Nonparametric Causal Survival Analysis under Clustered Interference

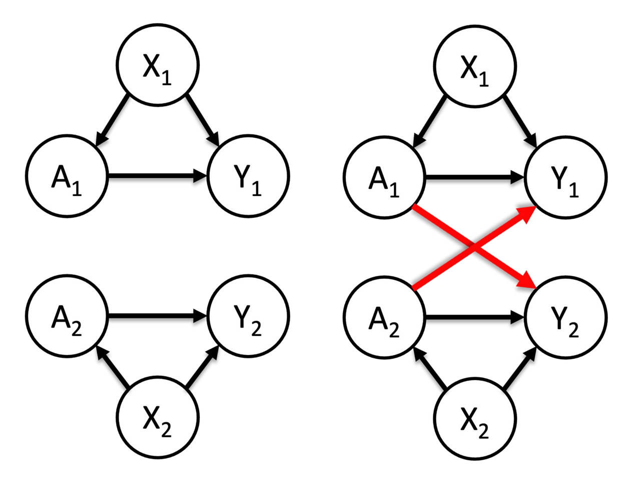
THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

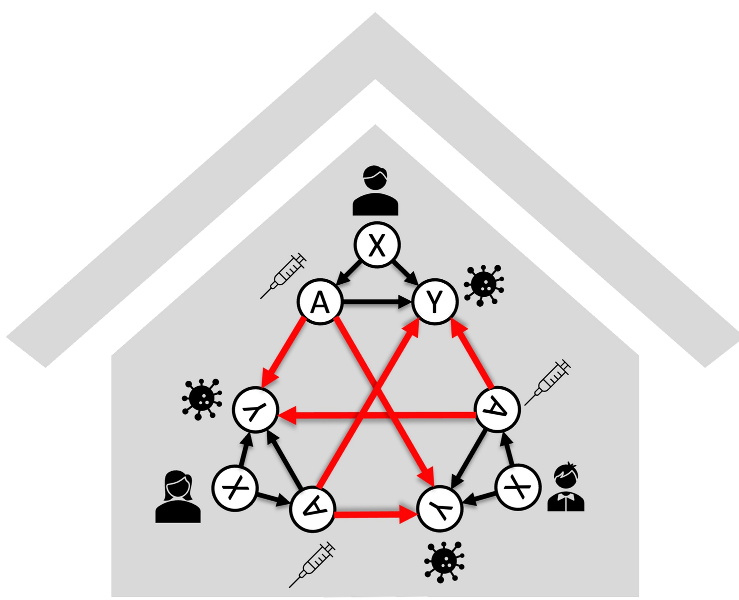
Chanhwa Lee[1], Donglin Zeng[2], Michael G. Hudgens[1]

[1]Department of Biostatistics, University of North Carolina at Chapel Hill, [2]Department of Biostatistics, University of Michigan, Ann Arbor

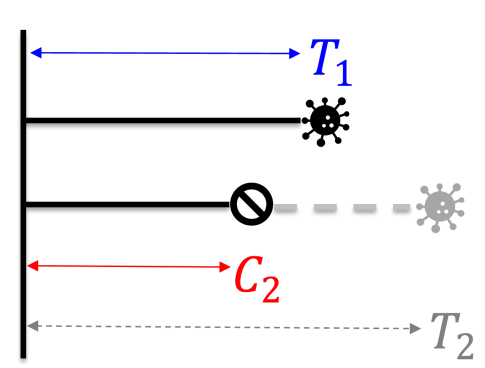
1 MOTIVATION

Clustered Interference [1]

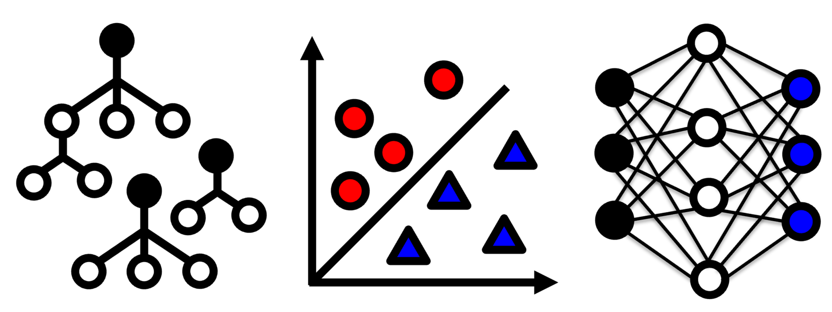




Right Censoring [2]



Nonparametric (Data-adaptive) [3]



2 BACKGROUND

Observed data

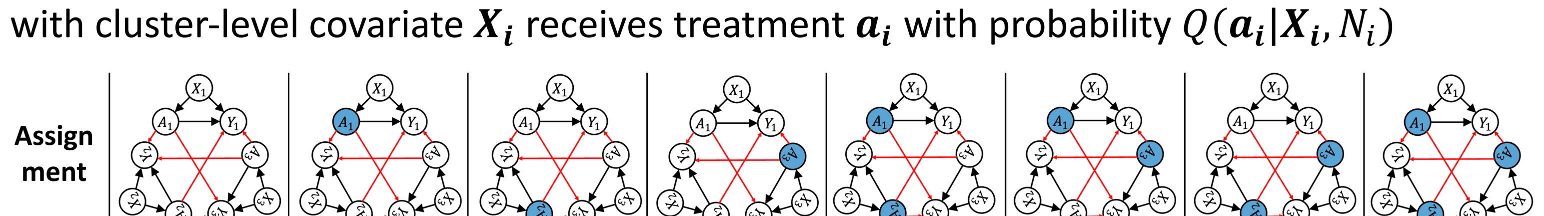
- Cluster $i \in \{1,\dots,m\}$, Unit $j \in \{1,\dots,N_i\}$
- $T_{ij} \in \mathbb{R}^+$: event time, $A_{ij} \in \{0,1\}$: treatment, $X_{ij} \in \mathbb{R}^p$: confounders
- $C_{ij} \in \mathbb{R}^+$: censoring time, $Y_{ij} = \min(T_{ij}, C_{ij})$, $\Delta_{ij} = 1(T_{ij} \le C_{ij})$

Potential outcome

- $T_{ij}(\boldsymbol{a}_i)$: Potential event time for unit j in cluster i when the cluster i receives $\boldsymbol{a}_i$
- $T_{ij}(\boldsymbol{a}_i) = T_{ij}(a_{ij}, \boldsymbol{a}_{i(-j)})$, $\boldsymbol{a}_{i(-j)} = (a_{i1},\dots,a_{i(j-1)}, a_{i(j+1)},\dots a_{iN_i})$
- No interference: $T_{ij}(a_{ij}, \boldsymbol{a}_{i(-j)}) = T_{ij}(a_{ij}, \boldsymbol{a'}_{i(-j)})$

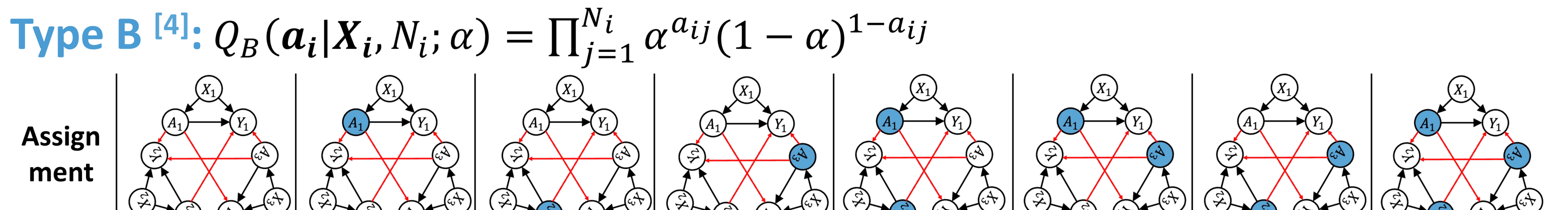
3 STOCHASTIC POLICY

Definition $Q(\cdot|\boldsymbol{X}_i, N_i)$: probability distribution on $\{0,1\}^{N_i}$ such that cluster of size N_i with cluster-level covariate $\boldsymbol{X}_i$ receives treatment $\boldsymbol{a}_i$ with probability $Q(\boldsymbol{a}_i|\boldsymbol{X}_i, N_i)$



Assignment	(graph)	(graph)	(graph)	(graph)	(graph)	(graph)	(graph)	(graph)
Prob	$Q(0,0,0\|X,N)$	$Q(1,0,0\|X,N)$	$Q(0,1,0\|X,N)$	$Q(0,0,1\|X,N)$	$Q(1,1,0\|X,N)$	$Q(1,0,1\|X,N)$	$Q(0,1,1\|X,N)$	$Q(1,1,1\|X,N)$

Type B [4]: $Q_B(\boldsymbol{a}_i|\boldsymbol{X}_i, N_i;\alpha) = \prod_{j=1}^{N_i} \alpha^{a_{ij}}(1-\alpha)^{1-a_{ij}}$



Assignment	(graph)	(graph)	(graph)	(graph)	(graph)	(graph)	(graph)	(graph)
Prob	$(1\text{-}0.7)^3$	$0.7(1\text{-}0.7)^2$	$0.7(1\text{-}0.7)^2$	$0.7(1\text{-}0.7)^2$	$0.7^2(1\text{-}0.7)$	$0.7^2(1\text{-}0.7)$	$0.7^2(1\text{-}0.7)$	0.7^3

CIPS [5]: $Q_{CIPS}(\boldsymbol{a}_i|\boldsymbol{X}_i, N_i;\delta) = \prod_{j=1}^{N_i} \pi_{ij,\delta}^{a_{ij}}(1-\pi_{ij,\delta})^{1-a_{ij}}$

- Propensity score of unit j in cluster i: $\pi_{ij} = P(A_{ij} = 1|\boldsymbol{X}_i, N_i)$
- Shifted (counterfactual) propensity score: $\pi_{ij,\delta}$ from $\frac{\pi_{ij,\delta}}{1-\pi_{ij,\delta}} = \delta \times \frac{\pi_{ij}}{1-\pi_{ij}}$

4 ESTIMANDS

- **Expected Overall Risk by time τ under policy Q**

$$\mu(\tau;Q) = \mathbb{E}\left\{\frac{1}{N_i}\sum_{j=1}^{N_i}\sum_{\mathbf{a}_i\in\mathcal{A}(N_i)} \mathbb{1}(T_{ij}(\mathbf{a}_i)\le\tau)Q(\mathbf{a}_i|\mathbf{X}_i,N_i)\right\}$$

- **Expected Risk by time τ when treated under policy Q**

$$\mu_1(\tau;Q) = \mathbb{E}\left\{\frac{1}{N_i}\sum_{j=1}^{N_i}\sum_{\mathbf{a}_{i(-j)}\in\mathcal{A}(N_i-1)} \mathbb{1}(T_{ij}(1,\mathbf{a}_{i(-j)})\le\tau)Q(\mathbf{a}_{i(-j)}|\mathbf{X}_i,N_i)\right\}$$

- $OE(\tau;Q,Q') = \mu(\tau;Q) - \mu(\tau;Q')$: compares two policies overall
  E.g. Difference in overall COVID19 risks when 50% vs. 30% of neighbors vaccinated
- $DE(\tau;Q) = \mu_1(\tau;Q) - \mu_0(\tau;Q)$: effect of treatment under policy Q
  E.g. Vaccine effect on COVID19 when 50% of neighbors vaccinated
- $SE_0(\tau;Q,Q') = \mu_0(\tau;Q) - \mu_0(\tau;Q')$: compares risk when untreated
  E.g. Unvaccinated unit's COVID19 risks when 50% vs. 30% of neighbors vaccinated

5 METHOD

1. Full data estimation equation from nonparametric EIF

$$\varphi^{F,*}_{ij}(\tau;\mathbf{Z}_i) := \sum_{\mathbf{a}_i\in\mathcal{A}(N_i)}\{w_j(\mathbf{a}_i,\mathbf{X}_i,N_i)+\phi_j(\mathbf{A}_i,\mathbf{X}_i,N_i;\mathbf{a}_i)\}\mathbb{P}(T_{ij}\le\tau|\mathbf{A}_i=\mathbf{a}_i,\mathbf{X}_i,N_i)$$
$$+\mathbb{P}(\mathbf{A}_i|\mathbf{X}_i,N_i)^{-1}w_j(\mathbf{A}_i,\mathbf{X}_i,N_i)\{\mathbb{1}(T_{ij}\le\tau)-\mathbb{P}(T_{ij}\le\tau|\mathbf{A}_i,\mathbf{X}_i,N_i)\}-\Psi(\tau;\mathbf{w})$$

2. Augmented-IPCW estimating equation

$$0 = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{N_i}\sum_{j=1}^{N_i}\left[\frac{\Delta_{ij}}{S^C_{ij}(Y_{ij}|\mathbf{A}_i,\mathbf{X}_i,N_i)}\varphi^{F,*}_{ij}(\tau;\mathbf{Z}_i)+\int_0^\infty\frac{\mathbb{E}\{\varphi^{F,*}_{ij}(\tau;\mathbf{Z}_i)|T_{ij}\ge r,\mathbf{A}_i,\mathbf{X}_i,N_i\}}{S^C_{ij}(r|\mathbf{A}_i,\mathbf{X}_i,N_i)}dM^C_{ij}(r)\right]$$

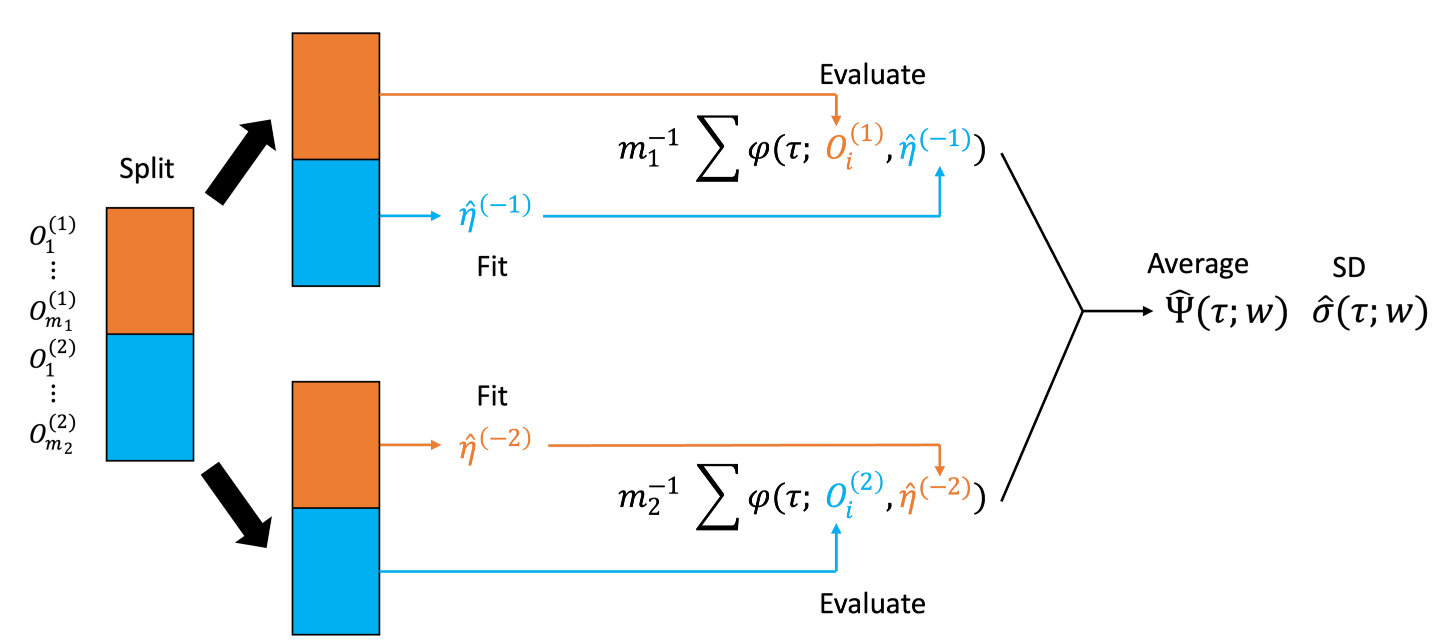
$$\varphi_{ij}(\tau;\mathbf{O}_i) = \sum_{\mathbf{a}_i\in\mathcal{A}(N_i)}\mathrm{OR}_{ij}(\tau;\mathbf{O}_i,\mathbf{a}_i) + \mathrm{IPCW\text{-}BC}ij(\tau;\mathbf{O}_i) + \mathrm{AUG}_{ij}(\tau;\mathbf{O}_i)$$

$$\mathrm{OR}_{ij}(\tau;\mathbf{O}_i,\mathbf{a}_i) = \{w_j(\mathbf{a}_i,\mathbf{X}_i,N_i)+\phi_j(\mathbf{A}_i,\mathbf{X}_i,N_i;\mathbf{a}_i)\}F^T_{ij}(\tau|\mathbf{A}_i=\mathbf{a}_i,\mathbf{X}_i,N_i),$$

$$\mathrm{IPCW\text{-}BC}_{ij}(\tau;\mathbf{O}_i) = \frac{w_j(\mathbf{A}_i,\mathbf{X}_i,N_i)}{\mathbb{P}(\mathbf{A}_i|\mathbf{X}_i,N_i)}\left\{\frac{\Delta_{ij}}{S^C_{ij}(Y_{ij}|\mathbf{A}_i,\mathbf{X}_i,N_i)}\mathbb{1}(Y_{ij}\le\tau)-F^T_{ij}(\tau|\mathbf{A}_i,\mathbf{X}_i,N_i)\right\},$$

$$\mathrm{AUG}_{ij}(\tau;\mathbf{O}_i) = \frac{w_j(\mathbf{A}_i,\mathbf{X}_i,N_i)}{\mathbb{P}(\mathbf{A}_i|\mathbf{X}_i,N_i)}\int_0^\tau\frac{S^T_{ij}(r|\mathbf{A}_i,\mathbf{X}_i,N_i)-S^T_{ij}(\tau|\mathbf{A}_i,\mathbf{X}_i,N_i)}{S^C_{ij}(r|\mathbf{A}_i,\mathbf{X}_i,N_i)S^T_{ij}(r|\mathbf{A}_i,\mathbf{X}_i,N_i)}dM^C_{ij}(r),$$

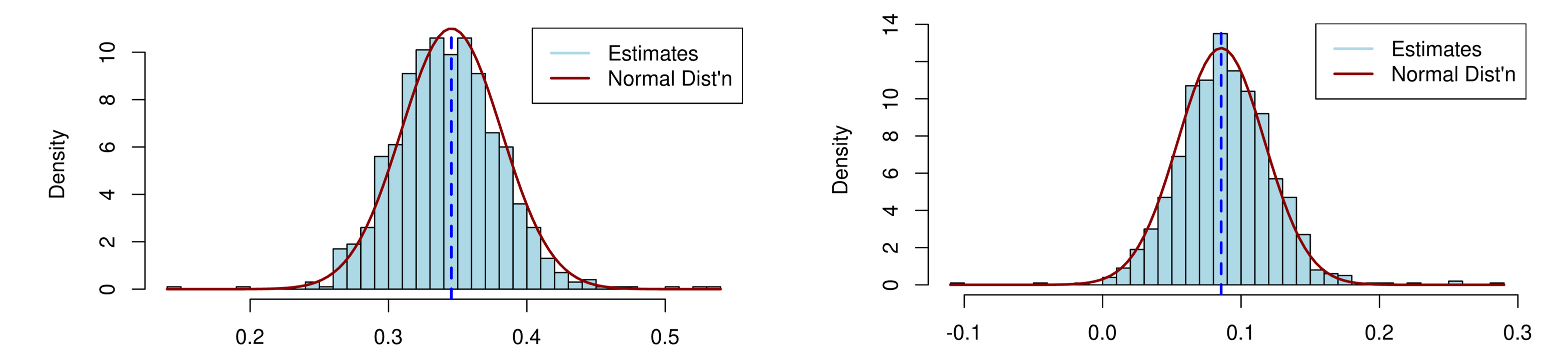
3. Sample Splitting [6] & Nonparametric Nuisance Function Estimation



6 RESULTS

Theory

- Consistent & Asymptotically Normal & Weak Convergence to Gaussian Process



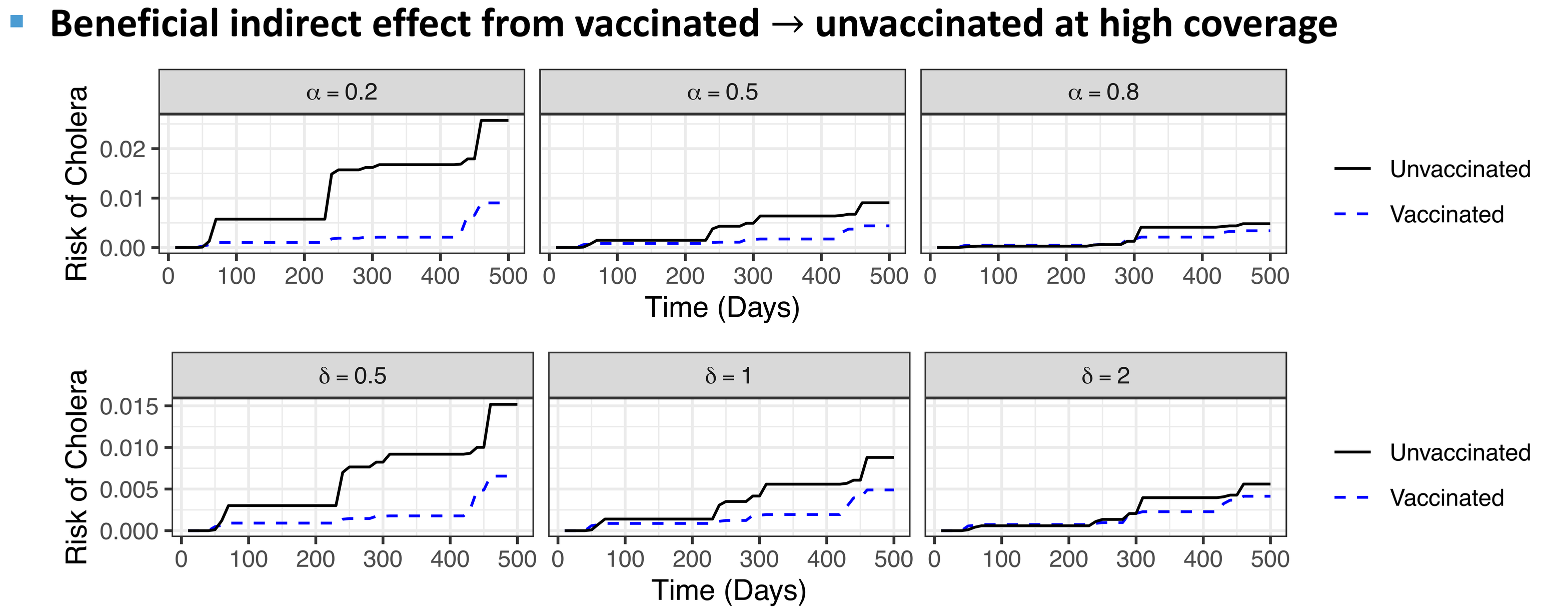
Simulation

- Flexible data-adaptive nuisance function estimation

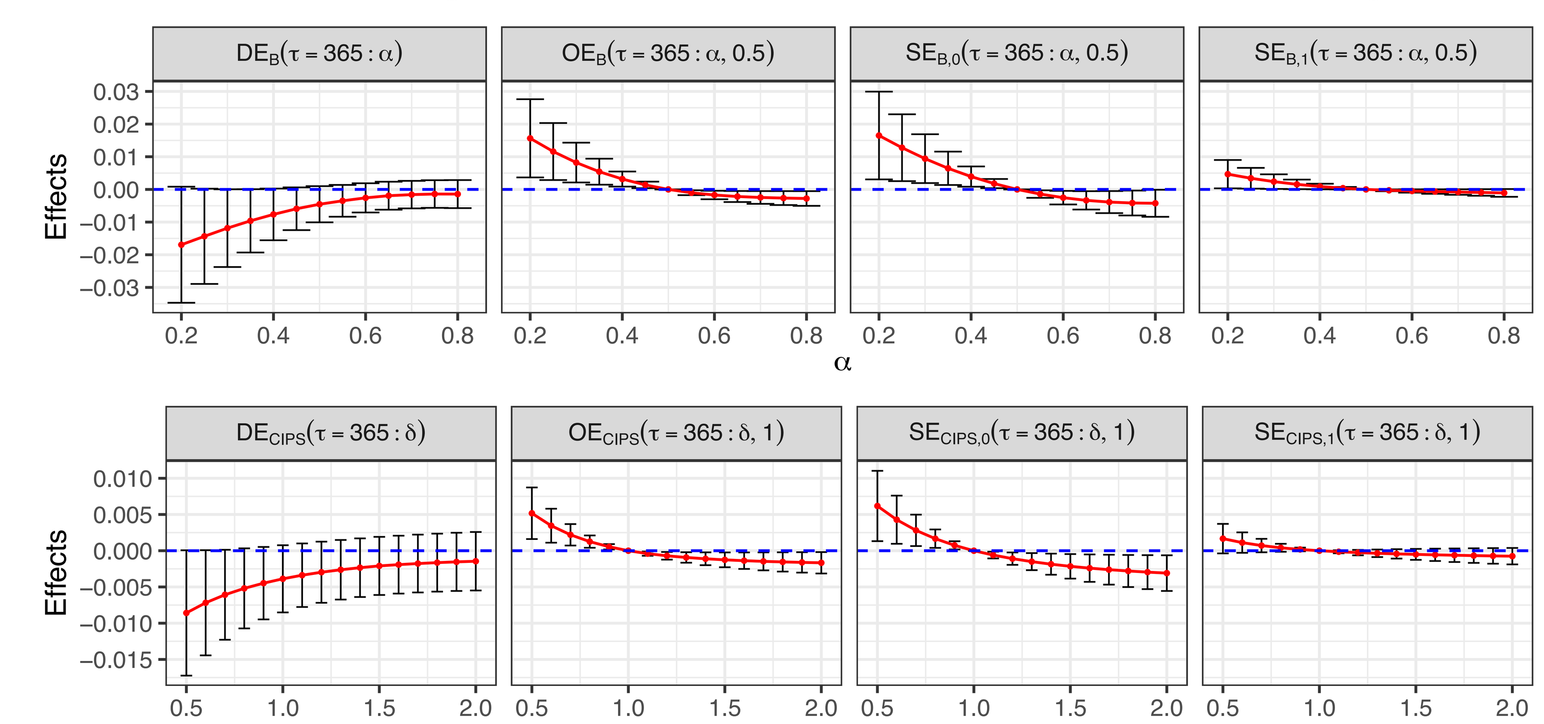
Estimand	Truth	Nonparametric (SL, RSF) Bias	RMSE	ASE	ESE	Cov %	Parametric (GLM, Cox) Bias	RMSE	ASE	ESE	Cov %	RMSE Ratio
$\mu_{\text{CIPS}}(0.3;0.5)$	0.345	-0.003	0.001	0.036	0.038	94.1%	-0.024	0.002	0.038	0.041	87.4%	0.642
$\mu_{\text{CIPS},1}(0.3;0.5)$	0.151	-0.001	0.001	0.025	0.027	93.9%	-0.016	0.001	0.027	0.028	88.3%	0.674
$\mu_{\text{CIPS},0}(0.3;0.5)$	0.524	-0.004	0.004	0.058	0.063	94.3%	-0.032	0.005	0.063	0.066	90.4%	0.744
$DE_{\text{CIPS}}(0.3;0.5)$	-0.373	0.002	0.004	0.058	0.064	94.1%	0.015	0.005	0.063	0.067	93.9%	0.853
$SE_{\text{CIPS},1}(0.3;0.5,1)$	0.021	0.001	0.001	0.021	0.024	94.9%	-0.003	0.001	0.023	0.025	95.1%	0.943
$SE_{\text{CIPS},0}(0.3;0.5,1)$	0.030	0.006	0.003	0.051	0.059	95.5%	-0.004	0.004	0.054	0.059	95.5%	0.986
$OE_{\text{CIPS}}(0.3;0.5,1)$	0.086	0.003	0.001	0.031	0.034	94.8%	-0.005	0.001	0.033	0.038	94.5%	0.800

Application (Cholera Vaccine Study)

- **Beneficial direct effect of vaccination at lower vaccine coverage**
- **Beneficial indirect effect from vaccinated → unvaccinated at high coverage**



- **Unvaccinated individuals can benefit from spillover effects from vaccinated individuals, and the magnitude of such benefit increases over vaccine coverage**



7 Discussion

- **Inference about treatment effects under clustered interference and censoring**
- **Can be applied to any stochastic treatment allocation policy**
- **Data-adaptive estimation with robust correction to yield CAN estimator**

Literature cited

1. Hudgens, M. G. and Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association* 103, 832–842.
2. Chakladar, S., Rosin, S., Hudgens, M. G., Halloran, M. E., Clemens, J. D., Ali, M., and Emch, M. E. (2022). Inverse probability weighted estimators of vaccine effects accommodating partial interference and censoring. *Biometrics* 78, 777–788.
3. Park, C. and Kang, H. (2022). Efficient semiparametric estimation of network treatment effects under partial interference. *Biometrika* 109, 1015–1031.
4. Tchetgen Tchetgen, E. J. and VanderWeele, T. J. (2012). On causal inference in the presence of interference. *Statistical Methods in Medical Research* 21, 55–75.
5. Lee, C., Zeng, D., and Hudgens, M. G. (2024). Efficient Nonparametric Estimation of Stochastic Policy Effects with Clustered Interference. Journal of the American Statistical Association
6. Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* 21, C1–C68.

Further information

Please address questions or comments to Chanhwa Lee at chanhwa@email.unc.edu.

Funding

This work was supported by a grant from the National Institutes of Health, USA (NIH grant NIH R01 AI085073).