Induced modified treatment policies for estimating effects of continuous exposures under interference

Salvador Balkus
sbalkus@g.harvard.edu
Harvard Biostatistics

Nima Hejazi
nhejazi@hsph.harvard.edu
Harvard Biostatistics

May 29, 2026

Environmental health studies

How to quantify causal effects involving…

- air pollution?
- wildfires?
- extreme heat?

All involve ***continuous*** *exposures*

Air pollution from factory

Canadian wildfire smoke blankets NYC

Anatomy of the data: The standard set-up

Observed data: A tuple of n-vectors, $O_1, \ldots, O_n$, sampled i.i.d., where

$$\mathbf{O} = (\mathbf{L}, \mathbf{A}, \mathbf{Y}) \sim \mathsf{P} \in \mathcal{P}$$

- $\mathbf{L}$: measured baseline covariates
- $\mathbf{A}$: continuous exposure
- $\mathbf{Y}$: outcome of interest

Question: How much would Y have changed had we intervened upon A?

Causal inference with a continuous exposure

- Let $Y(a)$ denote potential outcome Y after setting $A = a$
- Typically, interest lies in counterfactual mean $\mathbb{E}[Y(a)]$, the average value of Y had $A = a$ been assigned

- What goes wrong when A is continuous?
  - → cannot observe all possible $a \in \mathcal{A}$; hard to identify and estimate dose-response curve
  - → "Assigning" $A = a$ often impractical (e.g., how to "assign" wildfire smoke exposure?)

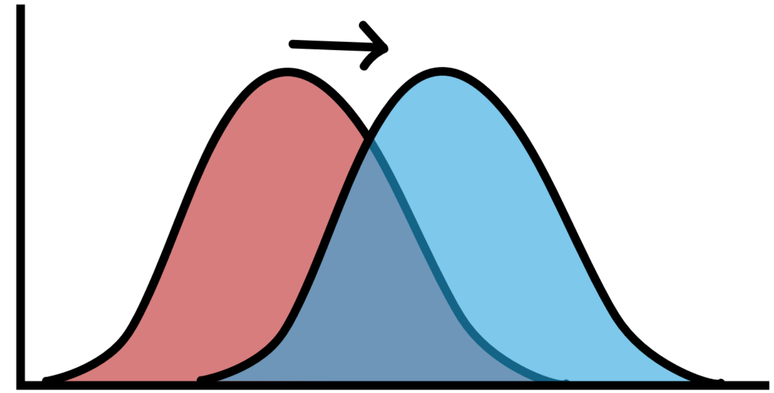
- <u>*Solution*</u>: Consider *modifying* the observed exposure.

Modified treatment policies

A user-specified function $d(A, L; \delta)$ that maps the observed exposure A to a post-intervention value A^d (Díaz and van der Laan 2012; Haneuse and Rotnitzky 2013).

- *Additive* MTP: $d(A, L; \delta) = A + \delta$
- *Multiplicative* MTP: $d(A, L; \delta) = A \cdot \delta$
- *Piecewise additive* MTP:

$$d(A, L; \delta) = \begin{cases} A + \delta \cdot L & A \in \mathcal{A}(L) \\ A & \text{otherwise} \end{cases}$$



Causal effect of a modified treatment policy

The counterfactual mean is

$$\mathbb{E}_{\mathsf{P}}\Big[Y(d(A, L; \delta))\Big] = \mathbb{E}_{\mathsf{P}}\Big[Y(A^d)\Big] \ ,$$

and the *population intervention effect* (PIE) is $\mathbb{E}[Y(A^d)] - \mathbb{E}[Y]$.

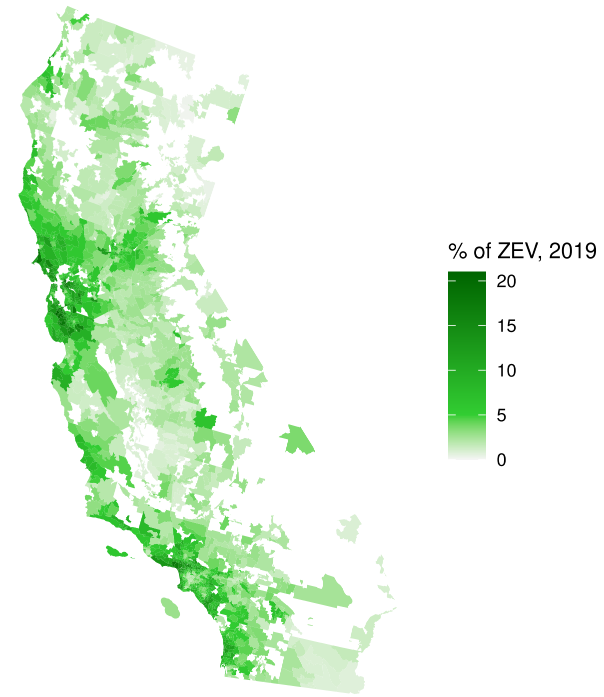
- "Average change in Y *caused* by modifying each A_i via $d(A, L; \delta)$."
- Causal, non-parametric analog of a linear regression coefficient.

Spatial data dependence

Motivating example: Electric vehicles

Question: What is the impact of zero-emissions vehicles (ZEV) on NO_2 air pollution in California?

- *Continuous exposure*: proportion of ZEVs registered in a ZIP code in California
- No known intervention can "set units' proportion of ZEVs to $A = a$".
- But can plausibly consider MTP effects: $\mathbb{E}[Y(A+1)]$ or $\mathbb{E}[Y(1.01 \cdot A)]$



How to **identify** and **estimate** causal effects of MTPs in spatial data?

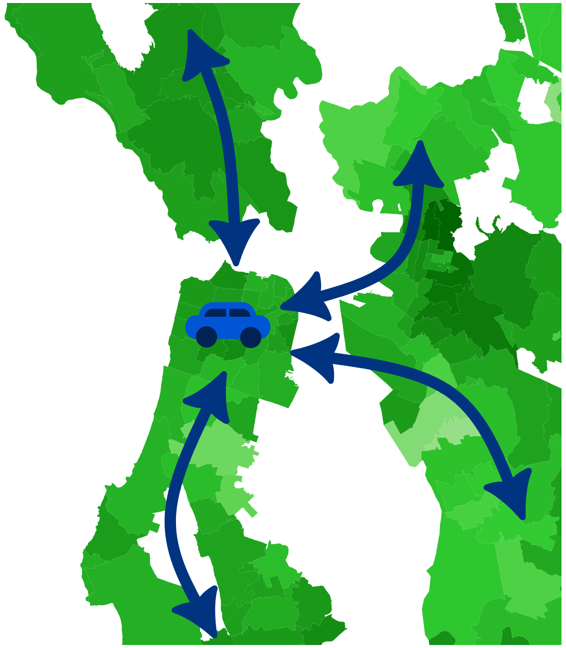
Must be…

- Policy-relevant—*intervention must define a population estimand*
- Flexibly estimable—*no reliance on restrictive parametric forms*
- Efficient—*attain the lowest admissable variance*

Interference

Hudgens and Halloran (2008): *Interference* occurs when potential outcome of unit i depends on exposures of *other units* $j \neq i$:

$$Y_i(a_i, a_j) \neq Y_i(a_i, a_j') \text{ if } a_j \neq a_j'$$



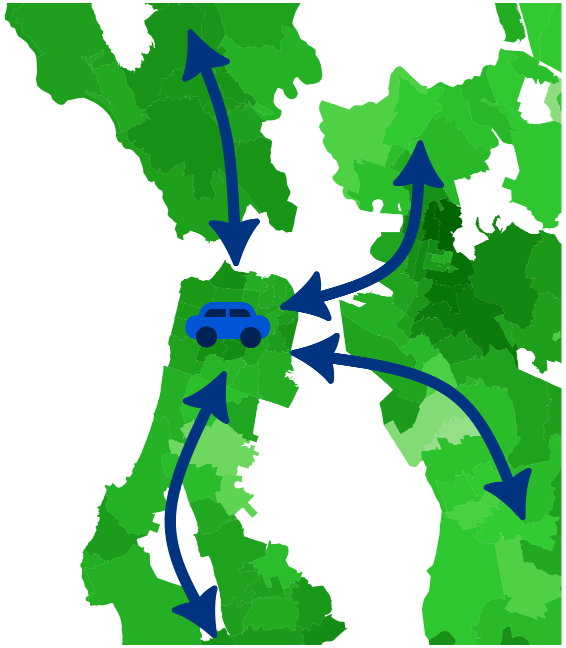
- Common in spatial data (due to dependence)
- SUTVA violated $\rightarrow$ Causal identification fails
- Correlated data $\rightarrow$ Challenges for estimation

Interference

Hudgens and Halloran (2008): *Interference* occurs when potential outcome of unit i depends on exposures of *other units* $j \neq i$

$$Y_i(a_i, a_j) \neq Y_i(a_i, a_j') \text{ if } a_j \neq a_j'$$

Network interference: Potential outcomes depend only on neighbors in a *known* adjacency matrix $\mathbf{F}$ (van der Laan 2014).



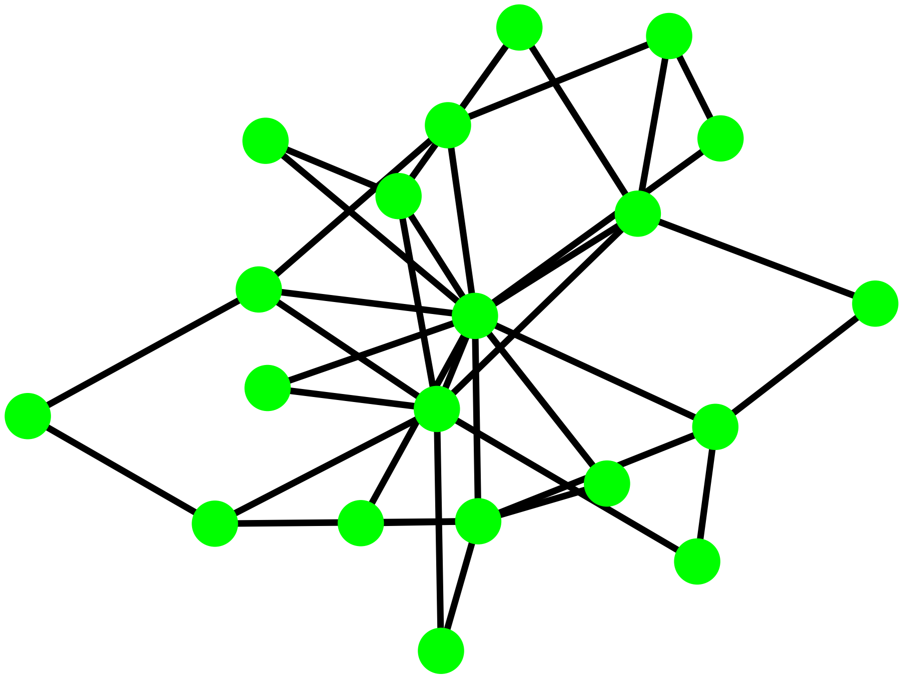
Effects of *induced* MTPs

Anatomy of the data: The dependent-data set-up

1. **Observed data**: A tuple of n-vectors, $O_1, \ldots, O_n$, where

$$\mathbf{O} = (\mathbf{L}, \mathbf{A}, \mathbf{Y})$$

2. **Network $\mathbf{F}$**: An adjacency matrix of each unit's neighbors (known).



Repairing identification under interference

Under interference, consider the following structural equation:

$$Y_i = f\Big(\boldsymbol{s}_A(A_j : j \in \mathbf{F}_i), \boldsymbol{s}_L(L_j : j \in \mathbf{F}_i)\Big)$$

- $\boldsymbol{s}$: "summary" of neighbors' exposures or ***exposure mapping***
- As shorthand, denote vector of $\boldsymbol{s}_A(A_j : j \in \mathbf{F}_i)$ as $\boldsymbol{s}(\mathbf{A})$
  - → Example: $\boldsymbol{s}(\mathbf{A})_i = \sum_{j \in \mathbf{F}_i} A_j$

Treating $\boldsymbol{s}(\mathbf{A})$ as the exposure instead of $\mathbf{A}$ restores SUTVA (Aronow and Samii 2017) → use $Y(\boldsymbol{s}(a))$ instead of $Y(a)$.

The induced MTP

But what happens if we apply the MTP *and then* summarize?

$$A \xrightarrow{d} A^d \xrightarrow{s} A^{s \circ d}$$

- We term the function $s \circ d$ the **induced MTP**.
- Population intervention effect (PIE) of an induced MTP:

$$\Psi_n(\mathsf{P}) = \mathbb{E}_{\mathsf{P}}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i(s(d(\mathbf{A}, \mathbf{L}; \delta))_i)\right] - \mathbb{E}_{\mathsf{P}}\left[Y\right]$$

- *Data-adaptive*; each Y_i depends on number of neighbors

Identification

Network analogs of standard assumptions (weaker)

- **A0** (SCM). Data generated from structural causal model:

$$L_i = f_L(\varepsilon_{L_i}); A_i = f_A(L_i^s, \varepsilon_{A_i}); Y_i = f_Y(A_i^s, L_i^s, \varepsilon_{Y_i}) \ ,$$

  with errors identically-distributed and $\varepsilon_i \perp\!\!\!\perp \varepsilon_j$ provided i, j are not neighbors in $\mathbf{F}$

- **A1** (Summary positivity). If $s(a), s(l) \in \text{supp}(A^s, L^s)$ then also $s(a^d), s(l) \in \text{supp}(A^s, L^s)$

- **A2** (No unmeasured confounding). $Y_i(a^s) \perp\!\!\!\perp A_i^s \mid \mathbf{L}$

Technical conditions on d and s

- **A3** (Piecewise smooth invertibility). The MTP d has a differentiable inverse on a countable partition of $\mathbf{supp}(A)$.

- **A4** (Summary **coarea**). s has Jacobian Js satisfying almost-everywhere

$$\underbrace{\sqrt{\det Js(a) Js(a)^\top}}_{\text{coarea}} > 0$$

What is coarea?

- Recall the change-of-variables (**area**) formula: If $Y = f(X)$, then

$$p_Y(y) = p_X(f^{-1}(x)) \cdot \frac{1}{|\det Jf(x)|}$$

- **Coarea** generalizes this when Jf may not be square (Negro 2022)

$$p_Y(y) = \int_{f(x)=y} \frac{p_X(x)}{\sqrt{\det Jf(x) Jf(x)^\top}} d\mu$$

Statistical estimand for induced MTP effects

Statistical estimand factorizes in terms of $\boldsymbol{A}^s$:

$$\psi_n = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{\mathsf{P}}\left[m(A_i^s, L_i^s) \cdot r(A_i^s, A_i^{s \circ d}, L_i^s) \cdot w(\mathbf{A}, \mathbf{L})_i\right] ,$$

with nuisance parameters $\boldsymbol{m}$ and $\boldsymbol{r}$, and *deterministic* weights $\boldsymbol{w}$:

$$m(a^s, l^s) = \mathbb{E}_{\mathsf{P}}[Y \mid A_i^s = a^s, L_i^s = l^s]$$

$$r(a^s, a^{s \circ d^{-1}}, l^s) = \frac{p(a^{s \circ d^{-1}} \mid l^s)}{p(a^s \mid l^s)}$$

$$w(\mathbf{a}, \mathbf{l}) = \sqrt{\frac{\det J(s \circ d^{-1})(\mathbf{a}) J(s \circ d^{-1})(\mathbf{a})^\top}{\det Js(\mathbf{a}) Js(\mathbf{a})^\top}}$$

Advantages of MTP effects in the network setting

- **Population-level** estimand, so the intervention is always compatible with the network as-observed
- Ameliorates **positivity violations** that would occur if enforcing static interventions directly on summaries $s(A)$
- Components of estimand depend on the data only through A^s and L^s

Estimation

Efficient estimation

Goal: Construct a doubly-robust, semi-parametric efficient estimator based on the **efficient influence function** $\phi(\mathsf{P})$

$$\frac{1}{n}\sum_{i=1}^{n}\phi(\mathsf{P}_{\hat{\eta}})(O_i) \ ,$$

where $\hat{\eta}$ is a set of nuisance estimators whose *product* converges at $o_{\mathsf{P}}(n^{-1/2})$ (i.e., only need $o_{\mathsf{P}}(n^{-1/4})$), typical in statistical learning)

- One-step de-biasing (Bickel et al. 1993; Pfanzagl and Wefelmeyer 1985)
- TMLE (van der Laan and Rose 2011; van der Laan and Rubin 2006)

Efficient estimation

The efficient influence function of ψ_n, a special case of the EIF for the counterfactual mean of a stochastic intervention (Ogburn et al. 2022), is

$$\bar{\phi}(\mathsf{P})(O_i) = \frac{1}{n}\sum_{i=1}^{n} w(\mathbf{A}, \mathbf{L})_i \cdot r(A_i^s, L_{s,i})(Y_i - m(A_i^s, L_i^s)) + \mathbb{E}(m(A_i^{s \circ d}, L_i^s; \delta), L_i^s) \mid \mathbf{L} = \mathbf{l}) - \psi_n \ ,$$

Efficient estimation

Ogburn et al. (2022)'s CLT: If $\hat{\psi}_n$ solves $\bar{\phi} \approx 0$ and $K_{\max}^2/n \to 0$, then, under mild regularity conditions,

$$\sqrt{C_n}(\hat{\psi}_n - \psi_n) \to \mathrm{N}(0, \sigma^2) \ ,$$

where $K_{\max}$ is the network's maximum degree.

The estimator $\hat{\psi}_n$ is asymptotically normal, but the rate depends on a factor $n/K_{\max}^2 < C_n < n$ ("automatically" contained within $\hat{\sigma}^2$)

Variance estimation

EIF was given in the form $\frac{1}{n}\sum_{i=1}^{n}\phi_{\mathsf{P}}(O_i)$, but must be centered at the means of units with the same number of neighbors $N(|\mathbf{F}_i|)$:

$$\varphi_i = \phi_{\hat{\mathsf{P}}_{n(O_j)}}(O_i) - \frac{1}{|N(|\mathbf{F}_i)||)|}\sum_{j\in N(|F_i|)}\phi_{\hat{\mathsf{P}}_{n(O_j)}}$$

Then, $\hat{\boldsymbol{\sigma}}^2 = \frac{1}{n^2}\sum_{i,j}\mathbf{F}_{ij}\varphi_i\varphi_j \underset{p}{\rightarrow} \boldsymbol{\sigma}^2$

Cross-fitting in dependent data

The usual "as-independent" cross-fitting still eliminates empirical process bias (Balkus et al. 2026b) even under various correlation structure (network, cluster, time-series, etc.)

$$\mathsf{P}_n\phi_{\hat\eta} = \underbrace{\mathsf{P}_n\phi_{\eta_0}}_{\text{CLT}} + \underbrace{\mathsf{P}(\phi_{\hat\eta} - \phi_{\eta_0})}_{\text{Nuisance product}} + \underbrace{(\mathsf{P}_n - \mathsf{P})(\phi_{\hat\eta} - \phi_{\eta_0})}_{\text{Empirical process, } o_{\mathsf{P}}(1/C_n)}$$

- Cross-fit empirical mean always unbiased
- If CLT holds pointwise over correlated units, then variance shrinks appropriate rate
- Therefore, $(\mathsf{P}_n - \mathsf{P})(\phi_{\hat\eta} - \phi_{\eta_0}) = o_{\mathsf{P}}(1/C_n)$

Estimation framework

1. Fit estimators $\hat{m}$ and $\hat{r}$ of nuisance parameters m and r via cross-fitting and super (ensemble machine) learning (Davies and van der Laan 2016; van der Laan et al. 2007).
2. Construct one-step or "network-TMLE" estimators (Zivich et al. 2022) from an estimated EIF based on $\hat{m}$ and $w \cdot \hat{r}$ (weighted density ratio)
3. Compute standard error and construct Wald-style confidence intervals based on empirical variance of the estimated EIF

Summary of advances

Existing approaches like Ogburn et al. (2022) require…

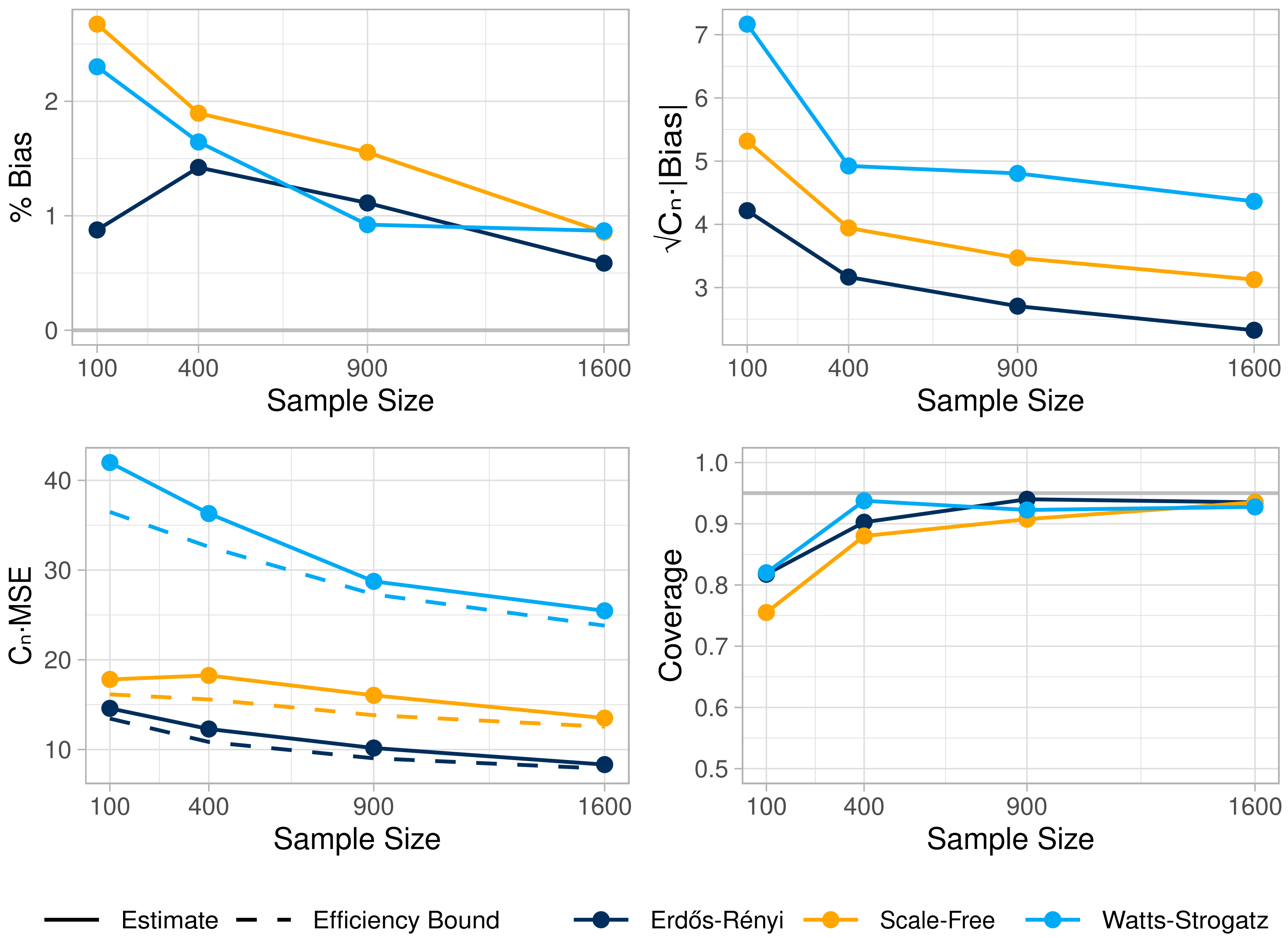
- Parametric regression (i.e. GLMs) to estimate nuisances
- Estimating a conditional mixture density of neighbor's exposure
- Repeating the TMLE procedure over many Monte Carlo samples, if the stochastic intervention is non-degenerate (**Sofrygin and Laan 2016**)

Our approach (**Balkus et al. 2026a**) builds on this by introducing…

- Tractable nuisance parameters estimable by nonparametric regression
- Avoids mixture density estimation
- No need for computationally-intensive Monte Carlo procedures
- New theory for cross-fitting under dependent data; eliminates entropy conditions

Empirical results

Asymptotic properties of Network-TMLE



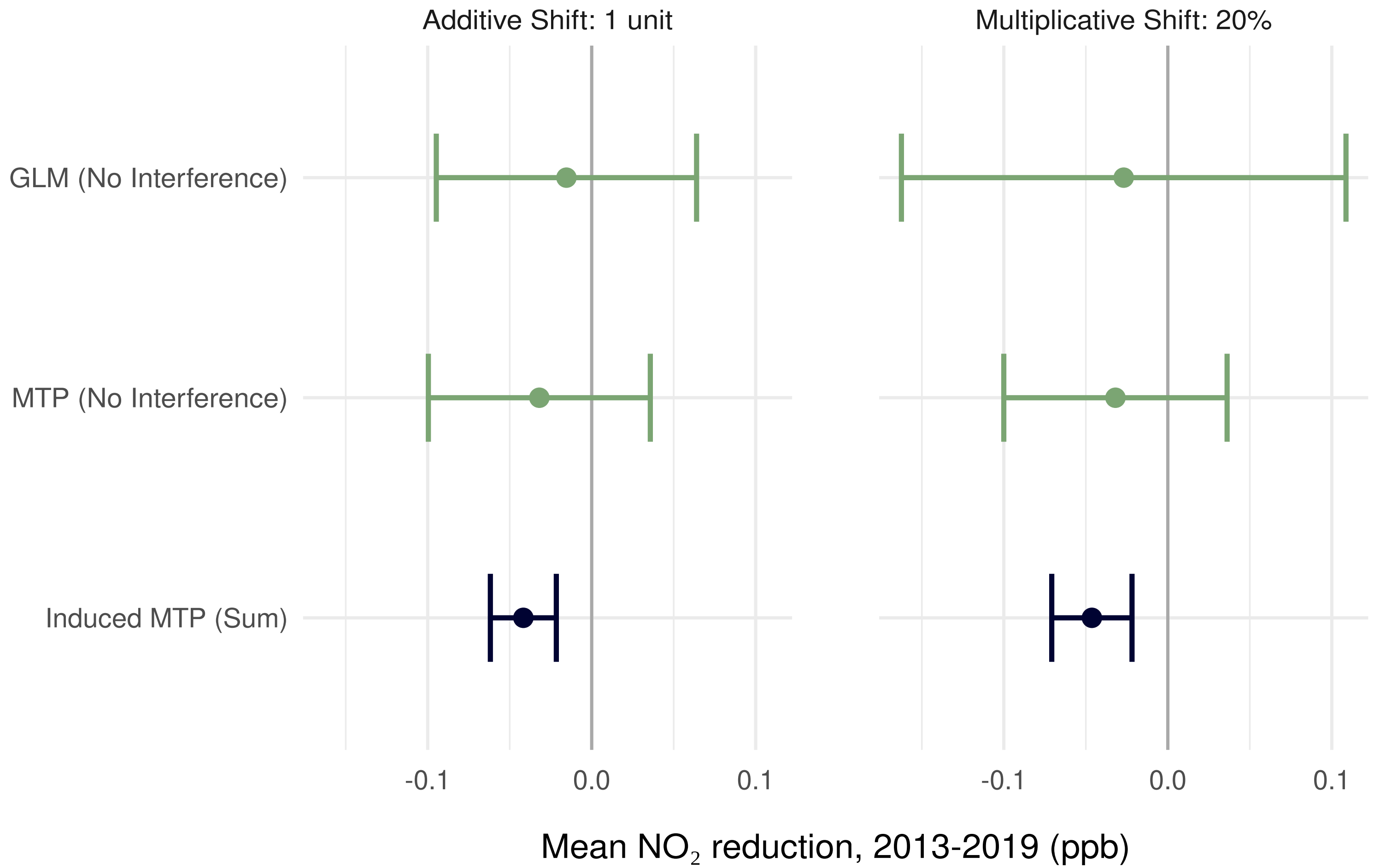
Versus competing methods on semisynthetic data

- Simulate A and Y as linear models from 16 socioeconomic and land-use ZIP-code level covariates from the ZEV-NO_2 California dataset.
- *How poor would estimates be if only mistake were ignoring interference?*

Method	Learner	% Bias	Variance	Coverage
Network-TMLE	Correct GLM	0.11	1.56	96.2%
Network-TMLE	Super Learner	1.03	1.56	94.0%
IID-TMLE	Correct GLM	20.42	2.11	54.8%
Linear Regression	—	20.62	2.12	55.0%

Data analysis

Effect of electric vehicles on NO_2 in California



- GLM (ignores interference): **ZEVs reduce NO_2 by 0.015 ppb**, totaling ~2.5% of average change in NO_2
- Induced MTP: **ZEVs reduce NO_2 by 0.042 ppb**, totaling ~7% of average change in NO_2

Further work

Further work

- General convergence results for nuisance learners under correlated data
- Estimating conditional density ratio nuisance $\boldsymbol{r}$ still may be difficult
  - → Amenable to undersmoothing or "Riesz learning"
- If summaries $\boldsymbol{s}$ unknown, can we learn them automatically?
- Same theory of the Longitudinal MTP (**Díaz et al. 2021**) should extend when reduced; useful for time-varying setting

Thank you! Questions?

salbalkus.github.io

@salbalkus @nshlab

DOI: 10.1093/jrsssb/qkag052

Funded by NIEHS T32 ES007142 and NSF DGE 2140743

References

Aronow, P. M., and Samii, C. (2017), "Estimating average causal effects under general interference, with application to a social network experiment," *The Annals of Applied Statistics*, Institute of Mathematical Statistics, 11. https://doi.org/10.1214/16-aoas1005.

Balkus, S. V., Delaney, S. W., and Hejazi, N. S. (2026a), "The causal effects of modified treatment policies under network interference," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, Oxford University Press (OUP). https://doi.org/10.1093/jrsssb/qkag052.

Balkus, S. V., Laith, H., and Hejazi, N. S. (2026b), "On the use of cross-fitting in causal machine learning with correlated units," *arXiv preprint arXiv:2601.10899*. https://doi.org/10.48550/ARXIV.2601.10899.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993), *Efficient and adaptive estimation for semiparametric models*, Springer.

Davies, M. M., and van der Laan, M. J. (2016), "Optimal spatial prediction using ensemble machine learning," *The International Journal of Biostatistics*, Walter de Gruyter GmbH, 12, 179–201. https://doi.org/10.1515/ijb-2014-0060.

Díaz, I., Williams, N., Hoffman, K. L., and Schenck, E. J. (2021), "Nonparametric causal effects based on longitudinal modified treatment policies," *Journal of the American Statistical Association*, Informa UK Limited, 118, 846–857. https://doi.org/10.1080/01621459.2021.1955691.

Díaz, I., and van der Laan, M. J. (2012), "Population intervention causal effects based on stochastic interventions," *Biometrics*, Wiley Online Library, 68, 541–549. https://doi.org/10.1111/j.1541-0420.2011.01685.x.

Haneuse, S., and Rotnitzky, A. (2013), "Estimation of the effect of interventions that modify the received treatment," *Statistics in Medicine*, Wiley, 32, 5260–5277. https://doi.org/10.1002/sim.5907.

Hudgens, M. G., and Halloran, M. E. (2008), "Toward causal inference with interference," *Journal of the American Statistical Association*, Informa UK Limited, 103, 832–842. https://doi.org/10.1198/016214508000000292.

Negro, L. (2022), "Sample distribution theory using coarea formula," *Communications in Statistics - Theory and Methods*, Informa UK Limited, 53, 1864–1889. https://doi.org/10.1080/03610926.2022.2116284.

Ogburn, E. L., Sofrygin, O., Díaz, I., and Laan, M. J. van der (2022), "Causal inference for social network data," *Journal of the American Statistical Association*, Informa UK Limited, 119, 597–611. https://doi.org/10.1080/01621459.2022.2131557.

Pfanzagl, J., and Wefelmeyer, W. (1985), "Contributions to a general asymptotic statistical theory," *Statistics & Risk Modeling*, 3, 379–388.

Sofrygin, O., and Laan, M. J. van der (2016), "Semi-parametric estimation and inference for the mean outcome of the single time-point intervention in a causally connected population," *Journal of Causal Inference*, Walter de Gruyter GmbH, 5. https://doi.org/10.1515/jci-2016-0003.

van der Laan, M. J. (2014), "Causal inference for a population of causally connected units," *Journal of Causal Inference*, Walter de Gruyter GmbH, 2, 13–74. https://doi.org/10.1515/jci-2013-0002.

van der Laan, M. J., Polley, E. C., and Hubbard, A. E. (2007), "Super learner," *Statistical Applications in Genetics and Molecular Biology*, De Gruyter, 6. https://doi.org/10.2202/1544-6115.1309.

van der Laan, M. J., and Rose, S. (2011), *Targeted learning: Causal inference for observational and experimental data*, Springer. https://doi.org/10.1007/978-1-4419-9782-1.

van der Laan, M. J., and Rubin, D. (2006), “Targeted maximum likelihood learning,” *The International Journal of Biostatistics*, De Gruyter, 2. https://doi.org/10.2202/1557-4679.1043.

Zivich, P. N., Hudgens, M. G., Brookhart, M. A., Moody, J., Weber, D. J., and Aiello, A. E. (2022), “Targeted maximum likelihood estimation of causal effects with interference: A simulation study,” *Statistics in Medicine*, Wiley, 41, 4554–4577. https://doi.org/10.1002/sim.9525.

Appendix

Simulation study: Data-generating process

Draw 400 iterations, estimate effect of MTP based on

$$\mathbf{L}_1 \sim \text{Beta}(3,2); \mathbf{L}_2 \sim \text{Poisson}(100); \mathbf{L}_3 \sim \text{Gamma}(2,4); \mathbf{L}_4 \sim \text{Bernoulli}(0.6)$$

$$m_L = \Big(1 + L_4\Big) \cdot \Big(-2(\mathbb{I}(L_1 > 0.3) + \mathbb{I}(L_2 > 90) + \mathbb{I}(L_3 > 5)) - (\mathbb{I}(L_1 > 0.5) + \mathbb{I}(L_2 > 100) + \mathbb{I}(L_3 > 10)) + 2(\mathbb{I}(L_1 > 0.7) + \mathbb{I}(L_2 > 110) + \mathbb{I}(L_3 > 15))\Big)$$

$$\mathbf{A} \sim \text{Normal}(m_L - 5, 1.0) \text{ and } \mathbf{A}^s = \Big[\sum_{j \in F_i} A_i\Big]_{i=1}^{n}$$

$$m_A = -2\mathbb{I}(A > -2) - \mathbb{I}(A > 1) + 3\mathbb{I}(A > 3); m_{A_s} = 3\mathbb{I}(A_s > 0) + \mathbb{I}(A_s > 6) + \mathbb{I}(A_s > 12)$$

$$\mathbf{Y} \sim \text{TruncNormal}(m_L \cdot (1 + 0.2 m_A + m_{A_s}) + 5, 2.0) \ ,$$

Effect of ZEV on $\mathbf{NO_2}$