Characterizing Retrospective Constraints of Survey Designs for Estimating Population-level Causal Effects

Sean Tomlin   Rebecca Andridge   Bo Lu

August 5, 2026

SRMS/SSS/GSS Student Paper Award
2026 Joint Statistical Meetings
Boston, MA

Evidence from General Use Population Surveys

- Large general use population surveys provide a unique data source for extracting causal evidence.
  - e.g., National Youth Tobacco Survey (NYTS)
- They are designed for *external validity*: population-level description and inference; unlike an internally valid RCT.
- Prior work on causal inference in complex survey design
  - Nattino et al. [2024] (henceforth NAL): exposure-sample selection from a finite population of potential outcomes.
  - NAL found the *available* sampling weight impacts analysis choice (e.g., use of survey weighted propensity score).

Retrospective Use for Causal Questions

- General use surveys release public-use files to secondary users.
- Survey managers construct design weights.
- Some undisclosed info may be crucial for constructing survey weights or mitigating bias during the original survey design.

Problem: Causal analyses can face constraints imposed by retrospective use of a survey design.

NYTS 2021: Prior E-cig Use and Future Intention to Smoke

What is the causal effect of **past e-cig use** on **intention to smoke** among never-smoking grade 6-12 students in the US?

Population ($\mathcal{P}$)	US students in grades 6–12 who have never tried cigarette smoking (cohort)
Exposure (Z)	Prior e-cigarette use
Outcome (Y)	Future intention to smoke
Confounders (X)	Age, sex, race, environmental, household, cognitive, exposure factors
Survey design ($S^* = S \cdot R = 1$)	design weights (nonresponse adjusted), sampling units, stratum, location

NYTS 2021: nationally representative, three-stage stratified cluster sample; 279 schools, $>$20,000 students.

NYTS 2021 Methodology Considerations

How might NYTS 2021 methodology impact causal effect estimation?

1. Location: "Where are you currently taking the survey?"
   - "On School Campus" reported higher e-cig prevalence than "At Home/Other Place." [Park-Lee et al., 2023]
2. Nonresponse adjustment in provided survey weights
   - Adjusts sex, grade, school level, census region, % of black students enrolled in school; *not* prior e-cig use.
   - The survey weights cannot be considered independent from prior e-cig use, due to correlation with NR-adjustment set.
3. Variance estimation*
   - Released design weights are fixed (e.g., no replicate weights provided).

Contributions

1. Formalize causal assumptions under retrospective use of complex survey designs.
2. Propose a post hoc weight class adjustment for certain practical scenarios.
3. Demonstration using NYTS 2021 data

Estimand and Causal Assumptions

Population Average Treatment Effect (PATE):

$$\text{PATE} := \frac{1}{N}\sum_{i=1}^{N}(Y_i^1 - Y_i^0) = E_{\mathcal{P}}\left(Y^1 - Y^0\right)$$

(A1) Ignorability.

$$(Z, S^*) \perp\!\!\!\perp (Y^1, Y^0) \mid H$$

(A2) Positivity.

$$0 < P(Z{=}z, S^*{=}1 \mid H) < 1$$

(A3) Consistency.

$$Y = Y^1 Z + Y^0(1 - Z)$$

Under (A1)–(A3), for $z = 0, 1$:

$$E(Y^z) = E\left\{\frac{Y\,S^*Z}{P(S^*{=}1, Z{=}z \mid H)}\right\}$$

Identification via inverse joint selection–exposure weighting.

A Few Structural Scenarios

The joint selection weight $w_i^{-1} = P(S^*{=}1, Z{=}z \mid H)$ factors two ways:

$$\underbrace{P(Z{=}z \mid S^*{=}1, H)\, P(S^*{=}1 \mid H)}_{\textbf{Path 1}} = \underbrace{P(S^*{=}1 \mid Z{=}z, H)\, P(Z{=}z \mid H)}_{\textbf{Path 2}}$$

- **Path 1** Propensity score in-sample $\times$ Sample selection probability.
- **Path 2** Sample selection adjusted for Z $\times$ Propensity score.

Partition the covariates into disclosed and undisclosed.

X	disclosed in the public-use file
W	undisclosed (used by managers to build weights)

Which path can we take to identify the PATE? We consider some structural scenarios about (Z, S^*, X, W).

Does the Released $P(S^*{=}1 \mid W)$ Identify the PATE?

Factor $P(Z{=}1, S^*{=}1 \mid X, W)$ two ways:

$$\textbf{Path 1:} \quad P(Z{=}1 \mid X, W, S^*{=}1)\, P(S^*{=}1 \mid X, W)$$

$$\textbf{Path 2:} \quad P(S^*{=}1 \mid X, W, Z{=}1)\, P(Z{=}1 \mid X, W)$$

X → **exposure,** W → **selection**: $Z \perp\!\!\!\perp (S^*, W) \mid X$ and $S^* \perp\!\!\!\perp (Z, X) \mid W$

$$\text{Path 1} = \text{Path 2} = P(Z{=}1 \mid X)\, P(S^*{=}1 \mid W)$$

Exposure-dependent selection: $Z \perp\!\!\!\perp W | X$ and $S^* \not\perp\!\!\!\perp Z | X, W$

Path 1: $P(Z{=}1 \mid X, W, S^*{=}1)$ does not reduce

Path 2: $P(S^*{=}1 \mid X, W, Z{=}1)\, P(Z{=}1 \mid X)$

The released $P(S^*{=}1 \mid W)$ does not recover $P(S^*{=}1 \mid X, W, Z{=}1)$

Weight Class Adjustment (WCA)

The released weight corrects for nonresponse, but was built without exposure Z, which may affect who responds.

Assumption: undisclosed design variables W affect *selection*, not exposure. Then, we can adjust the released weight as follows.

WCA on Quantiles of Design Weight

1. Group observations into K classes by their released weight.
2. Within each class, take the proportion reporting exposure:

$$\hat{P}(Z{=}z \mid \text{class}_i) = \frac{1}{|\text{class}_i|} \sum_{j \in \text{class}_i} \mathbb{I}(Z_j = z)$$

3. Scale each weight by this proportion.

The adjusted weight now reflects the exposure the original weight ignored.

WCA Estimator

The released weight is $\hat{\pi}(X_i, W_i) = 1/a_i$. The WCA weight combines confounding and selection corrections:

$$\hat{w}_i^{\text{WCA}} = \hat{w}_{ps_i} \cdot \frac{1}{\hat{\pi}(X_i, W_i)\, \hat{P}(Z_i{=}z \mid \text{class}_i)}, \qquad \hat{w}_{ps_i} = \frac{Z_i}{\hat{e}(X_i)} + \frac{1 - Z_i}{1 - \hat{e}(X_i)}$$

PATE estimator:

$$\hat{\tau}_{\text{WCA}} = \frac{\sum_i \hat{w}_i^{\text{WCA}} Y_i Z_i}{\sum_i \hat{w}_i^{\text{WCA}} Z_i} - \frac{\sum_i \hat{w}_i^{\text{WCA}} Y_i (1 - Z_i)}{\sum_i \hat{w}_i^{\text{WCA}} (1 - Z_i)}$$

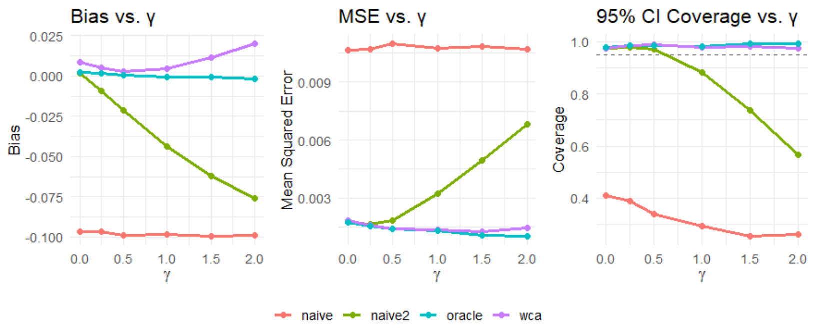
Consistency (fixed K)

$$\hat{\tau}_{\text{WCA}} \xrightarrow{p} \tau + (\text{bias from coarse } K)$$

$\sqrt{n}$-consistent to τ plus coarsening bias {which vanishes as K grows (or as within class distribution of (X,W) becomes more homogeneous}.

Simulation Study

Design. Selection depends on X and, through γ, on exposure ($\gamma > 0$: Z-dependent). Compare four estimators as γ grows: two naïve, WCA, and Oracle.



Result: WCA tracks the Oracle across all γ; the naïve estimators degrade as exposure-dependent selection grows.

Weight Class Adjustment in the NYTS 2021

- Conditioning on $S^*\!=\!1$ links survey setting and prior e-cig use (School Campus **16.0%** vs. Home/Other **12.3%**).
- We adjust `finwgt` on Location cells, as in the NYTS nonresponse adjustment [CDC's Office on Smoking and Health, 2021].

Location	Z	n	$\hat{c}$
School Campus	0	6,845	1.118
School Campus	1	1,074	0.924
Home/Other	0	5,609	0.906
Home/Other	1	814	0.940

$\hat{c}$ = sample share $\div$ weighted share.

Small adjustment, reflecting a well-designed survey.

WCA $\rightarrow$ adjusted weight $\rightarrow$ $\hat{e}(X)$ $\rightarrow$ balance $\rightarrow$ $\widehat{\text{PATE}}$

Population-level Covariate Balance

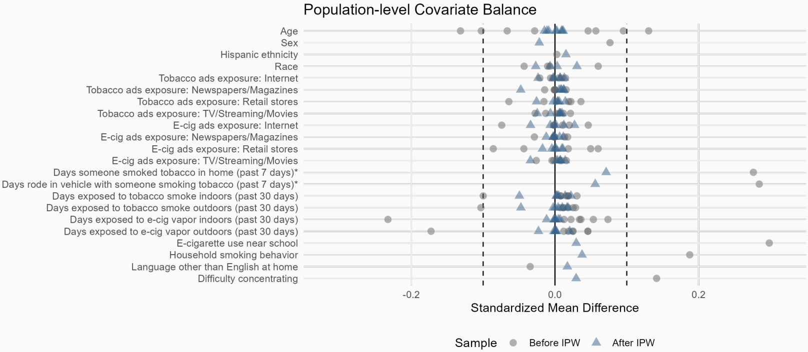


Figure 1: Population-level covariate balance before-after propensity score adjustment with WCA design weights.

Estimated PATE with Post Hoc WCA on Design Weights

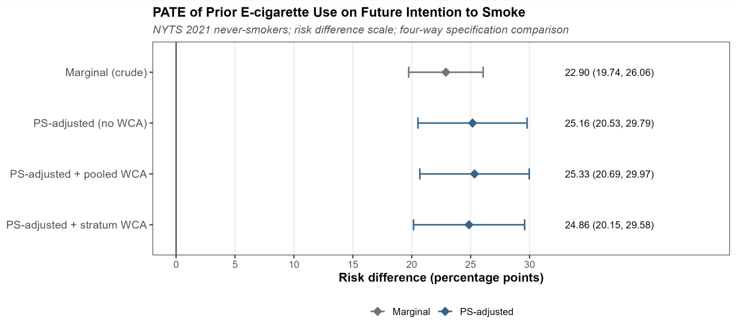


Figure 2: Estimated PATEs (risk difference in %) of prior e-cig use on future intention to smoke among never-smokers, NYTS 2021.

Result: $\widehat{\text{PATE}} = 25.33\%$ [95% CI: 20.69, 29.97]. Prior e-cig use increases the the risk of future intention to smoke among never-smoking US grade 6-12 students.

Variance Estimation as a Retrospective Constraint

Released weights are treated as fixed, but were estimated by survey managers. Analysts could propagate that uncertainty using survey bootstrap methods [Preston, 2009].

	SE (pp)	**95% CI (pp)**
Linearization (`svyglm`)	2.33	(20.7, 30.0)
Survey bootstrap	2.74	(20.0, 30.7)

$$\mathrm{Var}(\hat{\tau} \mid \text{weights fixed}) \;\leq\; \mathrm{Var}(\hat{\tau})$$

Survey bootstrapped SE ≈18% larger than treating survey weights as fixed.

Discussion

- We formalized some constraints to identification when estimating causal effects from complex survey design.
- Post-hoc WCA may be used in some scenarios
- Demonstration using NYTS 2021
- Ongoing: Sensitivity analysis with complex survey design, survey bootstrap methods, other constraints

References

CDC's Office on Smoking and Health. 2021 National Youth Tobacco Survey: Methodology Report. Technical report, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Atlanta, GA, 2021.

G. Nattino, R. Ashmead, and B. Lu. Causal inference with complex surveys: A unified perspective on sample selection and exposure selection. *The American Statistician*, pages 1–18, 2024.

E. Park-Lee, A. S. Gentzke, C. Ren, M. Cooper, M. D. Sawdey, S. S. Hu, and K. A. Cullen. Impact of survey setting on current tobacco product use: National youth tobacco survey, 2021. *Journal of Adolescent Health*, 72(3):365–374, 2023.

J. Preston. Rescaled bootstrap for stratified multistage sampling. *Survey Methodology*, 35(2):227–234, 2009.