# Beyond Exchangeability: Distribution-Shift-Aware Integration of External Control Data in Randomized Trials

Jiawei Shan

Department of Statistics
Department of Biostatistics & Medical Informatics
University of Wisconsin-Madison

IMSI Workshop
June 26, 2026

# Acknowledgement

Yiteng Tu (UCSD)

Guanbo Wang (Dartmouth)

Chao Ying (UW-Madison)

Jiwei Zhao (UW-Madison)

# Randomized trials: the gold standard, with caveats

- Randomized controlled trials (RCTs) are the gold standard for causal effects: randomization removes confounding *by design*.
- In practice, however, trial estimates are often not precise enough:
  - optimistic assumptions about event rates, follow-up, and accrual;
  - costly and slow to scale.
- This motivates augmenting an RCT with auxiliary external data, e.g., historical trials, registries, electronic health records, claims. (Pocock, 1976; Colnet et al., 2024)

# External controls (EC)

- For studies of an *experimental* intervention, the auxiliary data often contain only untreated subjects.
- These are used as external controls (EC).
- Central question:

*How can we use EC data to improve efficiency without introducing bias?*

| **RCT** (treated + control) |
| --- |
| randomized $T \in \{0, 1\}$<br>small / underpowered |

+

| **External Controls** |
| --- |
| untreated only ($T = 0$)<br>large, but different |

# The exchangeability assumption

- The standard route assumes exchangeability: conditional on covariates, untreated outcomes are comparable across trial and EC,

$$p(y_0 \mid \mathbf{x}) = q(y_0 \mid \mathbf{x}) \quad \Big(\text{or the weaker mean version } \mathbb{E}_p(Y_0 \mid \mathbf{x}) = \mathbb{E}_q(Y_0 \mid \mathbf{x})\Big).$$

- Then EC subjects directly inform untreated outcomes in the trial.
- A large literature borrows EC information under this assumption. (Dahabreh et al., 2019; Li et al., 2023; Valancius et al., 2024)

*$p(\cdot)$ and $q(\cdot)$ denote the distributions in the RCT and EC populations.*

# But exchangeability often fails

- Differences in eligibility criteria, standard of care, calendar time, recruitment, and measurement induce distribution shift.
- EC subjects are then *systematically different* from trial controls.
- Existing remedies are guarded / defensive:
  - select compatible controls; downweight, discount, or shrink external information. (Rosenman et al., 2023; Yang et al., 2023; Gao et al., 2024)
- Limitations:
  - they improve robustness but may discard or attenuate useful information;
  - they typically improve only for the control arm.

# This work: be *distribution-shift-aware*

*Don't assume the shift away; instead, model it explicitly and exploit it.*

1. A framework to integrate EC data without exchangeability, by exploiting distribution shifts.
2. Calibration-based augmented estimating functions from trial-only EIFs, using *both* covariate and concept shift $\Rightarrow$ improve both arms.
3. Practical augmented estimators by directly modeling / estimating the shifts.
4. A data-adaptive shrinkage estimator: consistent under shift-model misspecification, with guaranteed efficiency dominance over trial-only inference.

# Data structure and notation

- Source indicator $R$: $R = 1$ for RCT, $R = 0$ for EC.
- Covariate $\mathbf{X}$, outcome $Y$, treatment $T \in \{0, 1\}$; potential outcomes $Y = TY_1 + (1 - T)Y_0$.
- Observed data $\mathcal{O}$: $N$ i.i.d. copies of $(R, T, \mathbf{X}, Y)$, with $n/N \to \kappa \in (0, 1)$.

| | $R$ | $T$ | $Y_1$ | $Y_0$ | $\mathbf{X}$ |
|---|---|---|---|---|---|
| RCT $\mathcal{O}_1 = \{(r_i = 1, t_i, \mathbf{x}_i, y_i)\}_{i=1}^{n}$ | 1 | 1 | ✓ | | ✓ |
| | 1 | 0 | | ✓ | ✓ |
| EC $\mathcal{O}_2 = \{(r_i = 0, t_i = 0, \mathbf{x}_i, y_i)\}_{i=n+1}^{N}$ | 0 | 0 | | ✓ | ✓ |

# Estimand and the RCT design

- Target: the average treatment effect (ATE) in the trial population,

$$\tau = \tau_1 - \tau_0, \qquad \tau_t = \mathbb{E}(Y_t \mid R = 1), \quad t \in \{0, 1\}.$$

- **Assumption 1** (randomization in the trial):

$$\text{(i)} \;\; T \perp\!\!\!\perp (\mathbf{X}, Y_0, Y_1) \mid R = 1; \qquad \text{(ii)} \;\; 0 < \pi = \mathsf{pr}(T = 1 \mid R = 1) < 1.$$

- Simple gold-standard trial estimators: treated / control sample means

$$\widetilde{\tau}_1^{\text{gs}} = \frac{\sum_{i=1}^n t_i y_i}{\sum_{i=1}^n t_i}, \qquad \widetilde{\tau}_0^{\text{gs}} = \frac{\sum_{i=1}^n (1 - t_i) y_i}{\sum_{i=1}^n (1 - t_i)}.$$

# Trial-only efficient influence functions (the benchmark)

- Let $\mu_t(\mathbf{x}) = \mathbb{E}(Y \mid \mathbf{x}, t, R = 1)$. The trial-only EIFs for $\tau_1$ and $\tau_0$ (Robins et al., 1994)

$$\widetilde{\psi}_1 = \frac{RT}{\kappa\pi}\{Y - \mu_1(\mathbf{X})\} + \frac{R}{\kappa}\{\mu_1(\mathbf{X}) - \tau_1\},$$
$$\widetilde{\psi}_0 = \frac{R(1-T)}{\kappa(1-\pi)}\{Y - \mu_0(\mathbf{X})\} + \frac{R}{\kappa}\{\mu_0(\mathbf{X}) - \tau_0\}.$$

- Set $\widetilde{\psi} = \widetilde{\psi}_1 - \widetilde{\psi}_0$. The trial-only semiparametric efficiency bound of $\tau$ is

$$\widetilde{B} = \mathbb{E}(\widetilde{\psi}^2).$$

- This is the best we can do with the trial alone, and thus serves as our benchmark.

# Beyond exchangeability: model the shift

- Capture the discrepancy between the two sources by two density ratios:

$$\underbrace{a(\mathbf{x}) = \frac{q(\mathbf{x})}{p(\mathbf{x})}}_{\textbf{covariate shift}} \quad \text{and} \quad \underbrace{b(\mathbf{x}, y_0) = \frac{q(y_0 \mid \mathbf{x})}{p(y_0 \mid \mathbf{x})}}_{\textbf{concept shift}}.$$

- Exchangeability is the special case $b(\mathbf{x}, y_0) \equiv 1$.
- Define the induced source-probability weights:

$$k(\mathbf{x}) = \mathsf{pr}(R = 1 \mid \mathbf{x}) = \frac{\kappa}{\kappa + (1-\kappa)a(\mathbf{x})},$$

$$\rho(\mathbf{x}, y_0) = \mathsf{pr}(R = 1 \mid \mathbf{x}, y_0, T = 0) = \frac{\kappa(1-\pi)}{\kappa(1-\pi) + (1-\kappa)a(\mathbf{x})b(\mathbf{x}, y_0)}.$$

# Distribution balancing

*Lemma 1 (Balancing).* *Under Assumption 1, for any integrable* $g_1(\mathbf{x})$ *and* $g_2(\mathbf{x}, y_0)$,

$$\mathbb{E}\{g_1(\mathbf{X}) \mid R = 1\} = \mathbb{E}\Big\{ \tfrac{k(\mathbf{X})}{\kappa}\, g_1(\mathbf{X}) \Big\},$$
$$\mathbb{E}\{g_2(\mathbf{X}, Y_0) \mid R = 1\} = \mathbb{E}\Big\{ \tfrac{1-T}{\kappa(1-\pi)}\, \rho(\mathbf{X}, Y_0)\, g_2(\mathbf{X}, Y_0) \Big\}.$$

- Reweighting by $k$ and $\rho$ transports expectations from the *trial* data to the *pooled* population.
- This is the engine for augmenting trial-only estimating functions with EC data.

# EC-augmented estimating function: treated arm $\tau_1$

- For $\tau_1$, the EC sample contributes *only* covariate information (through $\mathbf{X}$).

$$\text{trial-only:} \quad \widetilde{\psi}_1 = \frac{RT}{\kappa\pi}\{Y - \mu_1(\mathbf{X})\} + \frac{R}{\kappa}\{\mu_1(\mathbf{X}) - \tau_1\},$$

$$\text{augmented:} \quad \psi_1 = \underbrace{\frac{RT}{\kappa\pi}\{Y - \mu_1(\mathbf{X})\}}_{\text{unchanged}} + \underbrace{\frac{k(\mathbf{X})}{\kappa}\{\mu_1(\mathbf{X}) - \tau_1\}}_{\text{covariate contribution}}.$$

- Single change: $\dfrac{R}{\kappa} \rightsquigarrow \dfrac{k(\mathbf{X})}{\kappa}$ — the covariate-shift weight that anchors the regression term to the *pooled* population.

# EC-augmented estimating function: control arm $\tau_0$

- For $\tau_0$, the EC sample contributes untreated-outcome *and* covariate information.

$$\text{trial-only:} \quad \widetilde{\psi}_0 = \frac{R(1-T)}{\kappa(1-\pi)}\{Y - \mu_0(\mathbf{X})\} + \frac{R}{\kappa}\{\mu_0(\mathbf{X}) - \tau_0\},$$

$$\text{augmented:} \quad \psi_0 = \underbrace{\frac{(1-T)\,\rho(\mathbf{X}, Y)}{\kappa(1-\pi)}\{Y - \mu_0(\mathbf{X})\}}_{\text{outcome contribution}} + \underbrace{\frac{k(\mathbf{X})}{\kappa}\{\mu_0(\mathbf{X}) - \tau_0\}}_{\text{covariate contribution}}.$$

- $R(1-T) \rightsquigarrow (1-T)\rho(\mathbf{X}, Y)$: concept-shift weight — borrows EC *untreated outcomes* over *all* untreated units.
- $R \rightsquigarrow k(\mathbf{X})$: covariate-shift weight (*same* correction as $\tau_1$) — anchors the regression term to the pooled population.

# Validity and guaranteed efficiency gain

*Proposition 1 (Efficiency gain).* *We have* $\mathbb{E}(\psi_t) = 0$ *for* $t \in \{0, 1\}$, *and*

$$\begin{aligned}
\mathbb{E}(\psi_1^2) - \mathbb{E}(\widetilde{\psi}_1^2) &= -\tfrac{1}{\kappa^2}\mathbb{E}\big[\{\mu_1(\mathbf{X}) - \tau_1\}^2\, k(\mathbf{X})\{1 - k(\mathbf{X})\}\big] \leq 0, \\
\mathbb{E}(\psi_0^2) - \mathbb{E}(\widetilde{\psi}_0^2) &= -\tfrac{1}{\kappa^2}\mathbb{E}\big[\{\mu_0(\mathbf{X}) - \tau_0\}^2\, k(\mathbf{X})\{1 - k(\mathbf{X})\}\big] \\
&\quad - \tfrac{1}{\kappa^2(1-\pi)^2}\mathbb{E}\big[(1 - T)\{Y - \mu_0(\mathbf{X})\}^2 \rho(\mathbf{X}, Y)\{1 - \rho(\mathbf{X}, Y)\}\big] \leq 0.
\end{aligned}$$

*Equalities hold if and only if* $\kappa = 1$ *(i.e. no EC data). In addition, let* $\psi = \psi_1 - \psi_0$, *and we have* $\mathbb{E}(\psi^2) - \mathbb{E}(\widetilde{\psi}^2) \leq 0$, *and the equality holds if and only if* $\kappa = 1$.

- Variance never increases; strictly smaller whenever EC data are present.
- EC improves both arms: covariate shift helps $\tau_1$; covariate + concept shift help $\tau_0$.

# Constructing the augmented estimator

- Posit parametric working models $k(\mathbf{x};\boldsymbol{\alpha})$, $\rho(\mathbf{x},y_0;\boldsymbol{\beta})$, $\mu_t(\mathbf{x};\boldsymbol{\gamma}_t)$, with $\sqrt{N}$-consistent nuisance estimators.
- Solve the empirical moment equations from $\psi_1, \psi_0$:

$$\widehat{\tau}_1 = \sum_{i=1}^{N} \left[ \tfrac{r_i t_i}{\widehat{\pi}} \{y_i - \mu_1(\mathbf{x}_i;\widehat{\boldsymbol{\gamma}}_1)\} + k(\mathbf{x}_i;\widehat{\boldsymbol{\alpha}})\, \mu_1(\mathbf{x}_i;\widehat{\boldsymbol{\gamma}}_1) \right] \Big/ \sum_{i=1}^{N} k(\mathbf{x}_i;\widehat{\boldsymbol{\alpha}}),$$

$$\widehat{\tau}_0 = \sum_{i=1}^{N} \left[ \tfrac{1-t_i}{1-\widehat{\pi}}\, \rho(\mathbf{x}_i,y_i;\widehat{\boldsymbol{\beta}}) \{y_i - \mu_0(\mathbf{x}_i;\widehat{\boldsymbol{\gamma}}_0)\} + k(\mathbf{x}_i;\widehat{\boldsymbol{\alpha}})\, \mu_0(\mathbf{x}_i;\widehat{\boldsymbol{\gamma}}_0) \right] \Big/ \sum_{i=1}^{N} k(\mathbf{x}_i;\widehat{\boldsymbol{\alpha}}).$$

- ATE estimator: $\widehat{\tau} = \widehat{\tau}_1 - \widehat{\tau}_0$.

# Consistency

*Theorem 2 (Consistency). Under standard regularity conditions:*

- $\widehat{\tau}_1$ *is consistent and asymptotically normal (CAN) if* $k(\mathbf{x}; \boldsymbol{\alpha}^*) = k(\mathbf{x})$*;*
- $\widehat{\tau}_0$ *is CAN if* $k(\mathbf{x}; \boldsymbol{\alpha}^*) = k(\mathbf{x})$ *and* $\rho(\mathbf{x}, y_0; \boldsymbol{\beta}^*) = \rho(\mathbf{x}, y_0)$.

- Consistency hinges on correctly modeling the shifts $k, \rho$ (equivalently $a(\mathbf{x}), b(\mathbf{x}, y_0)$).
- Robust to misspecification of the outcome models $\mu_t$, a robustness flavor inherited from the trial-only EIFs.

# Asymptotic distribution

*Theorem 3 (Asymptotic representation).* *If all working models are correctly specified, then*

$$\sqrt{N}(\widehat{\tau}_1 - \tau_1) = N^{-1/2} \textstyle\sum_{i=1}^{N} \psi_{1,i} + \underbrace{\Gamma_1 \sqrt{N}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)}_{\text{from estimating the shift}} + o_p(1),$$

$$\sqrt{N}(\widehat{\tau}_0 - \tau_0) = N^{-1/2} \textstyle\sum_{i=1}^{N} \psi_{0,i} + \underbrace{\Gamma_0 \sqrt{N}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*) + \Gamma_\rho \sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)}_{\text{from estimating the shift}} + o_p(1).$$

- Estimating the shift parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ affects efficiency;
- Estimating the outcome parameters $(\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1)$ does *not*.

# Two regimes for the shift

- Known shifts (from design or reliable external knowledge): the remainder terms vanish, and

$$\sqrt{N}(\widehat{\tau}_t - \tau_t) \xrightarrow{d} \mathcal{N}\{0, \mathbb{E}(\psi_t^2)\} \quad \Rightarrow \quad \text{efficiency gain over trial-only.}$$

- Validation data (auxiliary sample of size $m$, $m/N \to \xi$):

$$\sigma^2_{t,\text{val}} = \mathbb{E}(\psi_t^2) + \underbrace{\xi^{-1} \cdot (\text{shift-estimation variance})}_{=O(\xi^{-1})}.$$

The extra term vanishes as the validation sample grows $\Rightarrow$ recover the known-shift efficiency.

# Why shrinkage?

- In practice the shifts are unknown — they must be modeled *and* estimated from data.
- Two distinct risks for the augmented estimator $\widehat{\tau}_t$:
  - misspecified $k, \rho \Rightarrow$ bias (inconsistency);
  - even when correctly specified, $\widehat{\tau}_t$ could be less efficient than trial-only due to the extra variability of estimating the shift.
- We want a single estimator that is
  - consistent regardless of model specification, and
  - at least as efficient as the trial-only benchmark, keeping the gain whenever EC is informative.

# Adaptive shrinkage

- Combine the trial-only estimator $\widetilde{\tau}_t$ and augmented estimators $\widehat{\tau}_t$ along a one-parameter family:

$$\widehat{\tau}_t(\lambda_t) = \widetilde{\tau}_t + \lambda_t(\widehat{\tau}_t - \widetilde{\tau}_t), \quad \lambda_t \in \mathbb{R}.$$

  - $\widehat{\tau}_t(0) = \widetilde{\tau}_t$, and $\widehat{\tau}_t(1) = \widehat{\tau}_t$.

1. Consistency: if $\widehat{\tau}_t$ is biased (shift model misspecified), we want $\lambda_t \to 0$, so the combination collapses to the consistent trial-only $\widetilde{\tau}_t$.
2. Efficiency: if $\widehat{\tau}_t$ is consistent, then $\widehat{\tau}_t(\lambda_t)$ is consistent for *every* $\lambda_t$, so we are free to choose $\lambda_t$ to minimize the variance.

# Adaptive shrinkage: optimal weight

- If $\widehat{\tau}_t$ is consistent, $\widehat{\tau}_t(\lambda_t)$ is consistent for every $\lambda_t$ — so pick $\lambda_t$ to minimize the variance.
- Minimizing $\mathrm{var}\{\widehat{\tau}_t(\lambda_t)\}$ over $\lambda_t$ gives the optimal weight

$$\lambda_t^* = -\frac{\mathrm{cov}(\widehat{\tau}_t - \widetilde{\tau}_t,\ \widetilde{\tau}_t)}{\mathrm{var}(\widehat{\tau}_t - \widetilde{\tau}_t)},$$

and the resulting variance dominates *both* components:

$$\sigma_t^2 := \mathrm{var}\{\widehat{\tau}_t(\lambda_t^*)\} \leq \min\{\mathrm{var}(\widetilde{\tau}_t), \mathrm{var}(\widehat{\tau}_t)\}.$$

# Adaptive shrinkage: data-adaptive weight

- Define the gap $D_t = \widehat{\tau}_t - \widetilde{\tau}_t$, and $\mathrm{var}(D_t) = O(N^{-1})$.
  - $\widehat{\tau}_t$ consistent: $D_t = O_p(N^{-1/2})$;
  - otherwise: $D_t = c_t^* + O_p(N^{-1/2}) = O_p(1)$;
- To facilitate adaptive shrinkage driven by the consistency of $\widehat{\tau}_t$, define

$$\delta_{t,N} = \frac{\mathrm{var}(D_t)}{\mathrm{var}(D_t) + D_t^4} \longrightarrow \begin{cases} 1, & \widehat{\tau}_t \text{ consistent}, \\ 0, & \text{otherwise}. \end{cases}$$

- Shrinkage weight

$$\lambda_{t,N} = \delta_{t,N}\, \lambda_t^* \longrightarrow \begin{cases} \lambda_t^*, & \widehat{\tau}_t \text{ consistent}, \\ 0, & \text{otherwise}. \end{cases}$$

- The shrinkage estimator is $\widehat{\tau}_t^{(s)} = \widehat{\tau}_t(\widehat{\lambda}_{t,N})$.

# Guarantees: consistency + efficiency dominance

*Theorem 4 (Adaptive shrinkage).* *The shrinkage estimator $\widehat{\tau}_t^{(s)}$ is CAN regardless of model specification:*

(i) *if $\widehat{\tau}_t$ is inconsistent, then $\widehat{\tau}_t^{(s)}$ behaves like the trial-only $\widetilde{\tau}_t$;*

(ii) *if $\widehat{\tau}_t$ is consistent, then $\sqrt{N}\{\widehat{\tau}_t^{(s)} - \tau_t\} \xrightarrow{d} \mathcal{N}(0, \sigma_t^2)$, with $\sigma_t^2 \leq \min\{\mathrm{var}(\widetilde{\tau}_t), \mathrm{var}(\widehat{\tau}_t)\}$.*

## Corollary

The asymptotic variance of $\widehat{\tau}_t^{(s)}$ is never larger than that of the trial-only estimator $\widetilde{\tau}_t$ — for *any* choice of working models.

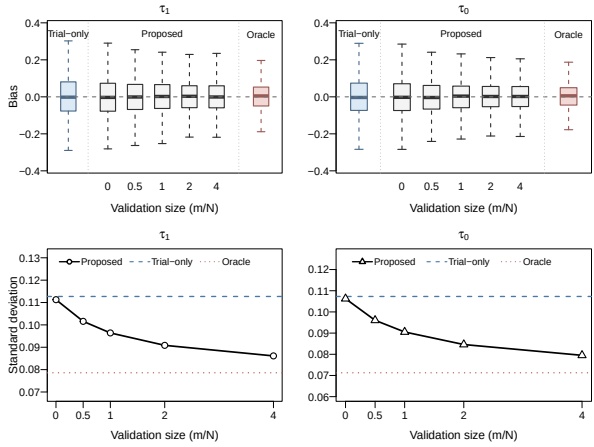
A safe estimator: efficiency dominance with guaranteed consistency.

# Simulations

- **Trial:** $X \sim \mathcal{N}(1.5, 0.8^2)$, $Y_t \mid x \sim \mathcal{N}\{\mu_t(x), 0.8^2\}$ with $\mu_1(x) = 2 + x + 0.6e^x$, $\mu_0(x) = 1 + 1.5x + 0.5e^x$.
- **External controls:** $X \sim \mathcal{N}(1, 1)$, $Y \mid x \sim \mathcal{N}\{0.5 + \mu_0(x), 0.8^2\}$.
  $\Rightarrow$ EC differs from the trial in both the $X$-marginal *and* $Y_0 \mid X$.
- Sizes $n = 1500$, $N = 3500$, $\pi = 0.5$.
- True shifts: $k(x; \boldsymbol{\alpha}) = \text{expit}(\alpha_0 + \alpha_1 x + \alpha_2 x^2)$, $\rho(x, y; \boldsymbol{\beta}) = \text{expit}(\beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 e^x + \beta_4 y)$.

# Experiment 1: efficiency gains under correct model specification

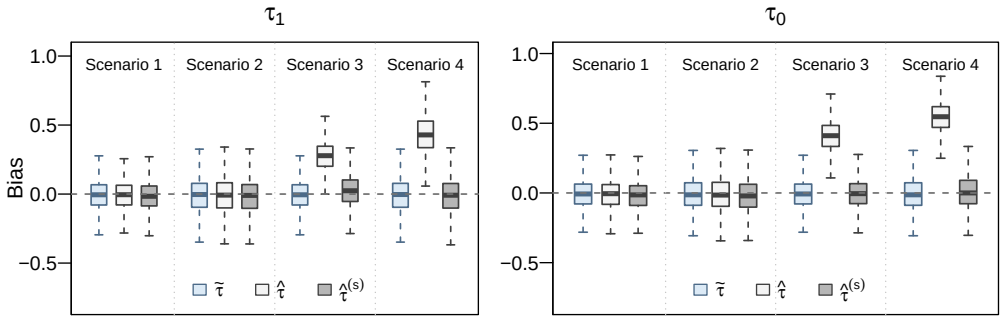
- Regimes: no info / validation data ($m/N \in \{0.5, 1, 2, 4\}$) / oracle (known shifts).




- All unbiased (top); proposed SD decreases from trial-only toward the oracle as validation grows (bottom).

# Experiment 2: robustness to misspecification

- Four scenarios: (1) all correct; (2) $\mu_t$ wrong; (3) $k, \rho$ wrong; (4) all wrong.

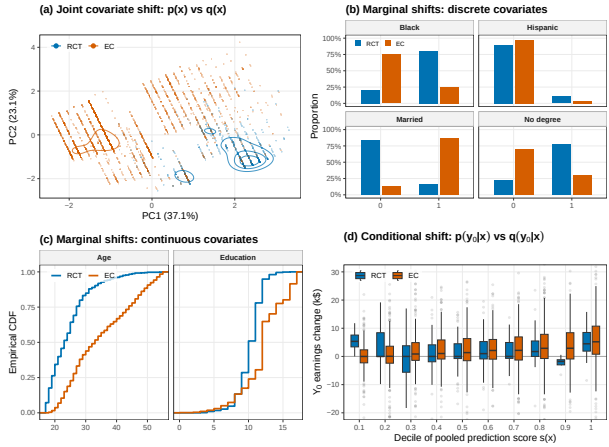



- Augmented $\widehat{\tau}$ is unbiased in (1)–(2) [robust to $\mu_t$], but biased in (3)–(4);
- the shrinkage $\widehat{\tau}^{(s)}$ stays centered at the truth in all scenarios.

# Real data: LaLonde NSW + PSID

- **Data:** NSW — a randomized job-training trial (RCT); PSID — an observational external-control group (untreated). (Lalonde, 1986; Dehejia and Wahba, 1999)
- **Outcome:** change in earnings, $\texttt{re\_dif} = \texttt{re78} - \texttt{re75}$.
- **Covariates (6):** age, education, black, hispanic, married, no-degree.
- **Design:** random $50/50$ split $\Rightarrow$ primary + validation sample; logistic working models for $k(\mathbf{x})$ and $\rho(\mathbf{x}, y_0)$.

# Distribution shift is real: NSW vs. PSID




- (a) PCA separation and (b,c) marginal covariate shift; (d) $Y_0$ differs within score deciles $\Rightarrow$ concept shift.

# Real data: results

| Estimand | | Trial-only $\widetilde{\tau}_t$ | Without validation set | | With validation set | |
|---|---|---|---|---|---|---|
| | | | Augmented $\widehat{\tau}_t$ | Shrinkage $\widehat{\tau}_t^{(s)}$ | Augmented $\widehat{\tau}_t$ | Shrinkage $\widehat{\tau}_t^{(s)}$ |
| $\tau_1$ | Est. (SD) | 2.418 (0.551) | 2.399 (0.568) | 2.444 (0.543) | 2.416 (0.546) | 2.411 (0.542) |
| $\tau_0$ | Est. (SD) | 1.567 (0.436) | 1.686 (0.429) | 1.787 (0.427) | 1.647 (0.355) | 1.653 (0.355) |
| $\tau$ | Est. (SD) | 0.851 (0.702) | 0.713 (0.711) | 0.907 (0.701) | 0.769 (0.649) | 0.759 (0.648) |

- Shrinkage SDs are $\leq$ trial-only across $\tau_1, \tau_0, \tau$ — the efficiency safeguard holds.
- With a validation set, the EC-augmented estimators show clearer variance reductions (especially $\tau_0, \tau$), and shrinkage preserves these gains.

# Takeaways

- We integrate external controls by modeling the distribution shift rather than assuming exchangeability.
- EC-augmented estimating functions exploit *both* covariate and concept shifts $\Rightarrow$ improve inference for *both* arms.
- Practical augmented estimators by directly modeling / estimating the shifts.
- Adaptive shrinkage estimators:
  - remains consistent under shift-model misspecification;
  - is asymptotically never less efficient than the trial-only benchmark.
- Finite-sample gains confirmed on synthetic experiments and the NSW/PSID benchmark.

# Thank you!

J. Shan, Y. Tu, G. Wang, C. Ying, and J. Zhao. Beyond exchangeability: Distribution-shift-aware integration of external control data in randomized trials. *arXiv preprint arXiv:2605.28785*, 2026

jiawei.shan@wisc.edu

https://jw-shan.github.io/

# References I


B. Colnet, I. Mayer, G. Chen, A. Dieng, R. Li, G. Varoquaux, J.-P. Vert, J. Josse, and S. Yang. Causal inference methods for combining randomized trials and observational studies: A review. *Statistical Science*, 39(1):165–191, 2024.

I. J. Dahabreh, S. E. Robertson, E. J. Tchetgen, E. A. Stuart, and M. A. Hernán. Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals. *Biometrics*, 75(2):685–694, June 2019. ISSN 0006-341X.

R. Dehejia and S. Wahba. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94(448):1053–1062, Dec. 1999. ISSN 0162-1459.

C. Gao, S. Yang, M. Shan, W. Y E, I. Lipkovich, and D. Faries. Improving randomized controlled trial analysis via data-adaptive borrowing. *Biometrika*, Dec. 2024. ISSN 0006-3444, 1464-3510.

R. Lalonde. Evaluating the econometric evaluations of training-programs with experimental-data. *American Economic Review*, 76(4):604–620, Sept. 1986. ISSN 0002-8282.

X. Li, W. Miao, F. Lu, and X.-H. Zhou. Improving efficiency of inference in clinical trials with external control data. *Biometrics*, 79(1):394–403, 2023. ISSN 1541-0420.

S. J. Pocock. The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases*, 29 (3):175–188, 1976.

J. M. Robins, A. Rotnitzky, and L. Zhao. Estimation of regression-coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866, Sept. 1994. ISSN 0162-1459.

# References II


E. T. Rosenman, G. Basse, A. B. Owen, and M. Baiocchi. Combining observational and experimental datasets using shrinkage estimators. *Biometrics*, 79(4):2961–2973, 2023.

J. Shan, Y. Tu, G. Wang, C. Ying, and J. Zhao. Beyond exchangeability: Distribution-shift-aware integration of external control data in randomized trials. *arXiv preprint arXiv:2605.28785*, 2026.

M. Valancius, H. Pang, J. Zhu, S. R. Cole, M. Jonsson Funk, and M. R. Kosorok. A causal inference framework for leveraging external controls in hybrid trials. *Biometrics*, 80(4):ujae095, 2024.

S. Yang, C. Gao, D. Zeng, and X. Wang. Elastic integrative analysis of randomised trial and real-world data for treatment heterogeneity estimation. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85 (3):575–596, July 2023. ISSN 1369-7412.