped Wedge Designs

Jim Hughes

(Patrick Heagerty, Avi Kenny, Emily Voldal, Fan Xia)

NIH OFFICE OF DISEASE PREVENTION
METHODS: MIND THE GAP SERIES

MAY 17, 2022



Outline

1. Background
2. What can go wrong in SW analyses?
... and what to do about it.
3. Current research



Stepped Wedge Design

		Time			
		1	2	3	4
cluster	1	0	1	1	1
	2	0	0	1	1
	3	0	0	0	1

- N clusters, T time periods
- Interest is in the “intervention effect”



A Common Analysis Approach

- Mixed effect model

$$Y_{ijk} = \mu + \beta(t_{ij}) + \delta x_{ij} + a_i + b_{ij} + c_i x_{ij} + e_{ijk}$$

- Model for time – $\beta(t_{ij})$
- Intervention effect (δ) – assumed immediate and constant
- Random effects for
 - Cluster $a_i \sim N(0, \tau^2)$
 - Time $b_{ij} \sim N(0, \gamma^2)$
 - Intervention $c_i \sim N(0, \eta^2)$
- Additional structure on e_{ijk} possible (e.g. decaying autocorrelation)

Highly Model Dependent!



What can go wrong?

$$Y_{ijk} = \mu + \beta(t_{ij}) + \delta x_{ij} + a_i + b_{ij} + c_i x_{ij} + e_{ijk}$$

- Model for time – $\beta(t_{ij})$ – is wrong
 - May bias the estimate of δ
- Wrong variance structure
 - Wrong standard errors (conservative or anti-conservative) (Voldal et al 2022, Kasza et al 2019)
- Treatment effect is not constant
 - Can give misleading estimate of δ



What can go wrong?

$$Y_{ijk} = \mu + \beta(t_{ij}) + \delta x_{ij} + a_i + b_{ij} + c_i x_{ij} + e_{ijk}$$

- **Model for time – $\beta(t_{ij})$ – is wrong**
 - May bias the estimate of δ
- Wrong variance structure
 - Wrong standard errors (conservative or anti-conservative) (Voldal et al 2022, Kasza et al 2019)
- Treatment effect is not constant
 - Can give misleading estimate of δ



What can go wrong?

$$Y_{ijk} = \mu + \beta(t_{ij}) + \delta x_{ij} + a_i + b_{ij} + c_i x_{ij} + e_{ijk}$$

- Model for time – $\beta(t_{ij})$ – is wrong
 - May bias the estimate of δ
- **Wrong variance structure**
 - Wrong standard errors (conservative or anti-conservative) (Voldal et al 2022, Kasza et al 2019)
- Treatment effect is not constant
 - Can give misleading estimate of δ



Variance component mis-specification

- Voldal et al. (Stat Med 2022)

Case 1: True random treatment, fit random time

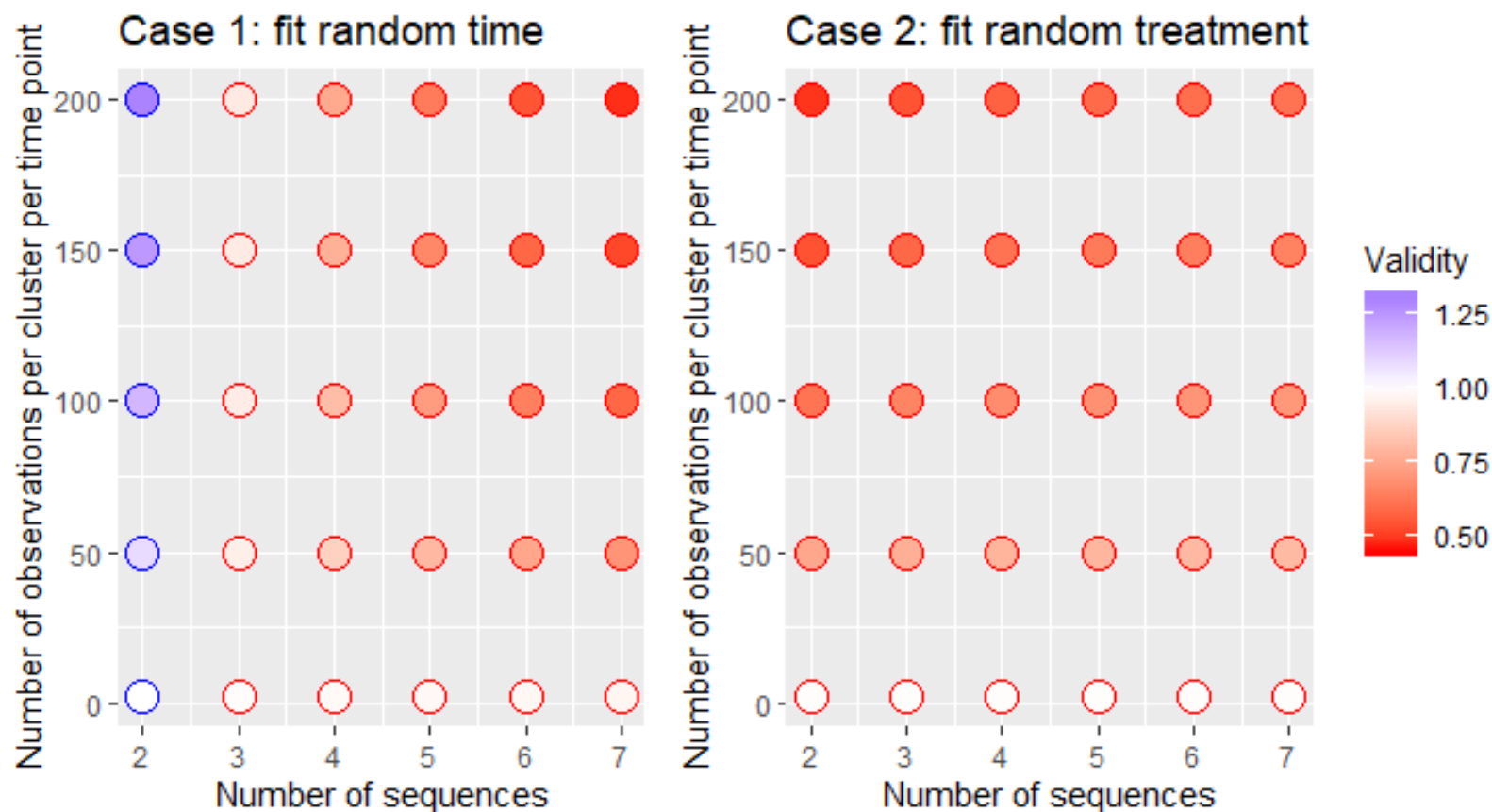
Case 2: True random time, fit random treatment

➤ Fixed effects unbiased, but variance of δ is wrong.

➤ Validity =
$$\frac{\text{Model-based variance from mis-specified model}}{\text{True mis-specified model variance}}$$



Validity





Variance component mis-specification

- Voldal et al. (Stat Med 2022)

Case 1: True random treatment, fit random time

Case 2: True random time, fit random treatment

- Fixed effects unbiased, but variance of δ is wrong.
- Recommendations
 1. Better to overfit, rather than underfit, variance components
 2. Use robust variance, when possible



What can go wrong?

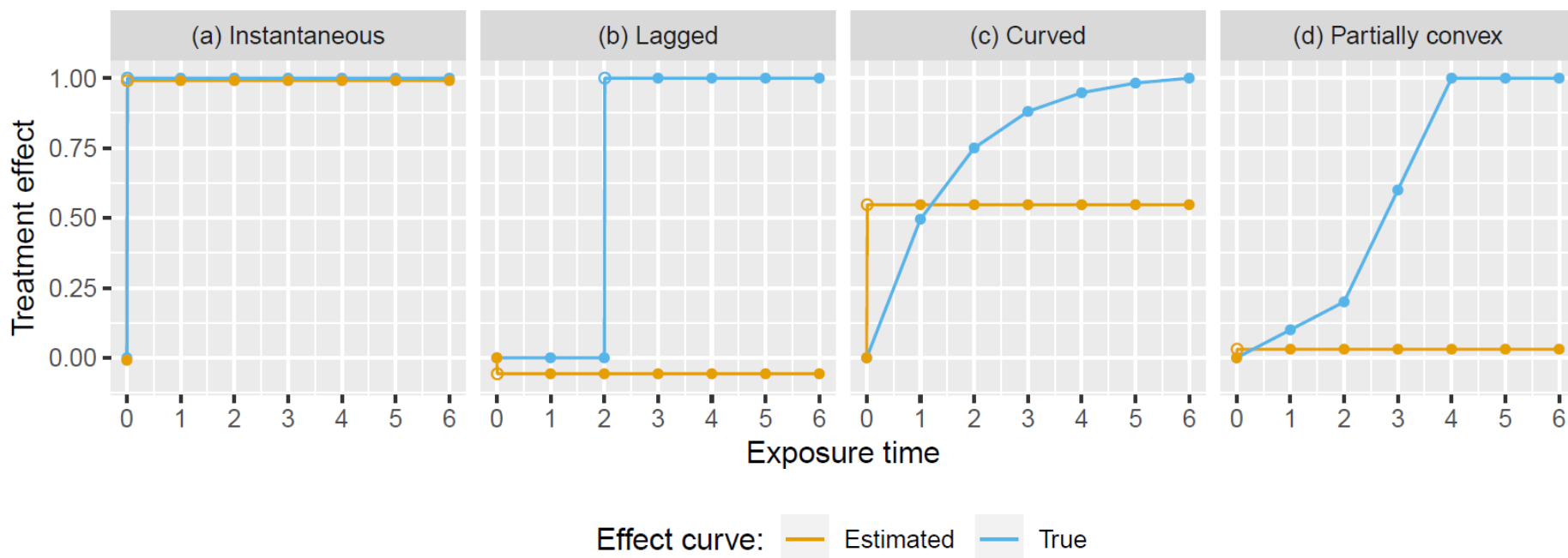
$$Y_{ijk} = \mu + \beta(t_{ij}) + \delta x_{ij} + a_i + b_{ij} + c_i x_{ij} + e_{ijk}$$

- Model for time – $\beta(t_{ij})$ – is wrong
 - May bias the estimate of δ
- Wrong variance structure
 - Wrong standard errors (conservative or anti-conservative) (Voldal et al 2022, Kasza et al 2019)
- **Treatment effect is not constant**
 - Can give misleading estimate of δ



Treatment effect is not constant

- What happens if you assume the treatment effect is immediate and constant (IT model), but it's not?



Kenny et al. (<https://arxiv.org/abs/2111.07190>)



An alternative TE estimator

- Carefully specify the estimand, Ψ
 - TE at a particular time after start of the intervention?
 - Average TE over an interval?
 - Average TE after a phase-in period?

- Fit a model of the form

$$Y_{ijk} = \cdots + \delta_{s_{ij}} x_{ij} + \cdots$$

and use the estimator

$$\hat{\Psi}_{[s_1, s_2]} \equiv \frac{1}{s_2 - s_1 + 1} \sum_{r=s_1}^{s_2} \hat{\delta}_r \quad 1 \leq s_1 \leq s_2 \leq T-1$$

- Gains robustness, but sacrifices efficiency



Recommendations

- 1) Fit flexible time effect (e.g. categorical time or spline)
- 2) Better to overfit than underfit random effects
- 3) If sufficient clusters, use a robust variance
 - With “small” numbers of clusters hypothesis tests based on robust variance may have inflated type I error rates (Thompson et al, 2021)
- 4) Avoid fitting IT model unless very confident that treatment effect is immediate and constant
 - Use an estimator of the form $\hat{\Psi}_{(s_1, s_2]}$



Current research directions



Desirable properties

- Robust to misspecification of model for time
- Robust to misspecification of variance structure
- Gives “sensible” results if treatment effect is not constant
 - Carefully define the estimand of interest
- As efficient as possible



Structural Nested Mean Models

- Developed by Robins and colleagues, originally to deal with complex causal questions
- Less restrictive assumptions than traditional regression models
- Exhibit “doubly robust” properties



Structural Nested Mean Models

$$E(Y_{ijk}) = \beta_j + \delta x_{ij}$$

$$V(Y_{ijk}) = \Sigma$$

The efficient score is

$$\sum_i (\mathbf{x}_i - \bar{\mathbf{x}})^T W_i \{(Y_i - \mathbf{x}_i \delta) - \beta\}$$

- $Y_i = (\bar{Y}_{i1}, \dots, \bar{Y}_{iT})$ $\mathbf{x}_i = (x_{i1}, \dots, x_{iT})$
- $\beta = (\beta_1, \dots, \beta_T)$
- $W_i^{-1} = \Sigma$



Structural Nested Mean Models

- Leads to the estimator

$$\hat{\delta} = \{\sum_i (x_i - \bar{x})^T W_i x_i\}^{-1} \sum_i (x_i - \bar{x})^T W_i (Y_i - \beta)$$

- Plug in estimates for W_i and β
- $\hat{\delta}$ is consistent and asymptotically normal *even if* W_i and β *are incorrectly specified* (“doubly robust”)!
- If assume $W_i = I$, $\beta = 0$ then (Hughes et al, 2020)

$$\hat{\delta} = \frac{\sum_i (x_i - \bar{x})^T Y_i}{\sum_i (x_i - \bar{x})^T x_i}$$



Structural Nested Mean Models I

$$\hat{\delta} = \frac{\sum_i (x_i - \bar{x})^T Y_i}{\sum_i (x_i - \bar{x})^T x_i}$$

- A “vertical” estimator that compares intervention to SOC within each time interval

		Time			
		1	2	3	4
cluster	1	0	1	1	1
	2	0	0	1	1
	3	0	0	0	1



Structural Nested Mean Models I

$$\hat{\delta} = \frac{\sum_i (x_i - \bar{x})^T Y_i}{\sum_i (x_i - \bar{x})^T x_i}$$

- $\text{Var}(\hat{\delta})$ may be derived asymptotically or by reference to the permutation distribution ($\mathcal{P}$) of $\hat{\delta}$

0	1	1	1
0	0	1	1
0	0	0	1

0	1	1	1
0	0	0	1
0	0	1	1

0	0	1	1
0	1	1	1
0	0	0	1

0	0	1	1
0	0	0	1
0	1	1	1

0	0	0	1
0	1	1	1
0	0	1	1

0	0	0	1
0	0	1	1
0	1	1	1

- Compute $\hat{\delta}$ for each permutation; compute $\text{Var}_{\mathcal{P}}(\hat{\delta})$
- Closed form expression for $\text{Var}_{\mathcal{P}}(\hat{\delta})$ possible!



Structural Nested Mean Models I

$$\hat{\delta} = \frac{\sum_i (x_i - \bar{x})^T Y_i}{\sum_i (x_i - \bar{x})^T x_i}$$

- Robust to time trend (assumed constant across all clusters, or clusters within predefined strata)
- Robust to variance structure mis-specification (assumed same across clusters; random treatment effect violates this)
- Relatively inefficient



Structural Nested Mean Models II

$$\hat{\delta} = \left\{ \sum_i (\mathbf{x}_i - \bar{\mathbf{x}})^T W_i \mathbf{x}_i \right\}^{-1} \sum_i (\mathbf{x}_i - \bar{\mathbf{x}})^T W_i (Y_i - \beta)$$

- Assume $W_i = I(1 - \rho) + J\rho$ and $\beta = 0$ giving

$$\hat{\delta}_\rho = \frac{\sum_i (x_i - \bar{x})^T \left(I - \frac{n\rho}{(1 + (nT - 1)\rho)} J \right) Y_i}{\sum_i (x_i - \bar{x})^T \left(I - \frac{n\rho}{(1 + (nT - 1)\rho)} J \right) x_i}$$



Structural Nested Mean Models II

$$\hat{\delta}_{\rho} = \frac{\sum_i (x_i - \bar{x})^T \left(I - \frac{n\rho}{(1 + (nT - 1)\rho)} J \right) Y_i}{\sum_i (x_i - \bar{x})^T \left(I - \frac{n\rho}{(1 + (nT - 1)\rho)} J \right) x_i}$$

- Since ρ is unknown, we estimate $\hat{\rho}$ (based on $Y_{ijk} - x_{ij}\hat{\delta}_{\rho} - \hat{\beta}$) and iterate.
- Can compute closed form for $\text{Var}_{\mathcal{P}}(\hat{\delta}_{\rho})$
- Improves efficiency - uses both vertical and horizontal estimates of intervention effect



Structural Nested Mean Models III

$$E(Y_{ijk}) = \beta_j + \delta_{s_{ij}}x_{ij}$$

$$V(Y_{ijk}) = \Sigma$$

Let $\delta = (\delta_1, \delta_2, \dots, \delta_{T-1})$

$$\hat{\delta} = \left\{ \sum_i (\Delta_i - \bar{\Delta})^T W_i \Delta_i \right\}^{-1} \sum_i (\Delta_i - \bar{\Delta})^T W_i (Y_i - \beta)$$

- Δ_i is a $T \times (T-1)$ matrix where row j indicates “exposure time” for observation ij
- $W_i = I(1 - \rho) + J\rho, \quad \beta = 0$



Simulation Results

- Simulation setting

$$Y_{ijk} = \mu + \beta_j + \delta h(s)x_{ij} + a_i + e_{ijk}$$

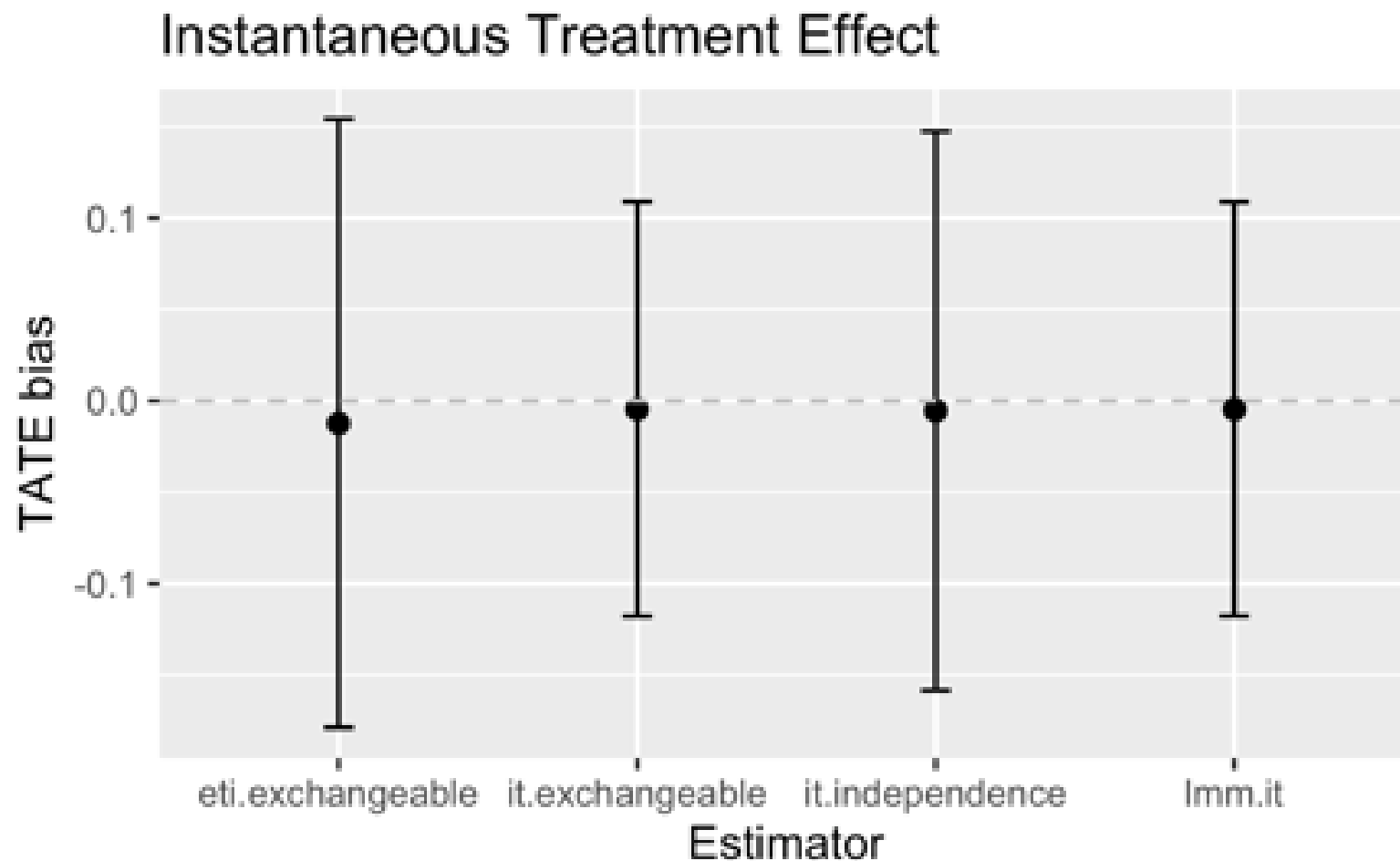
$$a_i \sim N(0, \tau^2)$$

$$e_{ijk} \sim N(0, \sigma^2)$$

- δ : maximum intervention effect ($\delta = 0.5$)
- $h(s)$: effect curve $0 \leq h(s) \leq 1$
- $N = 12, T = 7, n = 20, \sigma = 1, \tau = 0.25$ (ICC = 0.059)



Simulation Results





Next Steps

- More extensive simulations
 - alternative covariance structures
- More flexible/efficient choice for W_i
 - dependent on x (random treatment effects)
 - variable sample size
 - $W_i^{-1} = \Sigma$ for arbitrary Σ
 - Closed form for $\text{Var}_{\mathcal{P}}(\hat{\delta})$ may not be possible in general; use sandwich form instead
- Application to binary data (log and logit link)



What's happening?

		Time			
		1	2	3	4
cluster	1	0	1	2	3
	2	0	0	1	2
	3	0	0	0	1

$\delta(3)$
 $\delta(2)$
 $\delta(1)$

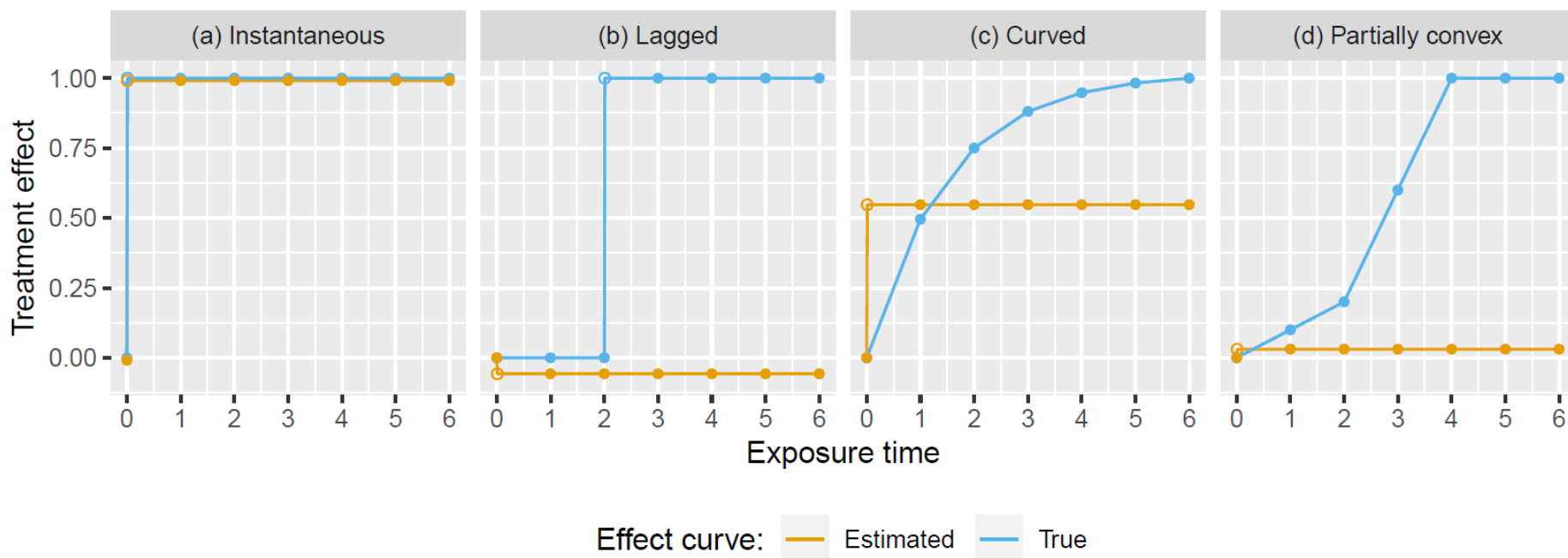
$$E(\hat{\delta}) = \sum_s w_s \delta_s$$

- weights sum to 1 (as expected) but ...
- weights can be > 1 and/or negative!



Treatment effect is not constant

- What happens if you assume the treatment effect is immediate and constant (IT model), but it's not?



Kenny et al. (<https://arxiv.org/abs/2111.07190>)



Treatment effect is not constant - Power

- Specify the SW design via design matrix $\mathbf{X}$ (including time-varying treatment effect)
- Use GLMM framework to specify variance Σ
- Specify the estimator $\hat{\Psi} \equiv \frac{1}{s_2 - s_1} \sum_{r=s_1+1}^{s_2} \hat{\delta}_r$

Then

$$\text{Var}(\hat{\Psi}) = \mathbf{H}^T (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{H}$$

and, for testing $H_0: \Psi = 0$,

$$\text{Power}(\Psi) = \Phi \left(\sqrt{\frac{\Psi^2}{\text{Var}(\hat{\Psi})}} - Z_{1-\frac{\alpha}{2}} \right)$$



Simulation Results – Type I error

- $T = 5$ time intervals
- $n = 10$ (average)
- $\delta = 0$
- $\sigma^2 = 1$ (error variance)
- $\tau^2 = 0.2$ (cluster variance)
- 10,000 replicates

		Treatment variance (η^2)					
		$V_{\delta=0}^1(\hat{\delta})$			$V_{\delta=\hat{\delta}}^1(\hat{\delta})$		
N	n	0	0.1	0.4	0	0.1	0.4
12	constant	0.05	0.05	0.05	0.06	0.07	0.07
	low var	0.05	0.05	0.05	0.06	0.07	0.07
	hi var	0.05	0.05	0.05	0.07	0.07	0.07
24	constant	0.05	0.05	0.06	0.06	0.06	0.06
	low var	0.05	0.05	0.06	0.06	0.06	0.06
	hi var	0.04	0.04	0.05	0.05	0.05	0.05
36	constant	0.05	0.05	0.06	0.05	0.06	0.06
	low var	0.05	0.05	0.06	0.06	0.06	0.06
	hi var	0.05	0.06	0.06	0.06	0.06	0.07

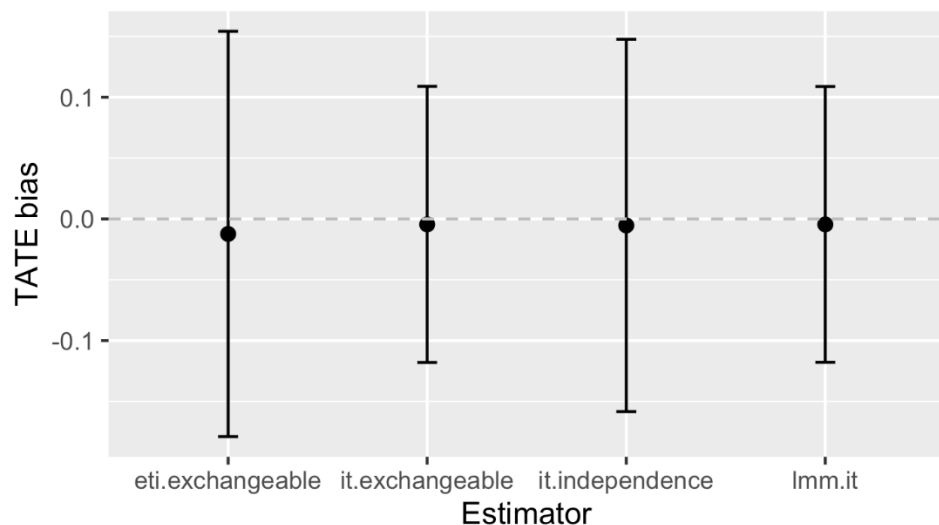
Hughes et al (2020)



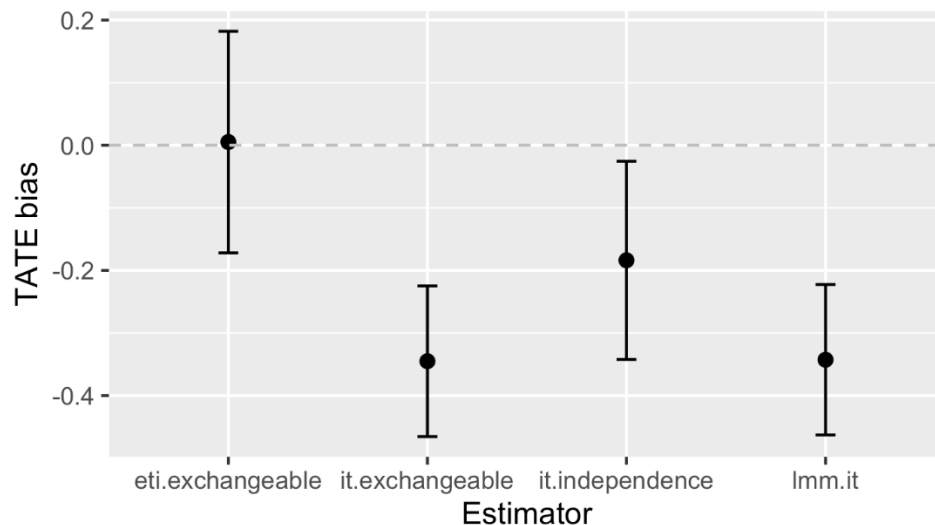
SCHOOL OF PUBLIC HEALTH • UNIVERSITY of WASHINGTON

Department of Biostatistics

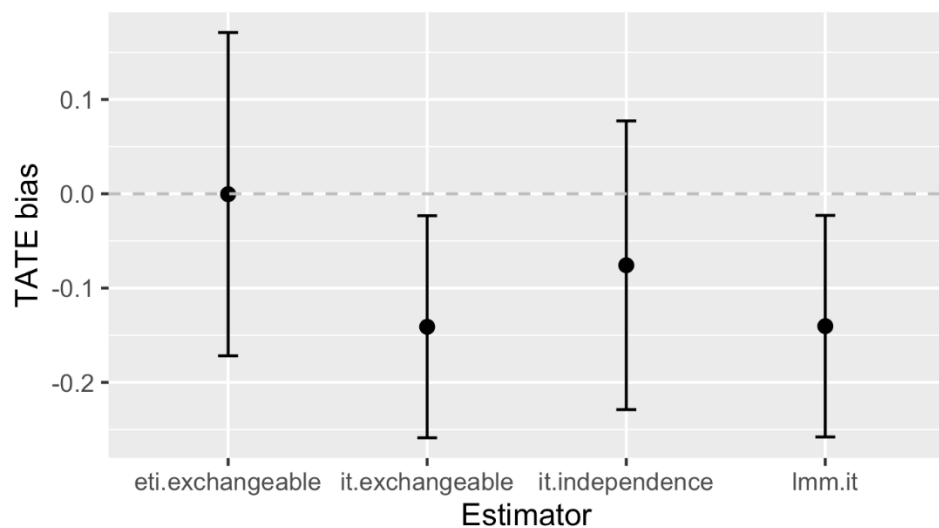
Instantaneous Treatment Effect



Lagged Treatment Effect



Curved Treatment Effect



Partially Convex Treatment Effect

