Transportability of Principal Causal Effects

Jared D. Huling

Division of Biostatistics *and Health Data Science*
University of Minnesota

Work with

Justin Clark

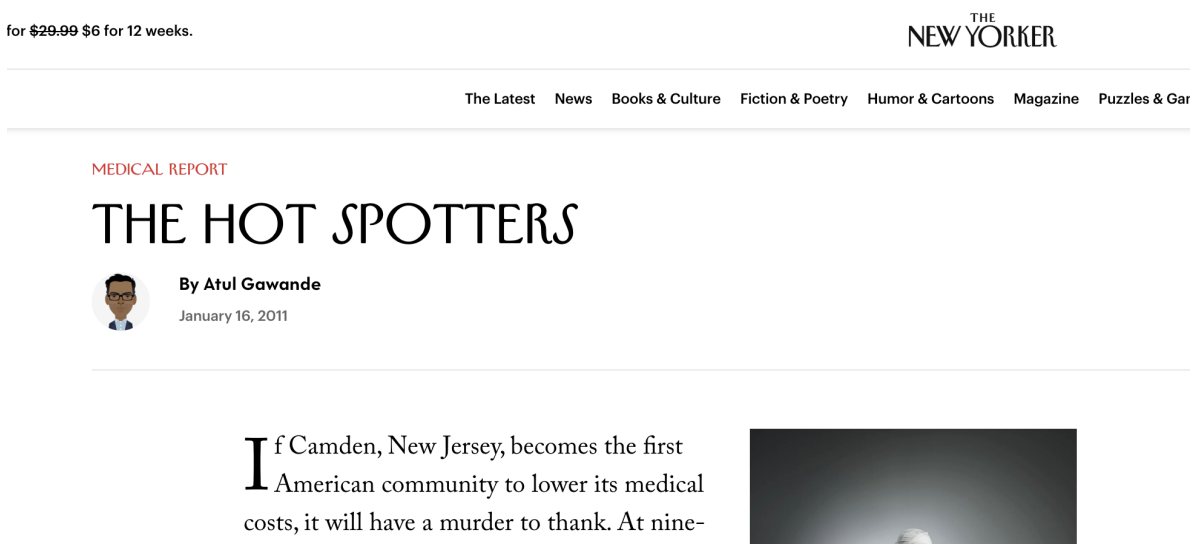
Ohio State University Department of Statistics

Acknowledgments

Justin Clark, PhD
(Analysis Group)

Healthcare Hotspotting for “superutilizers”

- “Superutilizers”: 50% of healthcare expenditures, yet needs often unmet
- Effective intervention for them could be a game-changer
- “Hotspotting” intervention for superutilizers
  - Includes intensive clinical and social components
  - Patients engaged by multidisciplinary team to address individual needs



Healthcare Hotspotting: a failure?

The NEW ENGLAND JOURNAL of MEDICINE

SPECIAL ARTICLE

Health Care Hotspotting — A Randomized, Controlled Trial

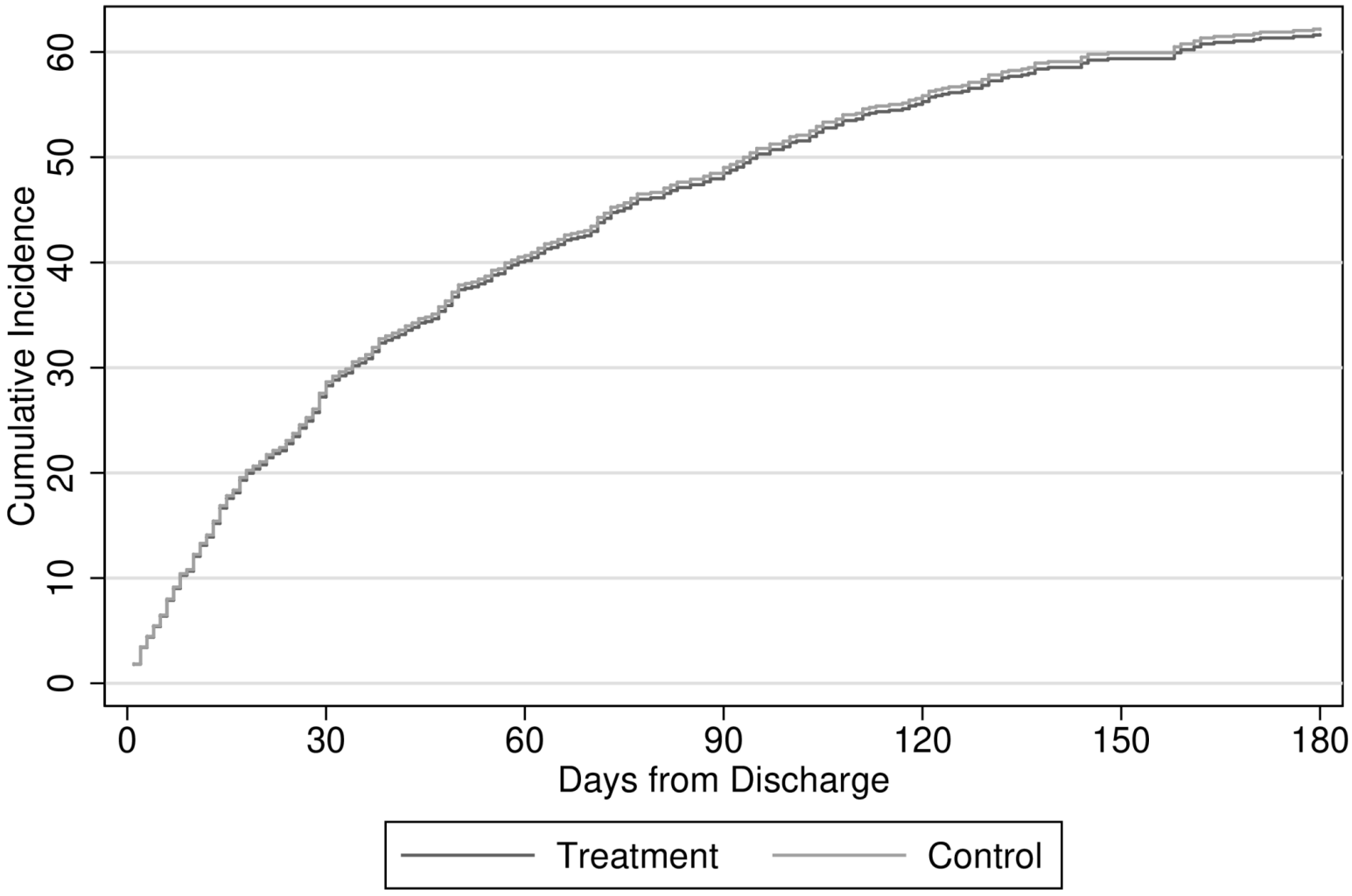
Amy Finkelstein, Ph.D., Annetta Zhou, Ph.D., Sarah Taubman, Sc.D., and Joseph Doyle, Ph.D.

https://www.nejm.org/doi/10.1056/NEJMsa1906848

- Failed to find differences in readmission rates between the two intervention arms

Healthcare Hotspotting: a failure?

Estimated cumulative incidence of readmission within 180 days



Source: Finkelstein, et al (2015)

Hotspotting may work for “high engagers”

Original Investigation | Health Policy

Hospital Readmissions by Variation in Engagement in the Health Care Hotspotting Trial

A Secondary Analysis of a Randomized Clinical Trial

Qiang Yang, PhD; Dawn Wiest, PhD; Anna C. Davis, PhD; Aaron Truchil, MS; John L. Adams, PhD

https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2809198

- Yang et al. stratified population by baseline probability of adherence
- Found effect among those with high baseline probability of engaging with hotspotting

Should a health system implement hotspotting?

- Benefit of health system interventions vary
  –often substantially–
  with patient characteristics (Huling et al., 2019)

- For hotspotting, engagement with intervention seems to drive much of this heterogeneity

- Patient mix of Camden trial does not represent typical health system
  - Mix of those who benefit?
  - Mix of high engagers?

- A *different* health system may have different patterns of patient engagement

Should a health system implement hotspotting?

Questions:

Would high engagers in your health system benefit?

***Practical**: should you invest in running an RCT for hotspotting that targets high engagers?*

Methodological challenges

- Need to account for population differences between Camden and new health system
- Need to account for engagement effects and possibly different engagement behavior in new health system

Generalization and nonadherence

- Methods for **generalizability/transportability of causal effects** important when effects heterogeneous: *average effect in new population may be different*

- Yet, methods for **generalizability** to new populations *implicitly assume adherence patterns identical in new population*

Transporting the ITT effect

- In RCTs, we typically estimate the **intention-to-treat** (ITT) effect, i.e., the effect of treatment *assignment*

- $Y(a)$: potential outcome under *assignment* to treatment $A = a$, not necessarily *receipt* or *use* of treatment $A = a$ ($a = 0$: control, $a = 1$: treatment).

- ITT effect in target population:

$$\mathbb{E}[Y(1) - Y(0) | \underbrace{R = 0}_{\text{in target population}}]$$

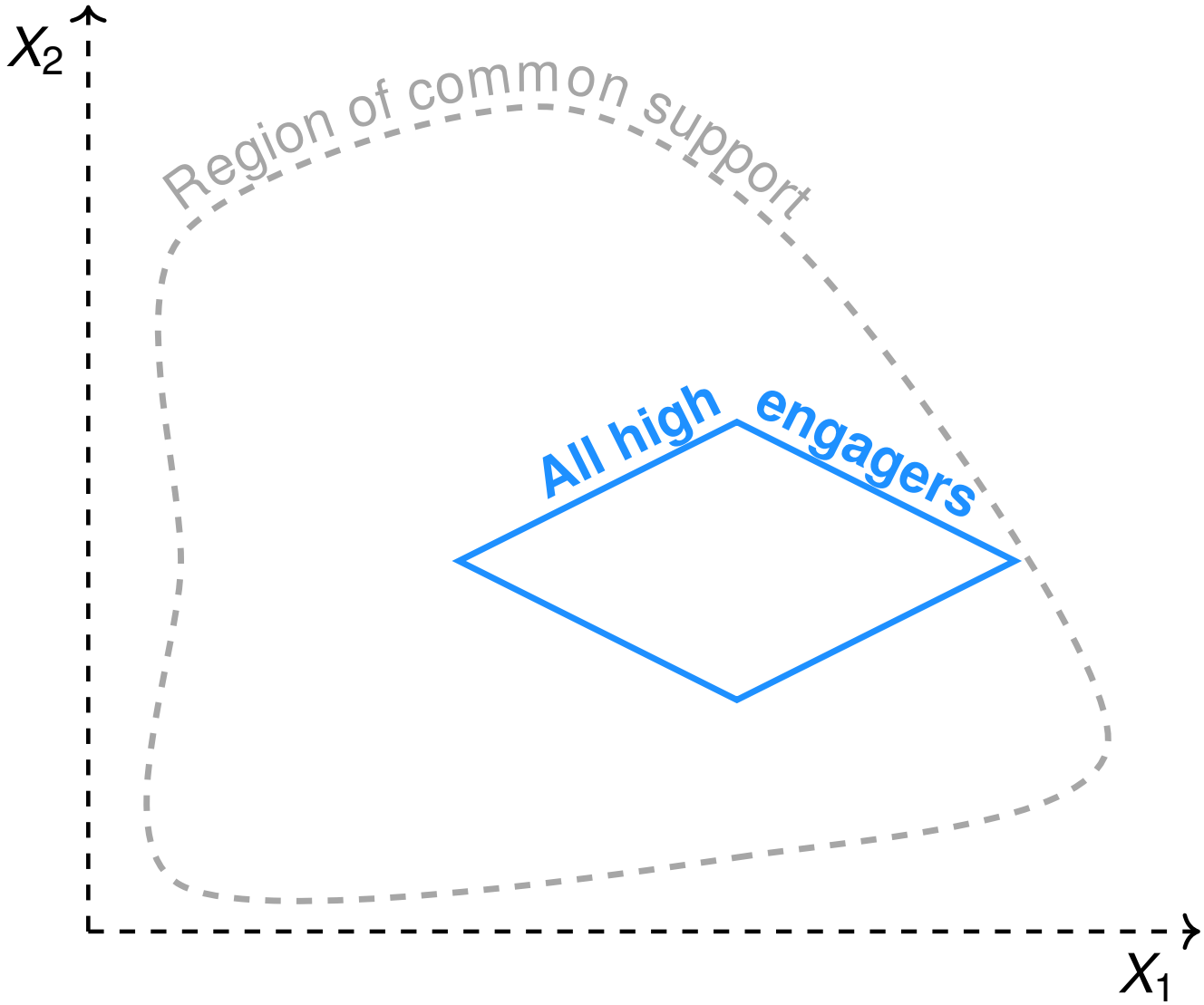
- May not reflect expected adherence in target pop'n

The big picture

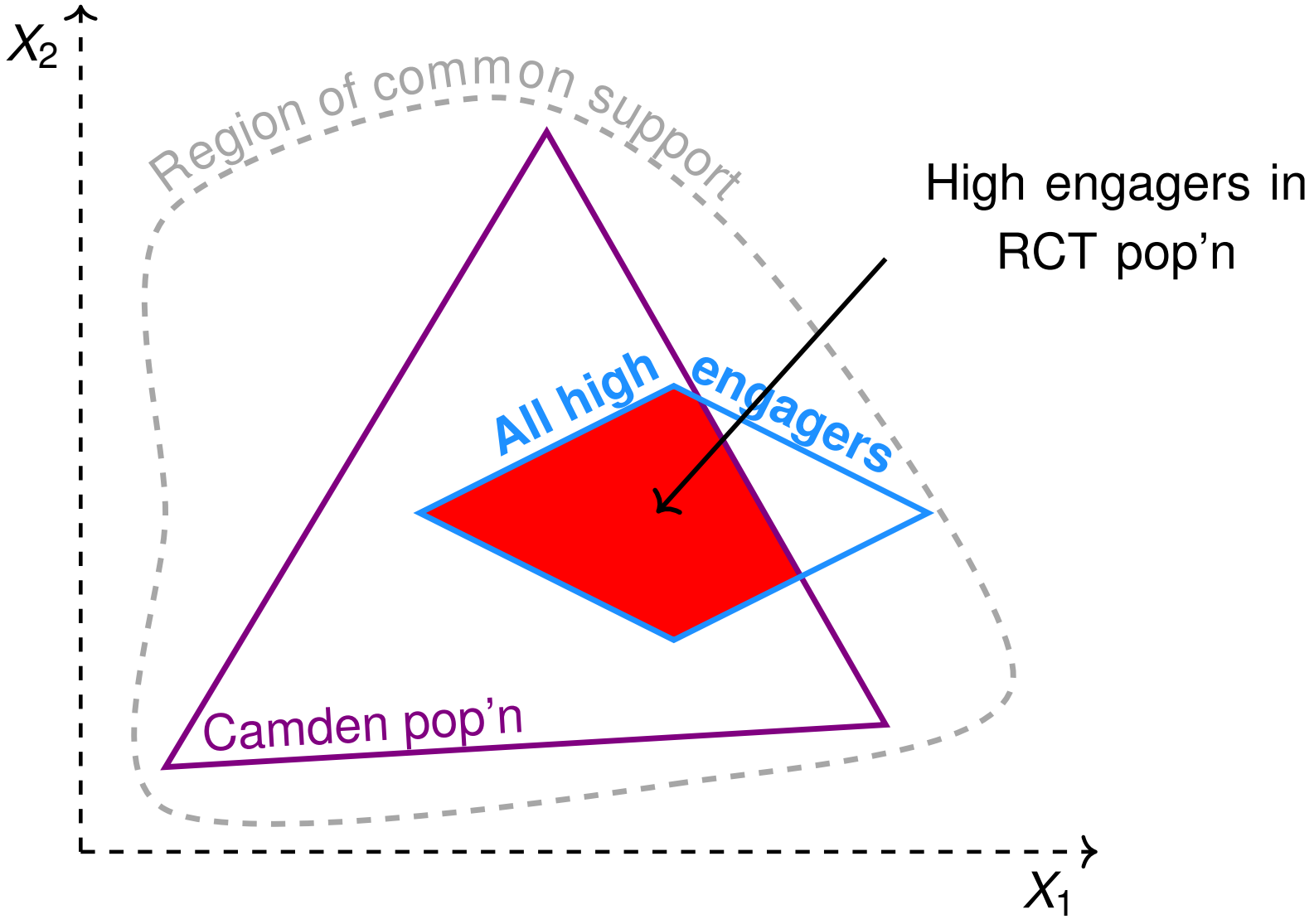
- **More *refined estimands* when generalizing to a new population often necessary**
  - E.g. generalizing effects among those who take treatment when assigned

- **The focus of this work more generally:**
  a "principal stratification" framework for estimands in subsets of new populations defined by post-assignment events
  - effects among "compliers", survivors, surrogate endpoints, etc

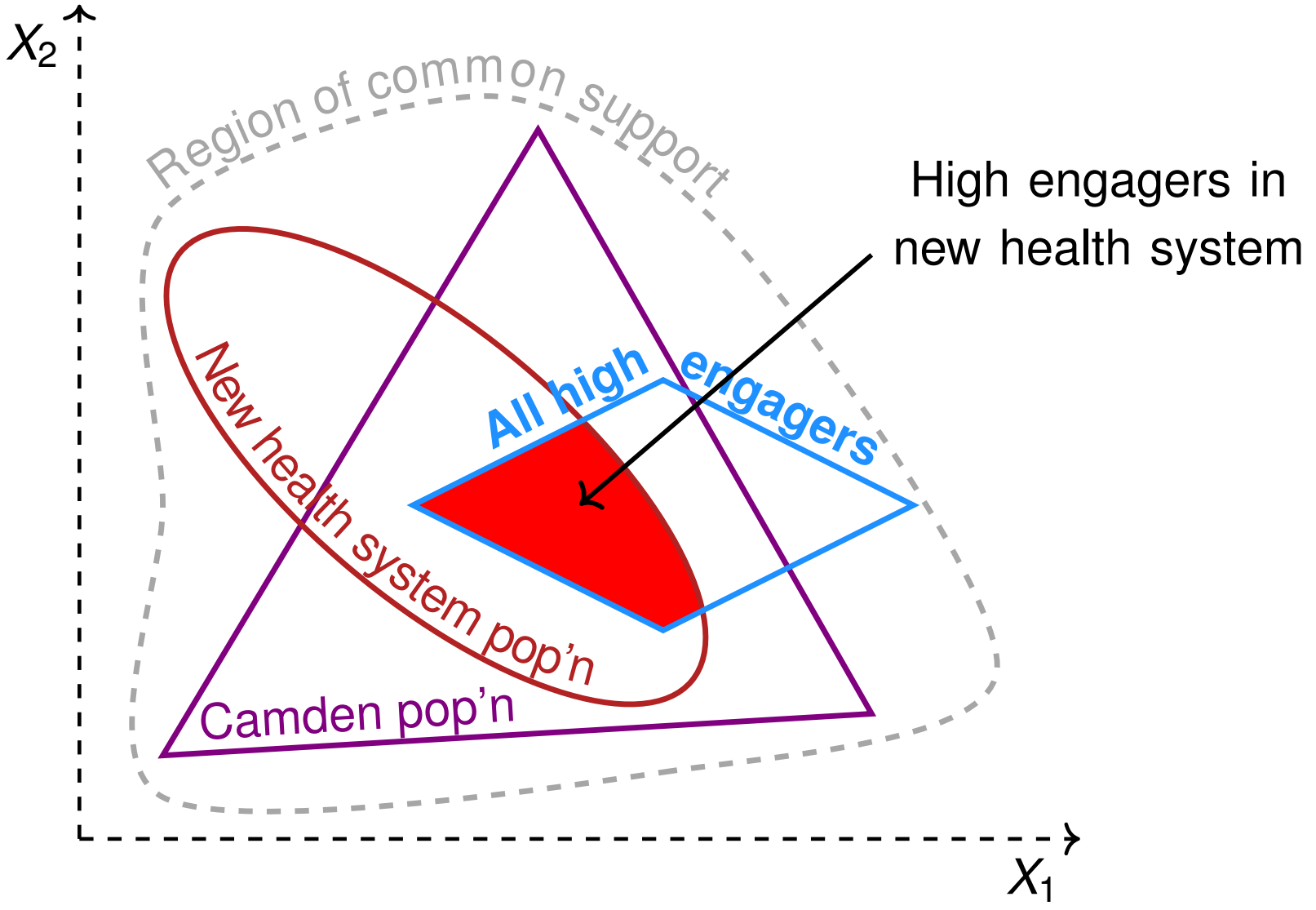
Stylistic depiction of high engagers population



Stylistic depiction of high engagers in RCT



Stylistic depiction of target population



Key challenges

1. How to formally define “compliers”?

2. What assumptions needed to identify compliers in observed data?

3. What assumptions needed to transport complier effects to a new population?

Background: Compliance

Principal Stratification: Frangakis and Rubin (2002) offers one approach to deal with compliance

Basic idea:

Trial participants exhibit latent compliance patterns, which are well-defined at baseline.

- We can then focus on effects within these strata.

Background: principal stratification

$(Y, C, A, X) \overset{i.i.d.}{\sim} P$

- Y: Outcome
- A: Treatment assignment
- X: Baseline covariates
- C: Treatment received

  $C = 1$: Receipt of active treatment

  $C = 0$: Non-receipt of treatment.

Background: principal stratification

Potential outcomes:

- $Y(1), Y(0)$
- $C(1), C(0)$

Units categorized into what treatment they'll take under being assigned/not assigned treatment:

	$C(1) = 0$	$C(1) = 1$
$C(0) = 0$	Never-takers	Compliers
$C(0) = 1$	Defiers	Always-Takers

Background: principal stratification

Treatment effect among compliers:

$$\mathbb{E}[Y(1) - Y(0) | \underbrace{C(1) = 1}_{\substack{\text{those who take} \\ \text{trt when assigned}}}, \underbrace{C(0) = 0}_{\substack{\text{those who don't} \\ \text{take trt when not} \\ \text{assigned}}}]$$

- Conditioning on $C(1)/C(0)$ okay, because they're well-defined *prior* to randomization

Background: principal stratification

The observed groups are a ***mixture*** of compliance patterns:

	$C = 0$	$C = 1$
$A = 0$	$C(0) = 0, C(1) \in (1, 0)$	$C(0) = 1, C(1) \in (1, 0)$
$A = 1$	$C(1) = 0, C(0) \in (1, 0)$	$C(1) = 1, C(0) \in (1, 0)$

Adapted from Jiang et al., *Multiply robust estimation of causal effects under principal ignorability* (2022)

Background: principal stratification

Treatment effect among compliers:

$$\mathbb{E}[Y(1) - Y(0) | C(1) = 1, C(0) = 0]$$

We *can't* identify this using **observed** compliance:

$$\mathbb{E}[Y| \underbrace{A = 1, C = 1}_{C(1)=1, C(0)\in(1,0)}] - \mathbb{E}[Y| \underbrace{A = 0, C = 0}_{C(1)\in(1,0), C(0)=0}]$$

because this compares two different populations

Background: assumptions

- **Methods based on parametric assumptions**
  - brittle, as they critically rely on parametric models, at great risk of bias due to misspecification
- **Methods based on exclusion restriction assumption**
  - require no direct effect of assignment
  - IV methods (Angrist et al., 1996), (Rudolph and Laan, 2017)
- **Methods based on principal ignorability assumption**
  - require baseline covariates to fully capture compliance status
  - use *principal scores* (conditional prob of being in principal strata) for identification (Jo and Stuart, 2009; Follmann, 2000; Ding and Lu, 2017)

Background: assumptions

- **Exclusion restriction may not hold in single-blind trials**
  - E.g., Hirano et al. (2000)
  - For hotspotting, *knowing* one is being given special care may change behavior/influence outcomes

- **We use a principal ignorability-based approach**
  - These assumptions will not always hold but *may* be plausible in trials with richly-collected data
  - Sensitivity analyses important

Our setting: observed data

$$(R \times Y, R \times A, R \times C, X, R) \overset{i.i.d.}{\sim} P$$

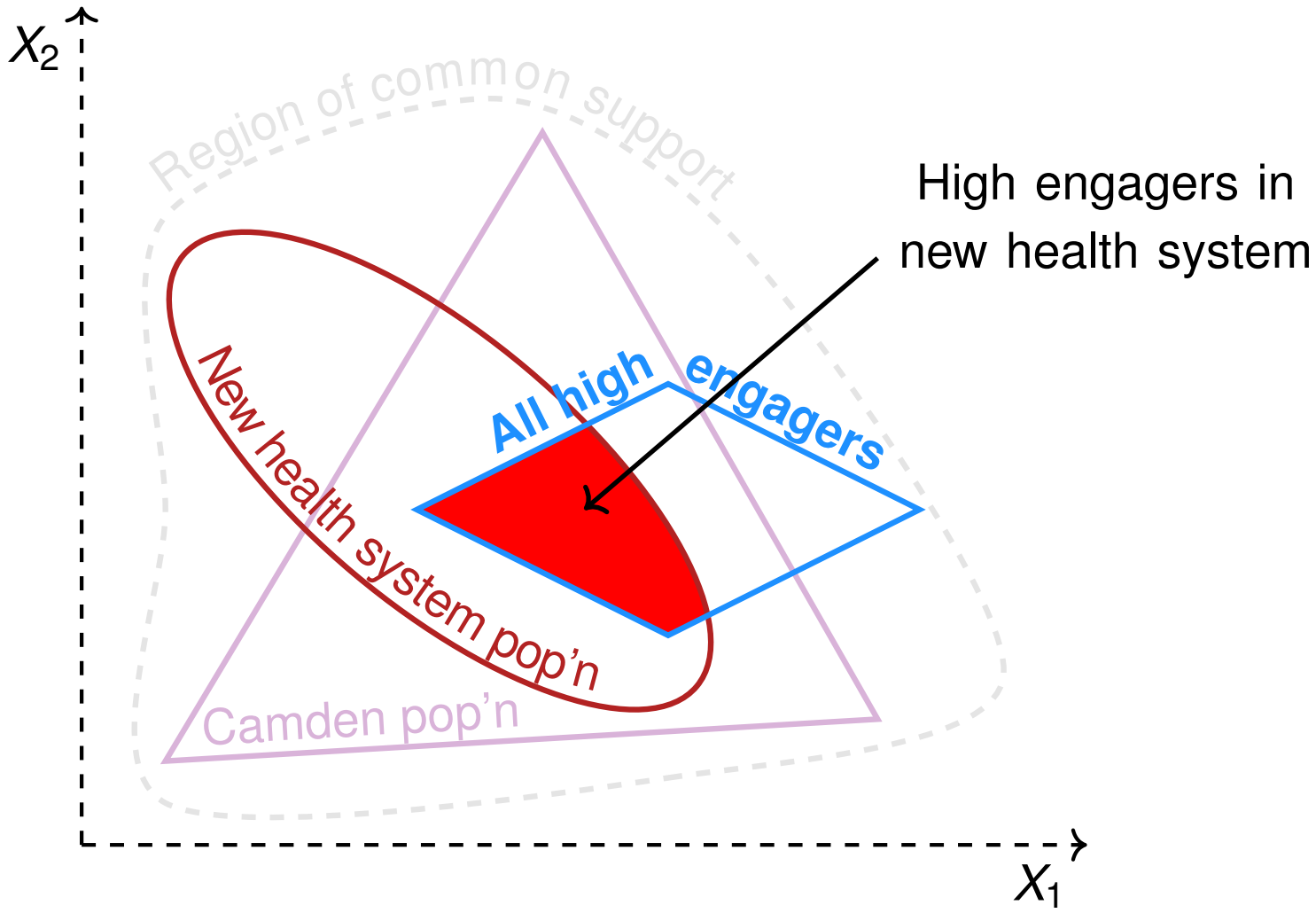
- R: Trial participation indicator
  ($= 1$: participated in trial, $= 0$ in target)
- Y: Outcome
- A: Treatment assignment
- C: Treatment received
  $C = 1$: Receipt of active treatment
  $C = 0$: Non-receipt of active treatment.
- X: Baseline covariates

Our setting: estimand

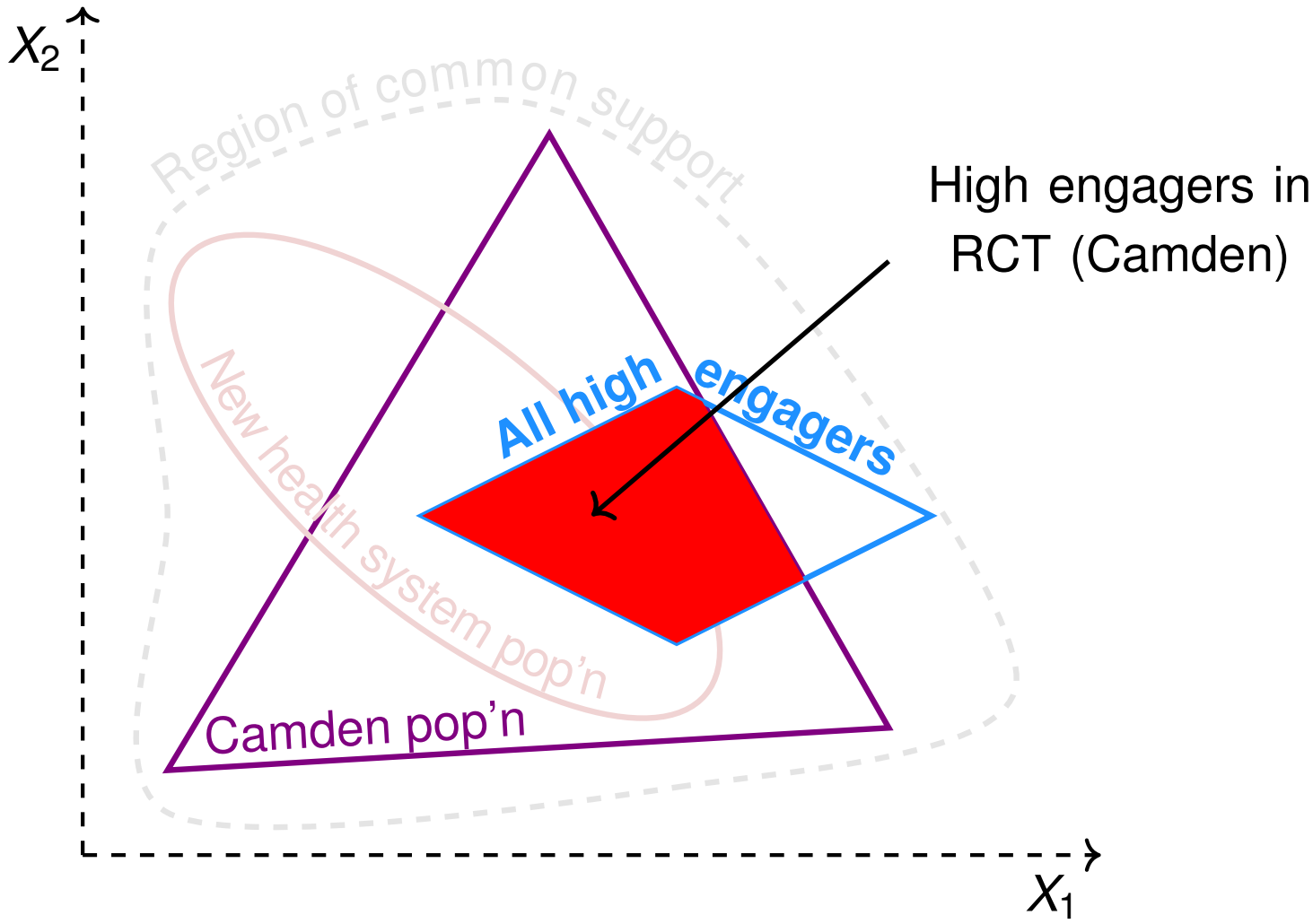
Transported principal causal effect:

$$\tau_{10}^{0} = \mathbb{E}[Y(1) - Y(0) | \underbrace{C(1) = 1, C(0) = 0}_{\text{among compliers}}, \underbrace{R = 0}_{\substack{\text{in the target} \\ \text{population}}}]$$

Stylistic depiction of compliers in target pop'n

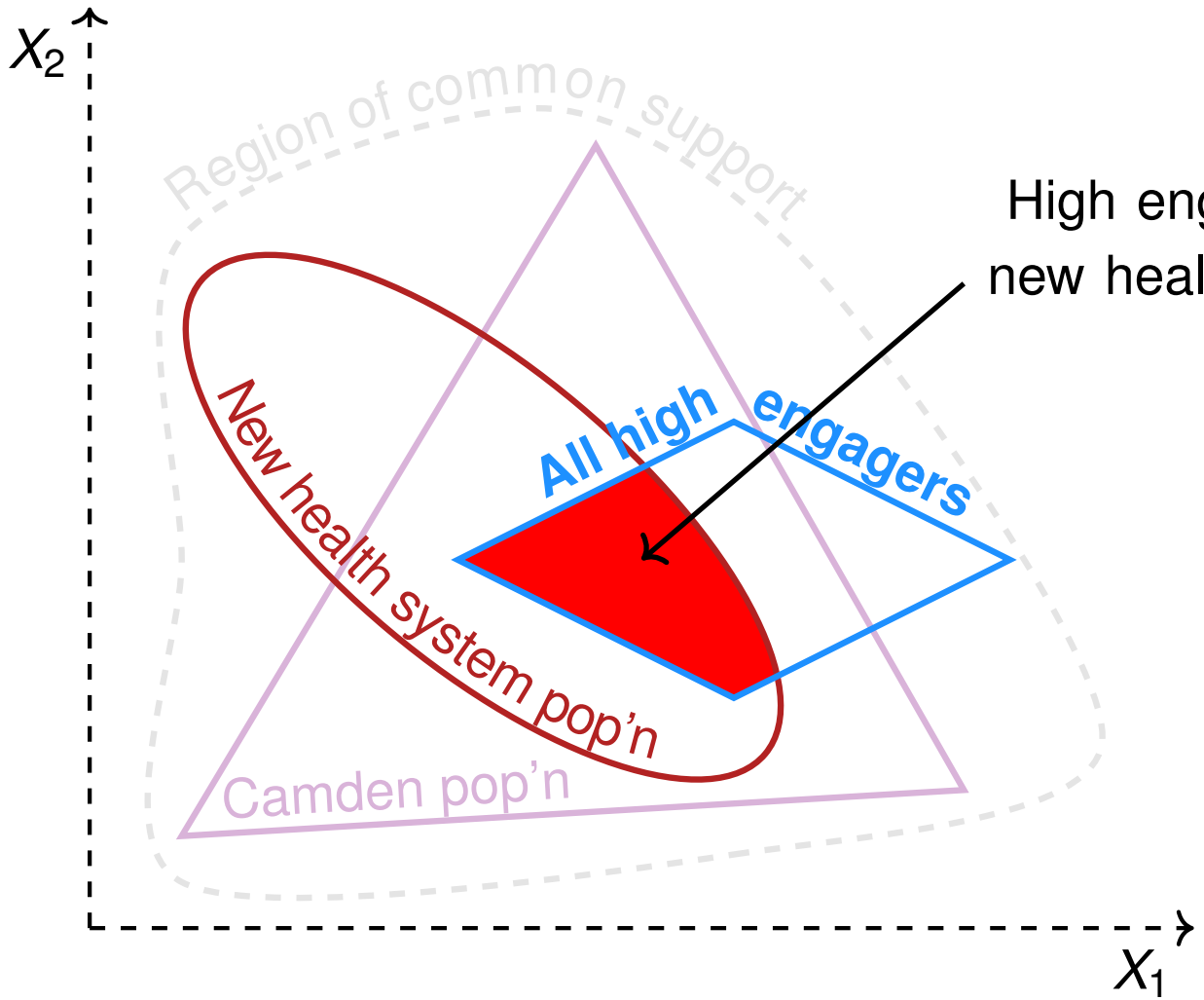


Stylistic depiction of compliers in RCT



$$\mathbb{E}[Y(1) - Y(0) | \underbrace{C(1) = 1, C(0) = 0}_{\text{among compliers}}, \underbrace{R = 1}_{\text{in the RCT population}}]$$

Stylistic depiction of compliers in target pop'n



$$\tau^0_{10} = \mathbb{E}[Y(1) - Y(0) | \underbrace{C(1) = 1, C(0) = 0}_{\text{among compliers}}, \underbrace{R = 0}_{\text{in the target population}}]$$

Assumptions for Transporting Principal Effects

Key question: Under what assumptions can we write

$$\begin{aligned}\tau_{10}^{0} &= \mathbb{E}[Y(1) - Y(0) | C(1) = 1, C(0) = 0, R = 0] \\ &= \mathbb{E}[Y(1) - Y(0) | \underbrace{U = 10}_{\text{compliers}}, R = 0]\end{aligned}$$

as a function of observed data?

Note: $U = c_1 c_0$ is shorthand for $C(1) = c_1$ and $C(0) = c_0$ (as in Ding and Lu (2017))

Assumptions for Transporting Principal Effects

1. Consistency:

$$Y_i = Y_i(1)A_i + Y_i(0)(1 - A_i)$$
$$C_i = C_i(1)A_i + C_i(0)(1 - A_i)$$

Follows from SUTVA. Connects potential outcomes and compliance behavior to observed outcomes and compliance.

2. Treatment Ignorability:

$$A \perp\!\!\!\perp (Y(1), Y(0), C(1), C(0))|X$$

Guaranteed by randomization.

Assumptions for Transporting Principal Effects

3. Monotonicity:

$$C(1) \geq C(0)$$

Rules out "defiers", i.e., trial participants who always take the opposite of the treatment assigned.

4. Principal Ignorability:

$$\mathbb{E}[Y(1)|U = 10, R = 1, X = x] = \mathbb{E}[Y(1)|U = 11, R = 1, X = x]$$
$$\mathbb{E}[Y(0)|U = 00, R = 1, X = x] = \mathbb{E}[Y(0)|U = 10, R = 1, X = x]$$

X is rich enough that conditional means same across some strata. Strong assumption!

Together, 3. and 4. allow one to identify mixtures in the RCT

Assumptions for Transporting Principal Effects

5. Exchangeability of Principal Strata:

$$R \perp\!\!\!\perp (C(1), C(0)) \mid X$$

Conditional on X, the distribution of latent compliance patterns is identical between the trial and target populations.

6. Mean Exchangeability in Principal Strata:

$$\begin{aligned} &\mathbb{E}[Y(a)|U = c_1c_0, R = 1, X = x] \\ &= \mathbb{E}[Y(a)|U = c_1c_0, R = 0, X = x] \end{aligned}$$

X rich enough to explain any differences in mean outcomes across trial/target populations within strata.

5+6 allow RCT data to identify compliers in target pop'n

Identification using outcome and principal score

Under these assumptions, we have:

$$\tau_{10}^{0} = \mathbb{E}[Y(1) - Y(0)|U = 10, R = 0]$$

$$= \frac{\mathbb{E}\Big[\overbrace{\{p_1(X) - p_0(X)\}}^{=P(U=10|X)}(1-R)\{\mu_{11}(X) - \mu_{00}(X)\}\Big]}{\underbrace{\mathbb{E}\big[\{p_1(X) - p_0(X)\}(1-R)\big]}_{=P(U=10,R=0)}}$$

with:

- $p_a(X) = P(C = 1|A = a, R = 1, X)$
- $\mu_{ac}(X) = \mathbb{E}[Y|A = a, C = c, R = 1, X]$

where from Ding and Lu (2017) the conditional prob of being a complier is identified as:

$$P(U = 10|X) = p_1(X) - p_0(X)$$

Weighted estimator identification

Under additional positivity assumptions:

$$\tau^0_{10} = \mathbb{E}[w_1(C, A, R, X)Y] - \mathbb{E}[w_0(C, A, R, X)Y]$$

with:

$$w_1(C, A, R, X) \propto$$

$$C \cdot A \cdot R \underbrace{\frac{p_1(X) - p_0(X)}{p_1(X)}}_{\text{reweight } C=1 \text{ group in arm } A=1 \text{ to compliers}} \underbrace{\frac{1}{\pi(X)}}_{\text{reweight to RCT sample}} \underbrace{\frac{1-\rho(X)}{\rho(X)}}_{\text{transp RCT to target}}$$

where

$$\pi(X) = P(A = 1 | R = 1, X)$$

$$\rho(X) = P(R = 1 | X)$$

Weighted estimator identification

Under additional positivity assumptions:

$$\tau_{10}^{0} = \mathbb{E}[w_1(C, A, R, X)\, Y] - \mathbb{E}[w_0(C, A, R, X)\, Y]$$

with:

$$w_0(C, A, R, X) \propto$$

$$(1 - C) \cdot (1 - A) \cdot R$$

$$\times \underbrace{\frac{p_1(X) - p_0(X)}{1 - p_0(X)}}_{\text{reweight } C = 0 \text{ group in arm } A = 0 \text{ to compliers}} \quad \underbrace{\frac{1}{1 - \pi(X)}}_{\text{reweight to RCT sample}} \quad \underbrace{\frac{1 - \rho(X)}{\rho(X)}}_{\text{transp RCT to target}}$$

Estimation

Target parameter:

$$\tau_{10}^{0} = \mathbb{E}[Y(1) - Y(0) | U = 10, R = 0]$$

How should we estimate it?

One option:

$$\hat{\tau}_{\text{OM}} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{p}_1(X_i) - \hat{p}_0(X_i)\right)(1 - R_i)\left\{\hat{\mu}_{11}(X_i) - \hat{\mu}_{00}(X_i)\right\}}{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{p}_1(X_i) - \hat{p}_0(X_i)\right)(1 - R_i)}$$

Another option:

Use a plug-in estimate $\hat{\tau}_{\text{IPW}}$ based on the *weighting* identification.

Issues in estimation

- $\hat{\tau}_{\text{OM}}$ and $\hat{\tau}_{\text{IPW}}$ both require estimation of several nuisance parameters
  - $p_1(\cdot)$, $p_0(\cdot)$, $\mu_{11}(\cdot)$, and $\mu_{00}(\cdot)$ for $\hat{\tau}_{\text{OM}}$
  - $p_1(\cdot)$, $p_0(\cdot)$, $\pi(\cdot)$, and $\rho(\cdot)$ for $\hat{\tau}_{\text{IPW}}$

- **Can posit parametric models:**
  - root-n convergence and asymptotic normality if all models are correctly-specified
  - otherwise: all bets off

Issues in estimation

- $\hat{\tau}_{\text{OM}}$ and $\hat{\tau}_{\text{IPW}}$ both require estimation of several nuisance parameters
  - $p_1(\cdot)$, $p_0(\cdot)$, $\mu_{11}(\cdot)$, and $\mu_{00}(\cdot)$ for $\hat{\tau}_{\text{OM}}$
  - $p_1(\cdot)$, $p_0(\cdot)$, $\pi(\cdot)$, and $\rho(\cdot)$ for $\hat{\tau}_{\text{IPW}}$

- **Can use *nonparametric* methods (ML, etc):**
  - Consistency can be guaranteed under fairly mild conditions
  - Yet, root-*n* convergence/asymptotic normality impossible under these conditions

Estimation

Can we do better?

Efficient estimation via EIFs

The Efficient Influence Function (EIF) provides useful information for estimation of a target parameter:

- The EIF provides an efficiency bound: its variance is the minimum possible *nonparametrically*
- The EIF can be used to derive efficient estimators
- EIF-based estimators tend to have relaxed reliance on nuisance parameter estimates

Efficient estimation via EIFs

Recall our parameter

$$\tau_{10}^{0} = \frac{\mathbb{E}\left[\{p_1(X) - p_0(X)\}(1-R)\{\mu_{11}(X) - \mu_{00}(X)\}\right]}{\mathbb{E}\left[\{p_1(X) - p_0(X)\}(1-R)\right]}$$

We've

- derived its efficient influence function φ_{eff}, and
- constructed an estimator $\hat{\tau}_{\text{EIF}}$ that has this influence function

Efficient influence function

The efficient influence function (EIF) is

$$\varphi_{\text{eff}} = \frac{1}{\mathbb{E}\left[\{p_1(X) - p_0(X)\}\{1 - \rho(X)\}\right]} \left\{\phi^0_{1,10} - \phi^0_{0,10} - \tau^0_{10}\lambda_{10}\right\},$$

where

- $\phi^0_{1,10}$,
- $\phi^0_{1,00}$, and
- λ_{10}

are all complex functions of all 4 nuisance parameters

Efficient influence function

EIF: $\varphi_{\text{eff}} = \{\phi^0_{1,10} - \phi^0_{0,10} - \tau^0_{10}\lambda_{10}\}/\mathbb{E}\left[\{p_1(X) - p_0(X)\}\{1-\rho(X)\}\right]$,

where

$$\phi^0_{1,10} \equiv \frac{e_{10}(X)}{p_1(X)}\{1-\rho(X)\}\psi_{Y_{1,1}C_{1,1}} - \mu_{11}(X)\{1-\rho(X)\}\left\{\psi_{C_{0,1}} - \frac{p_0(X)}{p_1(X)}\psi_{C_{1,1}}\right\} + e_{10}(X)\psi_{1-R}\mu_{11}(X),$$

$$\phi^0_{0,10} \equiv \frac{e_{10}(X)}{1-p_0(X)}\{1-\rho(X)\}\psi_{Y_{0,1}(1-C_{0,1})} - \mu_{00}(X)\{1-\rho(X)\} \times \left\{\psi_{1-C_{1,1}} - \psi_{1-C_{0,1}}\left(\frac{1-p_1(X)}{1-p_0(X)}\right)\right\} + e_{10}(X)\psi_{1-R}\mu_{00}(X),$$

$$\lambda_{10} \equiv \left[\psi_{C_{1,1}} - \psi_{C_{0,1}}\right]\{1-\rho(X)\} + e_{10}(X)\psi_{1-R},$$

$$\psi_{f(Y_{a,r},C_{a,r},X)} \equiv \frac{I(A=a)I(R=r)\left[f(Y,C,X) - \mathbb{E}\{f(Y,C,X)|X,A=a,R=r\}\right]}{P(A=a|R=1,X)P(R=r|X)} + \mathbb{E}\{f(Y,C,X)|X,A=a,R=r\}, \text{ and}$$

$$\psi_{f(R)} = f(R) - \mathbb{E}[f(R)|X],$$

Efficient influence function-based estimation

The mean of φ_{eff} is zero; applying this fact, we can solve for τ_{10}^0 to obtain:

$$\tau_{10}^0 = \frac{\mathbb{E}[\phi_{1,10}^0 - \phi_{0,10}^0]}{\mathbb{E}\left[\lambda_{10}\right]}.$$

We can then estimate the effect among compliers in the target population as

$$\hat{\tau}_{\text{EIF}} = \frac{\mathbb{E}_n[\hat{\phi}_{1,10}^0 - \hat{\phi}_{0,10}^0]}{\mathbb{E}_n\left[\hat{\lambda}_{10}\right]},$$

where $\hat{\phi}_{1,10}^0$, $\hat{\phi}_{0,10}^0$, and $\hat{\lambda}_{10}$ are plug-in estimators of their respective population quantities.

Efficient Estimation

$\hat{\tau}_{\text{EIF}}$ is a complex function of

- $\hat{p}_a(X) = \widehat{P}(C = 1|A = a, R = 1, X)$
- $\hat{\rho}(X) = \widehat{P}(R = 1|X)$
- $\hat{\mu}_{ac}(X) = \widehat{\mathbb{E}}[Y|A = a, C = c, R = 1, X]$
- $\hat{\pi}(X) = \widehat{P}(A = 1|R = 1, X)$

Not all of the above need to be correctly specified for consistency of $\hat{\tau}_{\text{eff}}$.

"Weak robustness" properties of $\hat{\theta}_{\text{eff}}$

We showed that $\hat{\tau}_{\text{EIF}}$ has the following "robustness" property:

Theorem: Our estimator $\hat{\tau}_{\text{EIF}}$ is **consistent** for

$$\tau^0_{10} = \frac{\mathbb{E}\left[\{p_1(X) - p_0(X)\}(1-R)\{\mu_{11}(X) - \mu_{00}(X)\}\right]}{\mathbb{E}\left[\{p_1(X) - p_0(X)\}(1-R)\right]}$$

if *any* of the following 3 holds:

1. the models for π, ρ, and p_a are consistent, **or**
2. the models for π, ρ, and μ_{ac} are consistent, **or**
3. the models for μ_{ac} and p_a are consistent

"Strong" robustness

- Suppose nuisance functions ($\hat{\pi}$, $\hat{p}_a$, $\hat{\mu}_{ac}$, $\hat{\rho}$) are estimated from a separate, independent sample (more later)
- Suppose the following hold: $||\hat{\phi}^*_{1,10} - \phi^0_{1,10}|| = o_p(1)$, $||\hat{\phi}^*_{0,10} - \phi^0_{0,10}|| = o_p(1)$, and $||\hat{\lambda}_{10} - \lambda_{10}|| = o_p(1)$ where $||\cdot|| = \{\mathbb{E}|\cdot|^2\}^{1/2}$

Then

$$\hat{\tau}_{\text{EIF}} - \tau^0_{10} = \mathbb{E}_n\big(\varphi_{\text{eff}}\big) + \overbrace{O_p\left(R^1_n + R^2_n + R^3_n\right)}^{\text{need this to go away}} + o_p(1/\sqrt{n}),$$

where

$$R^1_n = ||\mu_{ac} - \hat{\mu}_{ac}|| \left\{||p_a - \hat{p}_a|| + ||p_{1-a} - \hat{p}_{1-a}||\right\},$$

$$R^2_n = ||\mu_{ac} - \hat{\mu}_{ac}|| \left\{||\pi - \hat{\pi}|| + ||\rho - \hat{\rho}||\right\}, \text{ and}$$

$$R^3_n = \left\{||p_a - \hat{p}_a|| + ||p_{1-a} - \hat{p}_{1-a}||\right\} \left\{||\pi - \hat{\pi}|| + ||\rho - \hat{\rho}||\right\}$$

Asymptotic normality

- If $R_n^1 = o(1/\sqrt{n})$, $R_n^2 = o(1/\sqrt{n})$, and $R_n^3 = o(1/\sqrt{n})$, then

$$\sqrt{n}\left(\hat{\tau}_{\text{EIF}} - \tau_{10}^0\right) \xrightarrow{d} N(0, \text{Var}\left\{\varphi_{\text{eff}}\right\})$$

- **Takeaway:** root-n consistency, asymptotic normality, and efficiency even if all nuisance params estimated at *slower* than parametric rates

$$R_n^1 = ||\mu_{ac} - \hat{\mu}_{ac}|| \left\{||p_a - \hat{p}_a|| + ||p_{1-a} - \hat{p}_{1-a}||\right\},$$

$$R_n^2 = ||\mu_{ac} - \hat{\mu}_{ac}|| \left\{||\pi - \hat{\pi}|| + ||\rho - \hat{\rho}||\right\}, \text{ and}$$

$$R_n^3 = \left\{||p_a - \hat{p}_a|| + ||p_{1-a} - \hat{p}_{1-a}||\right\} \left\{||\pi - \hat{\pi}|| + ||\rho - \hat{\rho}||\right\}$$

Cross-fitting/sample splitting

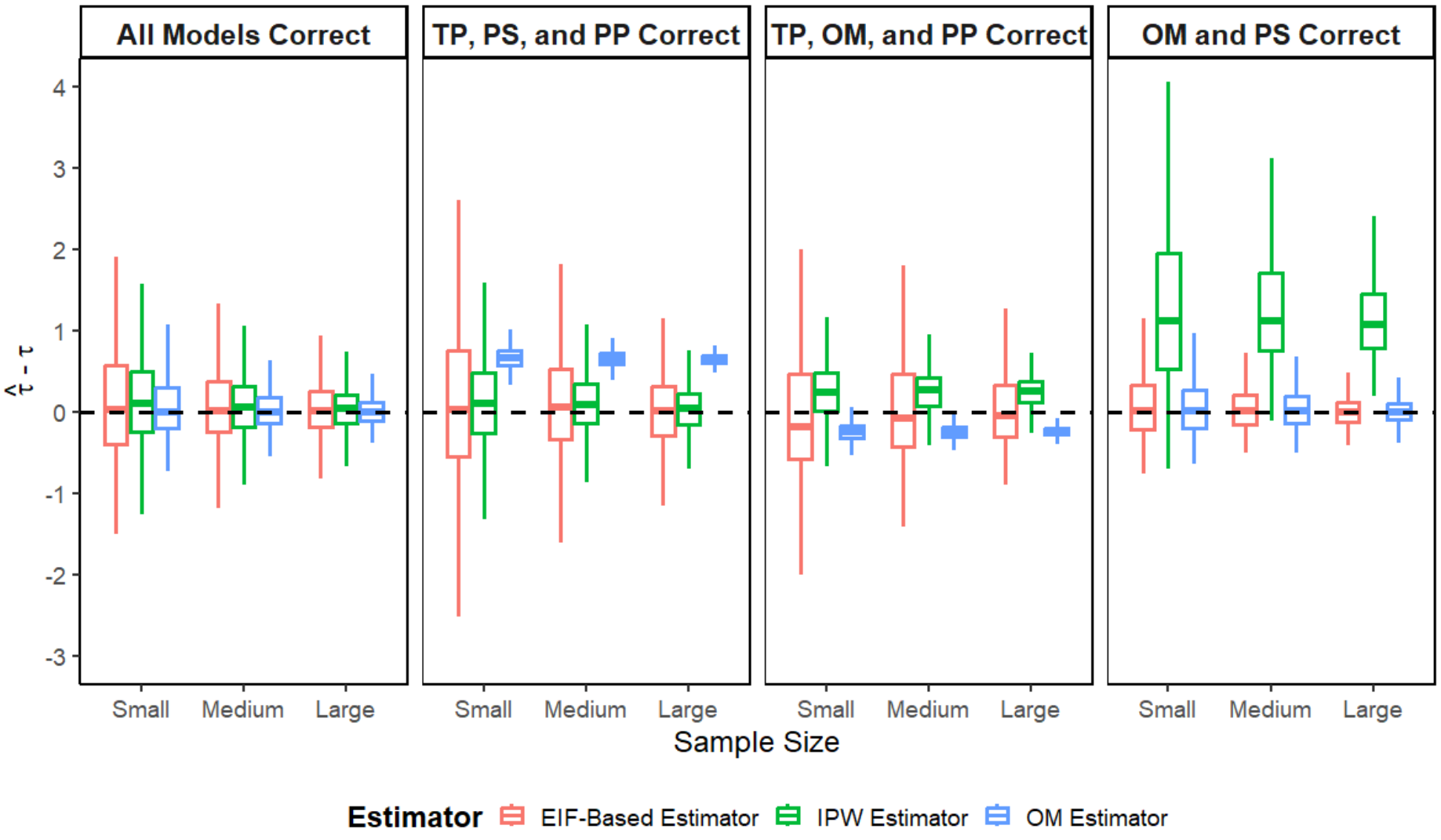
- **Reality:** we never have an independent sample to estimate nuisance parameters
  - Not actually a problem!
- **Instead, just**
  1. randomly partition data into K pieces
  2. leave one fold out, use the remaining $K - 1$ folds to estimate nuisance params
  3. predict the held out fold's nuisance param values
  4. loop through all folds
- **Now:** all samples have nuisance params estimated on held out data
- **Result:** all the previous asymptotic results hold with trivial modification

Simulation setup

We vary:

- sample size
- ratio of sample size in trial to target populations
- whether each nuisance parameter model is misspecified

Simulation results



Transporting hotspotting effects

- Data from **Camden hotspotting RCT** (613 participants)
- **Target population** data from *EHR from a large, Midwestern, academic health system*
  - same inclusion/exclusion criteria applied (463 patients)
- **Outcomes**: indicators of 30, 90, and 180 readmission

Hotspotting descriptives

Characteristic	**Trial Sample** (N=613)	**Target Pop'n Sample** (N=463)
# Hospitalizations in prior 6 mos	1.80 (1.67)	1.86 (1.44)
Length of index admission (days)	7.01 (5.99)	7.54 (9.66)
Male	315 (51.4%)	215 (46.4%)
Race		
Non-White	496 (80.9%)	178 (38.4%)
White	117 (19.1%)	285 (61.6%)
Marital status		
Married, civil union, or cohabitating	170 (27.7%)	115 (24.8%)
Single, divorced, or widowed	443 (72.3%)	348 (75.2%)
Diagnoses at/prior to index		
Heart Failure	272 (44.4%)	180 (38.9%)
HIV/AIDS	15 (2.4%)	12 (2.6%)
COPD/Emphysema	172 (28.1%)	154 (33.3%)
Diabetes	306 (49.9%)	246 (53.1%)
Arthritis	81 (13.2%)	70 (15.1%)
Mental Health Condition	501 (81.7%)	416 (89.8%)
Substance Abuse	322 (52.5%)	271 (58.5%)

An estimand for hotspotting complier effects

Transported principal causal effect among those who would engage if enrolled in hotspotting:

$$\tau_{10}^{0} = \mathbb{E}[Y(1) - Y(0) \mid \underbrace{C(1) = 1}_{\text{among engagers if enrolled}}, \underbrace{R = 0}_{\text{in the target population}}]$$

- The flexibility of the principal stratification setup allows us to define a variety of clinically-relevant estimands
- All previous estimators/efficiency theory directly applies

Relaxed assumptions for hotspotting data

- For the hotspotting data, those assigned "control" cannot access hotspotting $\implies$ one-sided compliance:

	$C = 0$	$C = 1$
$A = 0$	$C(0) = 0, C(1) \in (1, 0)$	N/A
$A = 1$	$C(1) = 0, C(0) = 0$	$C(1) = 1, C(0) = 0$

Relationship between potential and observed compliance under one-sided compliance.

- With one-sided compliance, causal assumptions can be relaxed

Relaxed assumptions for hotspotting data

The weaker variants of our original Assumptions 3-5 are:

3*. Strong monotonicity:

$$C(0) = 0$$

4*. Principal ignorability:

$$\mathbb{E}[Y(0)|U = 00, R = 1, X] = \mathbb{E}[Y(0)|U = 10, R = 1, X]$$

5*. Stratum exchangeability:

$$R \perp\!\!\!\perp C(1) \mid X.$$

Transporting hotspotting effects

Population	180-Day TE	90-Day TE	30-Day TE
Engagers in Target	8.73 (-3.45, 20.91)	7.68 (-4.92, 20.29)	-10.50 (-21.26, 0.25)
Engagers in Trial	6.94 (0.48, 13.40)	4.93 (-1.78, 11.64)	-4.73 (-10.60, 1.13)
Distilled Sample of Trial Pop'n	6.73 (-0.73, 14.18)	4.98 (-2.72, 12.69)	0.31 (-6.56, 7.18)

- Estimates of the treatment effect (TE) of hotspotting on readmission and corresponding 95% confidence intervals.
- Estimates are risk differences with positive values indicating an increase in the probability of readmission.

Transporting hotspotting effects: problems

- Sample sizes are relatively small
- We don't have access to all important variables in the target population (e.g. ethnicity, arrest history)
- Can we develop methods that use all variables in the RCT data even if not in target population?
  - Under what assumptions do these new possible methods work?
  - Are results with these new methods robust to violations of these new assumptions?

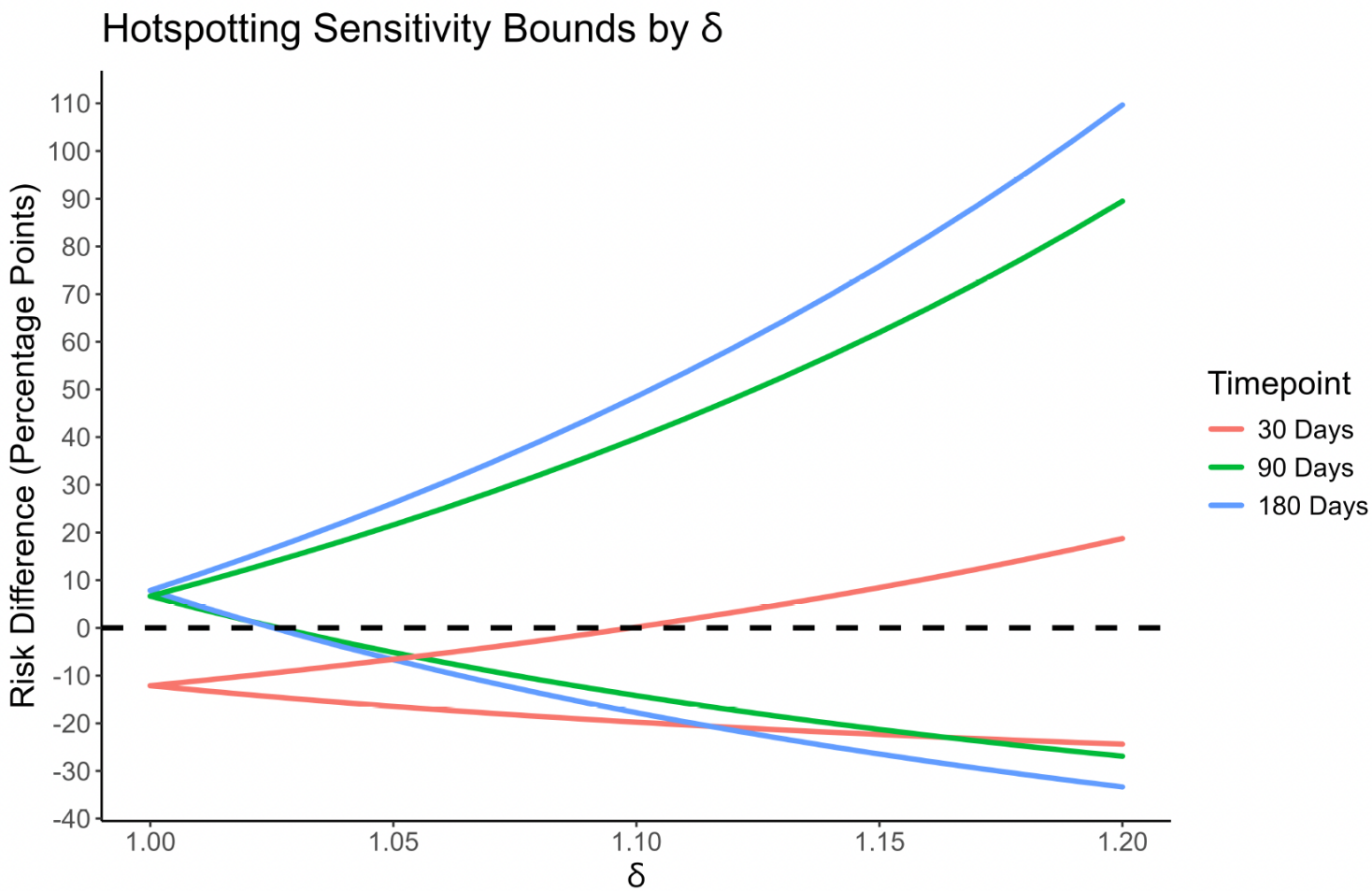
Transporting effects using all RCT data

Population	180-Day TE	90-Day TE	30-Day TE
$\hat{\tau}_{OM}$	7.41 (-35.49, 50.31)	7.24 (-36.84, 51.31)	-11.67 (-55.99, 32.65)
$\hat{\tau}_{OM,new}$	7.83 (2.22, 13.45)	6.65 (2.22, 11.08)	-12.11 (-19.14, -5.07)

- $\hat{\tau}_{OM}$ is the outcome model estimator described before limited to variables jointly available in both datasets
- $\tau_{OM,new}$ is an outcome-model style estimator that makes use of variables available in RCT but not target population

Sensitivity analysis

Sensitivity of results due to violations of new assumptions needed for $\tau_{OM,new}$



Discussion

- Standard methods transport ITT effect
- Refined estimands such as effect among compliers may help elucidate treatment effects in target populations for complex scenarios
- We introduced a basic set of assumptions for transporting effects in principal strata
- Derived efficient influence function and explored robustness to nuisance parameter misspecification
- Lots more to be done:
  - Can assumptions be relaxed?
  - Sensitivity analysis?
  - Finite sample behavior of plug-in estimators

Thanks!

jaredhuling.org

References I

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455.

Ding, P. and Lu, J. (2017). Principal stratification analysis using principal scores. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(3):757–777.

Follmann, D. A. (2000). On the effect of treatment among would-be treatment compliers: An analysis of the multiple risk factor intervention trial. *Journal of the American Statistical Association*, 95(452):1101–1109.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58(1):21–29.

Hirano, K., Imbens, G. W., Rubin, D. B., and Zhou, X.-H. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics*, 1(1):69–88.

References II

Huling, J. D., Yu, M., and Smith, M. (2019). Fused comparative intervention scoring for heterogeneity of longitudinal intervention effects. *Annals of Applied Statistics*, 13(2):824.

Jo, B. and Stuart, E. A. (2009). On the use of propensity scores in principal causal effect estimation. *Statistics in Medicine*, 28(23):2857–2875.

Rudolph, K. E. and Laan, M. J. (2017). Robust estimation of encouragement design intervention effects transported across sites. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(5):1509–1525.