Integrating diverse evidence sources in clinical research: bridging RCTs and RWD

The interface of statistics, AI/ML, and real-world evidence

Shu Yang

Department of Statistics, North Carolina State University

· syang24@ncsu.edu

NC STATE UNIVERSITY

Roadmap

1. Overview of RWD/E in clinical and regulatory studies
2. The interface of RCT, RWD/E, statistics, and AI/ML
3. Combining conventional wisdom & new techniques: data integration and AI + statistical rigor
4. Summary and path forward

The 21st Century Cures Act & FDA’s RWE Framework

2016 · PUBLIC LAW

21st Century Cures Act

Signed into law December 13, 2016 — designed to accelerate medical product development and bring innovations to patients faster and more efficiently.

2018 · FDA GUIDANCE

RWE Framework

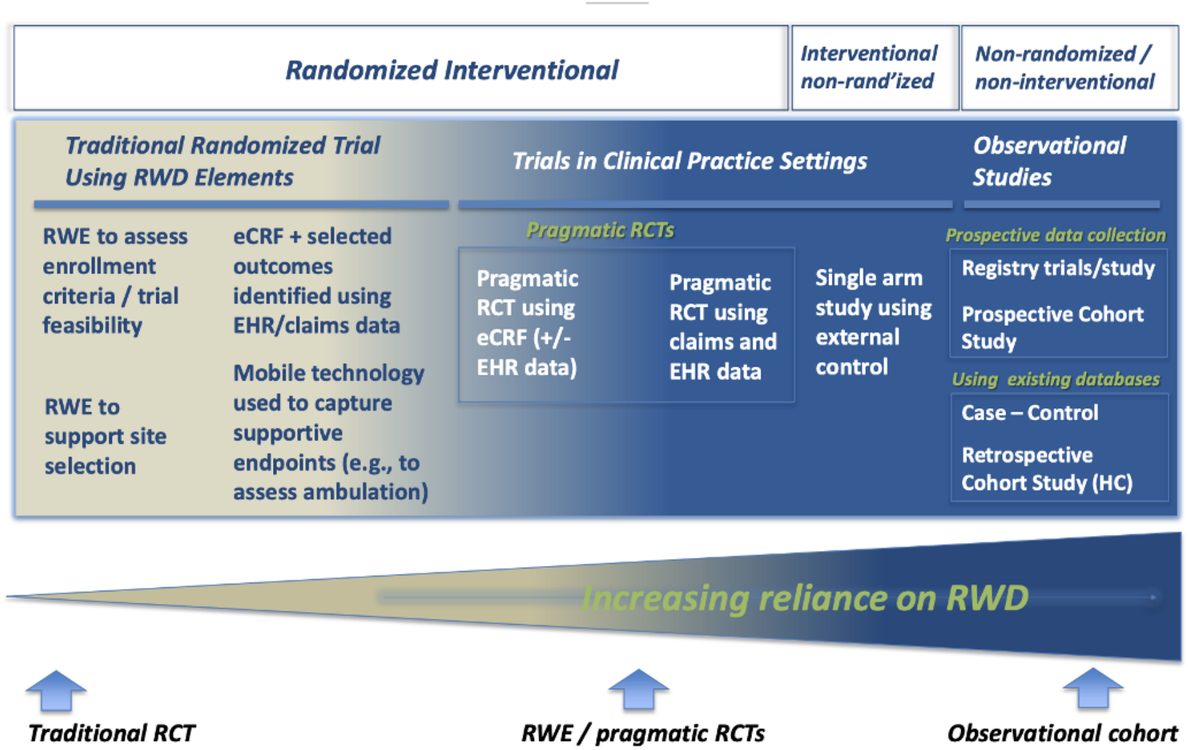
A structured framework (FDA, 2018) for evaluating the potential use of real-world evidence — new indications, label expansions, and post-market safety.

RWD in Clinical Research: an Evolving Landscape

Real-world data now spans a **continuum** of sources feeding regulatory evidence:

- electronic health records
- claims & billing
- disease registries
- digital health / wearables
- patient-generated data

From peripheral to central: analyzed rigorously, RWD answers questions beyond RCT scope.



A spectrum of designs, not a binary of "trial vs. no trial."

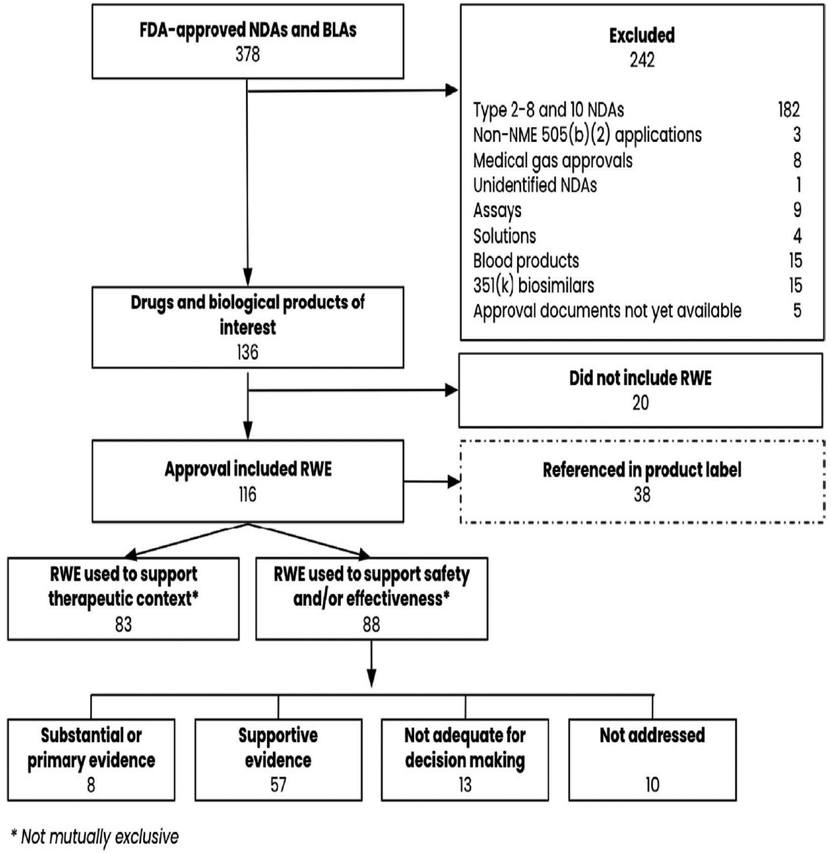
Stein (2019), Regulatory use of real-world evidence

FDA has Approved Applications with RWE

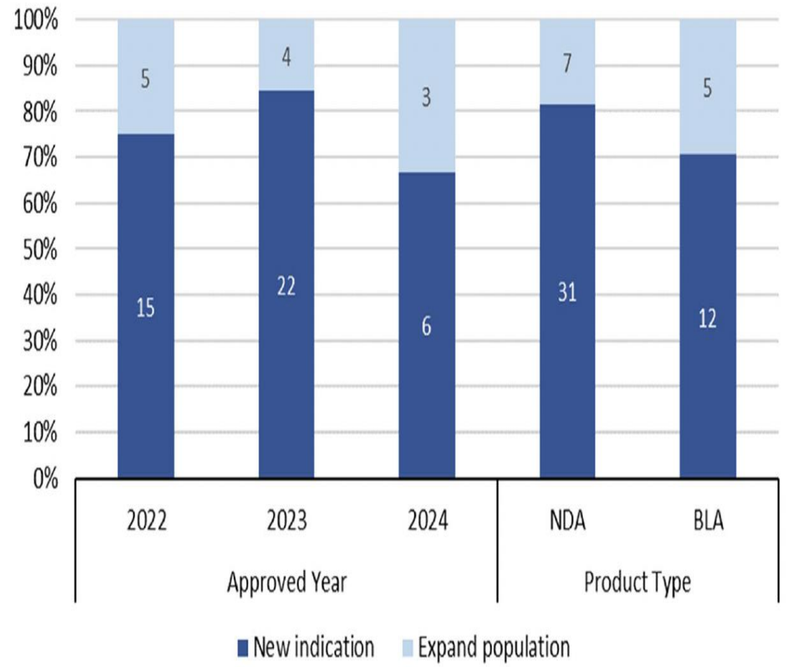
31%

of FDA new-drug & biologic approvals (2019–2021) were supported by real-world evidence.

Signals regulatory willingness and readiness.



Inclusion of RWE in FDA-approved NDAs and BLAs between Jan 2019 and June 2021 (Purpura et al. , 2022)



Purpose of Approvals with RWE by Year and Product Type. Abbreviation: NDA: new drug application; BLA: biologics license application (Deng et al., 2025)

Roadmap

1 Overview of RWD/E in clinical and regulatory studies

2 The interface of RCT, RWD/E, statistics, and AI/ML

3 Combining conventional wisdom & new techniques: data integration and AI + statistical rigor

4 Summary and path forward

RCT vs. RWD

	RCT	RWD
Randomization	✓	✗**
Representativeness	✗*	✓
Patient diversity	✗	✓
Sample size	small	large
Follow-up	short	long
Cost	high	low

✗* e.g., only ~8% of cancer patients participate in oncology trials.

✗** e.g., unmeasured confounding without randomization.

Conventional gold standard: **RCT** . Newly emerging: **RWD** . The two are complementary.

Statistics vs. AI/ML — and their complementary advantage

ASPECT	STATISTICS	AI / ML	COMPLEMENTARY
Primary goal	Inference & explanation	Prediction & pattern recognition	Combine statistical inference with predictive power
Assumptions	Explicit, often strong model assumptions	Minimal distributional assumptions	Validate flexible models with rigorous theory
Data types	Structured, tabular	High-dimensional & unstructured (image, text, sensor)	Broaden scope to diverse data sources
Interpretability	High — parameters & effect sizes	Often low (“black box”)	Use statistics for explainability of ML
Uncertainty	CIs, hypothesis tests, p-values	Limited; resampling / Bayesian approx.	Combine ML predictions with valid inference
Model flexibility	Predefined functional forms	Captures nonlinearities & interactions	Improve fit while preserving interpretability
Strength in practice	Regulatory acceptance, reproducibility	State-of-the-art predictive performance	Apply the right tool to dataset size & complexity

AI Role Across the Drug-Development Lifecycle

UPSTREAM

Feature extraction

Phenotyping, endpoint extraction from clinician notes and images, and covariate enrichment.

MIDSTREAM

Flexible modeling

Flexible nuisance modeling, high-dimensional propensity & outcome models, treatment-effect heterogeneity, synthetic controls and digital twins.

DOWNSTREAM

Living evidence

Automated reporting & archiving, model-drift monitoring, continuously updating pipelines for surveillance.

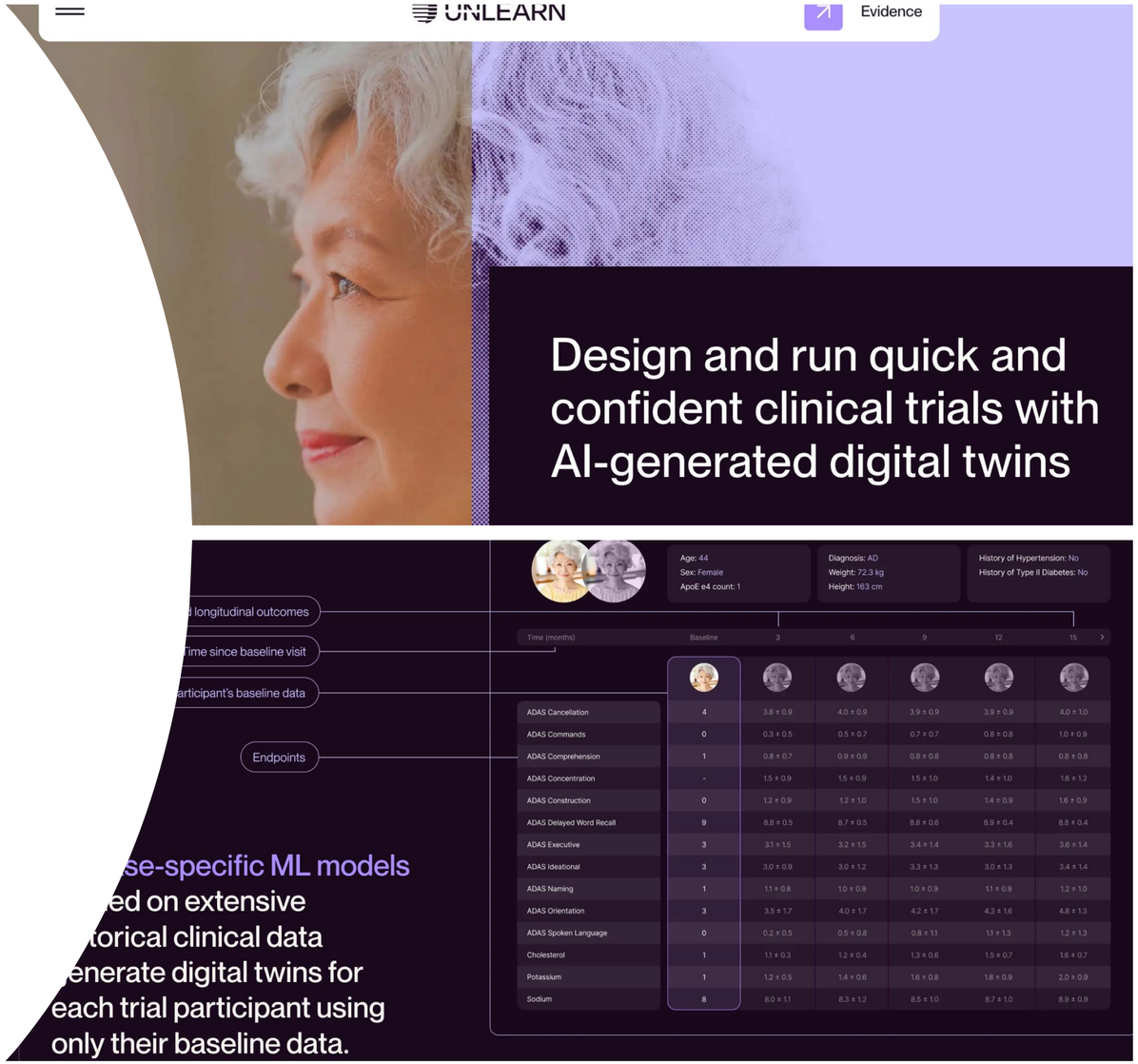
Predictive power and automation alone do not confer **causal validity** or **regulatory credibility** that is the statistician's domain.

AI-generated Digital Twins

Core idea: combining mechanistic disease models with AI/ML to forecast outcomes

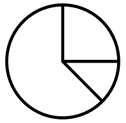
Example: Unlearn's trial using disease progression + Neural Networks for patient-level simulations

Key challenges: Model evaluation and potential bias if RWD for training are biased.



NC STATE UNIVERSITY

Bias Concerns with RWD



- **Baseline incomparability / covariate shift ← observed bias**
- **Selection bias**

✓ External data can differ from the RCT in baseline disease characteristics.



- **Outcome incomparability / outcome drift ← hidden bias**
- **Unmeasured confounding**:



✓ Critical patient disease characteristics may not have been assessed or may have been assessed differently.

- **Lack of concurrency**:



✓ RCT and external data may have been conducted in different time frames or care settings.

- **Measurement error**:

✓ Data collection intervals and quality may lack consistency.



- **Outcome validity**:

✓ Outcomes in external data sources may be measured differently than in RCTs or they may not be well-defined or reliable.

Innovation must not come at the cost of validity. The enabler is …

FDA (2019). Rare diseases: Natural history studies for drug development.
FDA (2023). Considerations for the Design and Conduct of Externally Control trials.

Roadmap

1 Overview of RWD/E in clinical and regulatory studies

2 The interface of RWD/E, statistics, and AI/ML

3 **Enablers:** Combining conventional wisdom & new techniques: data integration and AI + statistical rigor

4 Summary and path forward

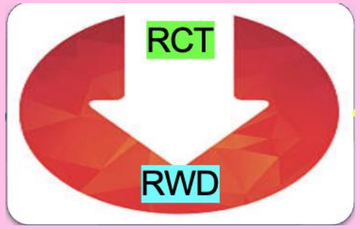
Research questions that can be addressed by data integration

Causal inference methods for combining RCT & RWD

Yang & Wang (2022) ASA Biop · Colnet et al. (2023) Stat Sci · Zhu et al. (2026) arXiv HCTs · github.com/IntegrativeStats

Leveraging RWD to improve the generalizability of RCT findings

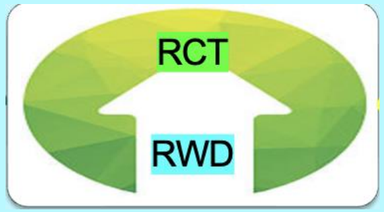
ATE · average treatment effect; ITR · individual treatment regimes



ATE	Lee et al. (2021–2024) Biom, JCI, JBS
ITR	Wu & Yang (2022) JCGS; Chu et al. (2023) ICML; Zhao et al. (2025) JRML
ITR with summary stat	Chu et al. (2023) Biometrika

Leveraging RWD to improve RCT analysis efficiency

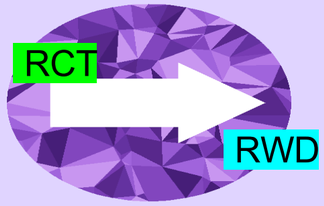
ATE · average treatment effect; HTE · heterogeneity of treatment effect



ATE: control var	Yang & Ding (2020) JASA
HTE: Test-then-pool	Yang et al. (2023) JRSSB; Gao & Yang (2023) EJS
ATE/HTE: Bias modeling	Wu & Yang (2022); Cheng et al. (2023); Yang et al. (2025); Mao et al. (2025)
ATE: Selective borrowing	Gao et al. (2024–2025); Zhu et al. (2025); Liu et al. (2025)

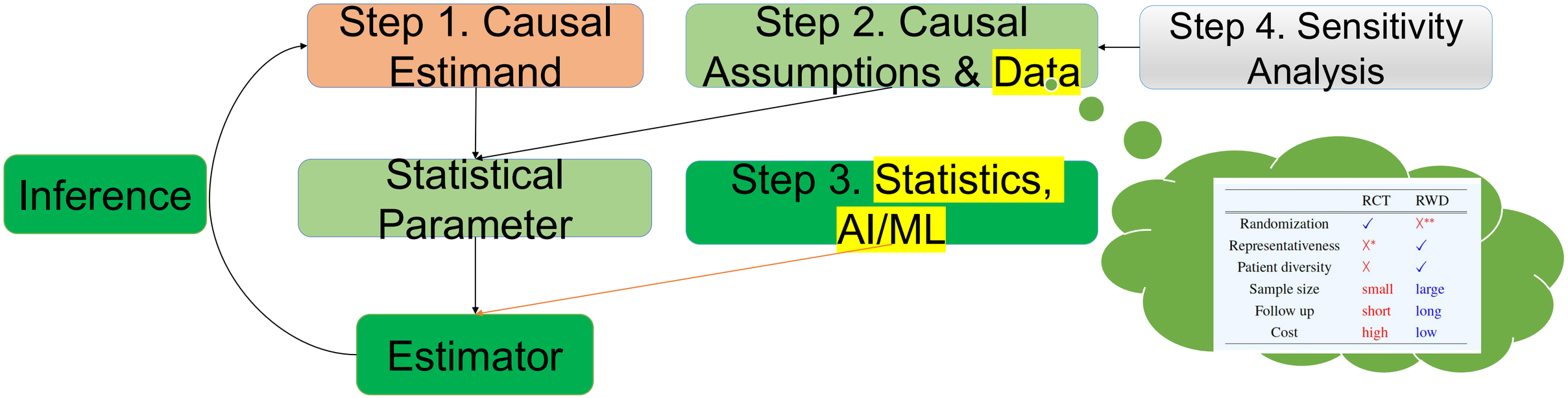
Leveraging RWD to enable long-term treatment effect estimation

Short-term RCT + long-horizon RWD; Proximal DiD (Park et al., 2026)



Proximal DiD	Park et al. (2026)
Long-term safety	Monitor adverse events beyond the RCT window with RWD
Long-term effectiveness	Estimate durable effects by linking short-term RCT to long-horizon RWD

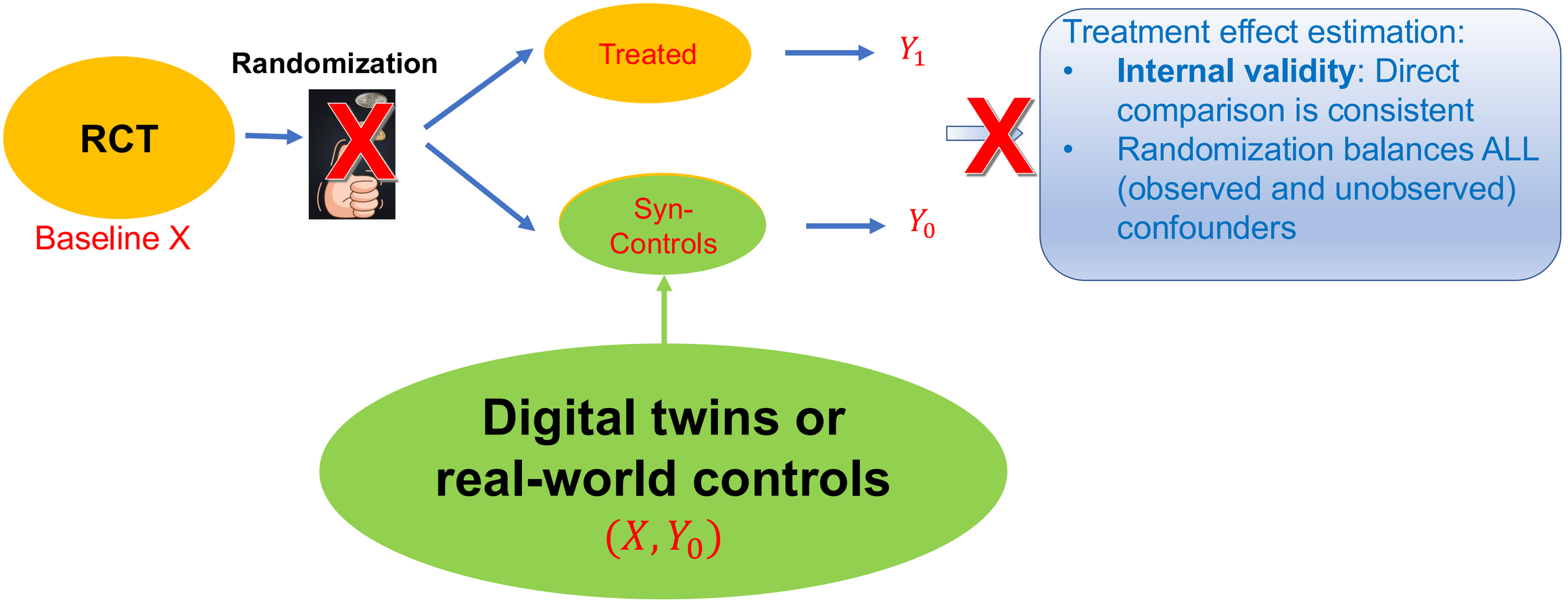
A Causal Roadmap for Evidence Integration



- Clarify inferential goals and estimands in an integration context
- Identify and address bias: covariate shift and outcome drift.
- Articulate assumptions and identification strategies explicitly.
- Leverage statistical principles to guide design and analysis choices.

Single-Arm Trial: Synthetic Controls

Readiness: low

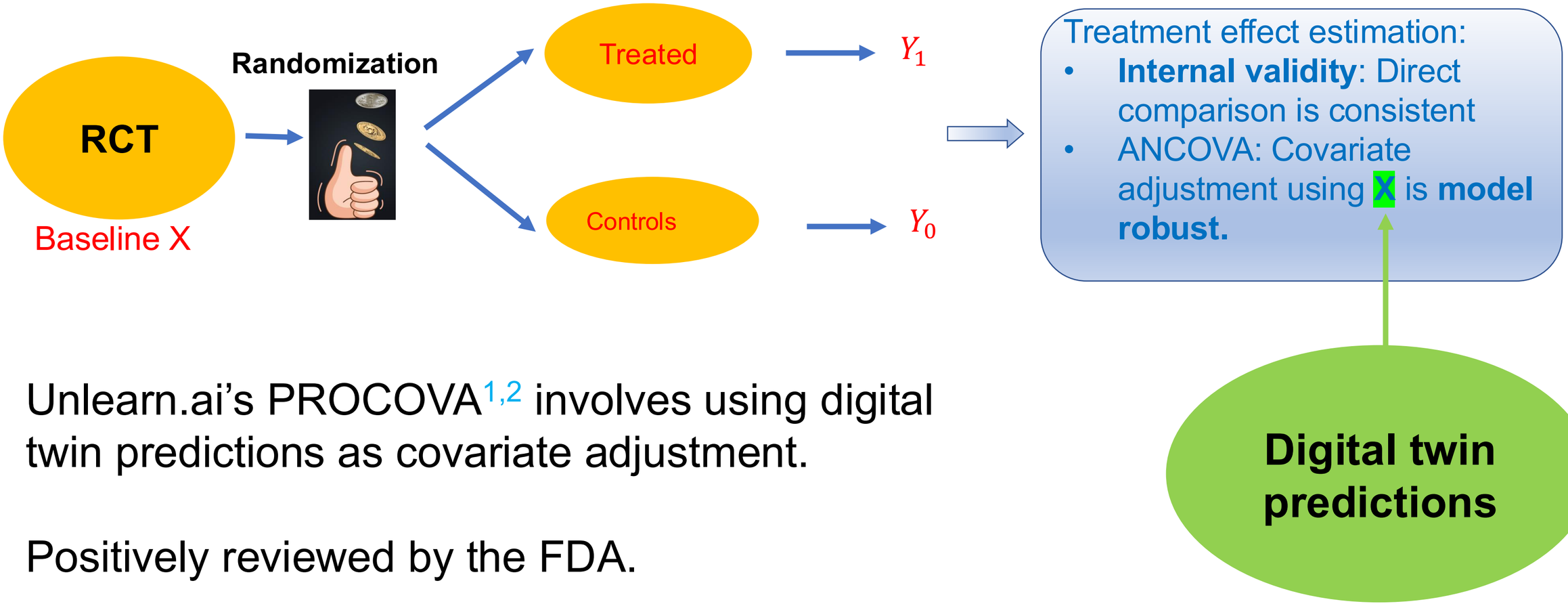


Treatment effect estimation:

- **Internal validity**: Direct comparison is consistent
- Randomization balances ALL (observed and unobserved) confounders

Super covariates

Readiness: High

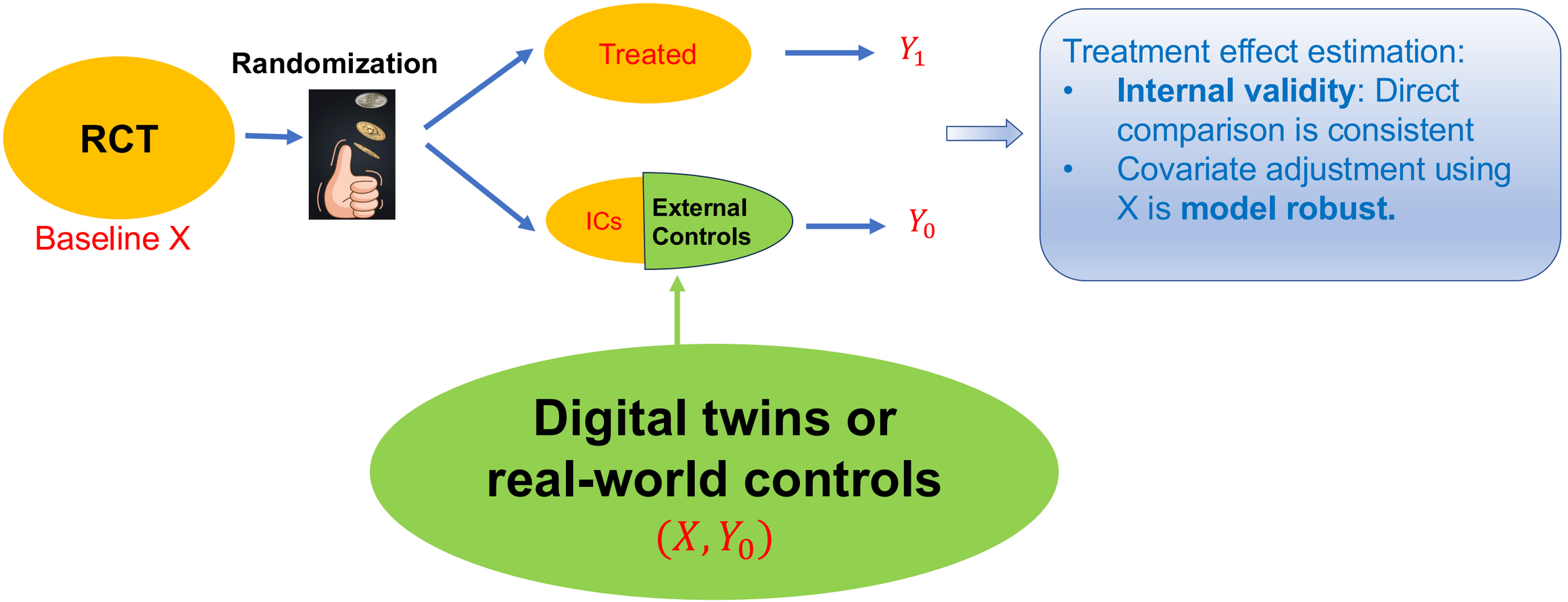


- Unlearn.ai's PROCOVA[1,2] involves using digital twin predictions as covariate adjustment.
- Positively reviewed by the FDA.

[1] Ross et al. (2024) Enhancing longitudinal clinical trial efficiency with digital twins and PROCOVA-MMRM
[2] Liu-Zhu-Han-Yang (2025) COADVISE: covariate adjustment with variable selection in randomized controlled trials

Hybrid Controlled Trials / Trial Augmentation

Readiness: Mod- High



Treatment effect estimation:

- **Internal validity**: Direct comparison is consistent
- Covariate adjustment using X is **model robust.**

HCTs: Basic Setup

- $Y(a)$ the potential outcome had the patient received treatment a
- Causal Estimand

$$\tau = E\{Y(1) - Y(0) \mid RCT\}$$

Data

- $\{(X_i, A_i, Y_i, S_i = 1) : i \in RCT\}$
- $\{(X_j, A_j = 0, Y_j, S_j = 0) : j \in EC\}$

Goals

- Statistical methods that borrow EC to improve RCT analysis while mitigating all possible biases from EC

THE GOVERNING QUESTION

“How to **borrow** comparable external controls, and **discard or down-weight** incomparable ones?”

Bayesian Methods

Informative priors that **adaptively down-weight** external data when it disagrees with the concurrent controls:

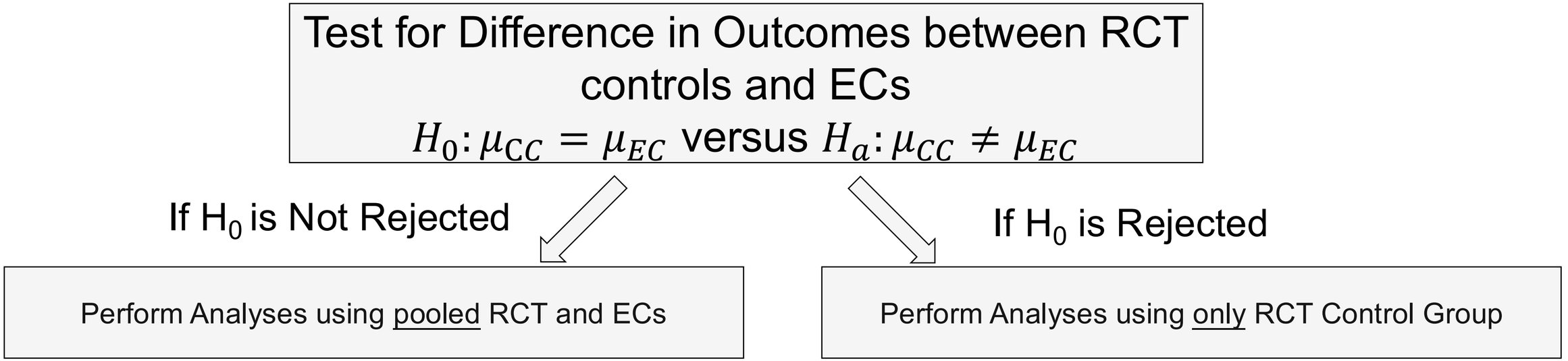
- Power prior (Ibrahim et al, 2000)
- Commensurate prior (Hobbs et al, 2012)
- Robust meta analytic predictive (MAP) prior (Schmidli et al, 2014)
- Elastic prior (Jiang et al, 2021)
- …

NC STATE UNIVERSITY

Frequentist Methods (bias-aware)

- Test then Pool (Viele, 2012; Yang et al. 2023 JRSSB)
- Confounding/bias function modeling (Yang et al. 2025 Bernoulli; Cheng et al. 2024 UAI; Wu & Yang 2022 CLEAR)
- Semiparametric efficient estimator with selective borrowing (Gao et al. 2024, 2025)
- Randomization-based inference with conformal selective borrowing (Zhu et al. 2025, Liu et al. 2026)

Test-Then-Pool



Pros: Intuitive; most common in practice

Cons: Yang et al 2023[1] address post-selection inference and show that TTP requires **strong signals** of violation and **moderate sample size** of RCT controls. **Wasteful** if some ECs are comparable

[1] Yang et al 2023 JRSSB Elastic integrative analysis of randomized trial and real-world data for treatment heterogeneity estimation

MOTIVATING THE NEXT METHOD

“Can data from **biased** external controls still be used to **improve** the RCT analysis?”

Cut or Filter

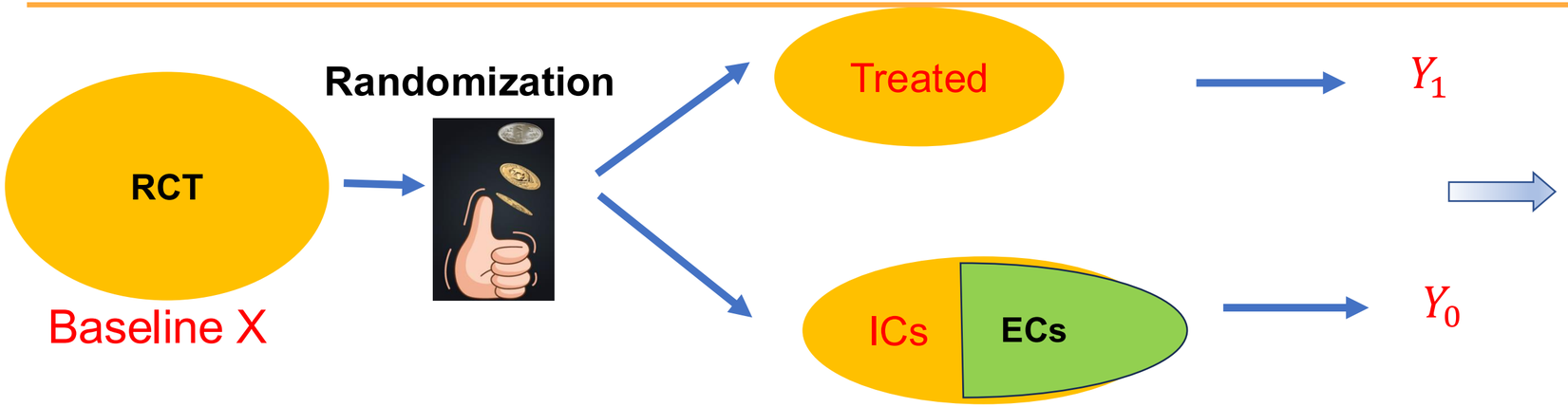


Creates a mathematical 'bias function' to model and subtract the bias, correcting the external data.



Uses advanced methods to 'selectively' borrow only the comparable parts of the external data.

Bias Function Modeling



Treatment effect estimation:

- Simple ANCOVA utilizing X and is **NOT** model robust anymore.

Model in RCT → A working ANCOVA model $\bar{\mu}_{A,1}(X;\beta_*)$

Model in EC → Define the bias function[1] quantifying the difference between the EC outcome mean and the working ANCOVA model

$$\boldsymbol{\mu_{0,0}(X)} = \bar{\mu}_{0,1}(X;\beta_*) + b_0(X)$$

[1] Wu&Yang 2022 CLEAR; Cheng et al 2023 UAI; Yang et al 2025 Bernoulli

ANCOVA augmented by the bias function

$$\bar{\mu}_{A,S}(X;\beta,\delta) = \underbrace{\beta_0 + \beta_A A + \beta_X^T X}_{ANCOVA} + \underbrace{(1-S)\delta^{\mathrm{T}} g(X)}_{bias\ function}$$

- Least-squared estimation with regularization provides a consistent estimator for the average treatment effect:

$$\left(\hat{\beta},\hat{\delta}\right) = \underset{\beta,\delta}{\operatorname{argmin}} \left[\sum_{i=1}^{N} \left\{ Y_i - \bar{\mu}_{A_i,S_i}(X_i;\beta,\delta) \right\}^2 + N \sum_{j=1}^{K_1} P_{\lambda_{1,j}}(|\beta|) + N \sum_{j=1}^{K_2} P_{\lambda_{2,j}}(|\delta|) \right]$$

Pros and Cons

Pros:

- Assumption lean
- The bias function captures all possible EC biases
- Always gain efficiency compared to the RCT-only ANCOVA analysis; the gain is large when the true bias function is simple

Cons:

- May not be statistically optimal.

PUSHING FURTHER

“Statistical **optimality**?”

with selective borrowing

- **Assumption 1.** (RCT validity)
  Randomization of treatment assignment and positivity
  (holds by default)
- **Assumption 2.** (EC outcome comparability)

$$E\{Y_j(0) \mid X_j, S_j = 0\} = E\{Y_j(0) \mid X_j, S_j = 1\} \text{ for } j \in EC$$

  Allows for possible covariate shift but not for other sources of bias.
  (an idealistic assumption)

Two objectives / steps[1]

- **Step 1 (address covariate shift)**. If EC outcomes are comparable (Assumption 2 holds), develop an efficient and robust estimator of the ATE.
- **Step 2 (address outcome drift).** Assumption 2 may be violated. Design selective borrowing to mitigate such biases.

[1] Gao et al 2024 BMK Improving RCT analysis with data-adaptive borrowing
Gao et al, 2025 ICML Doubly protected estimation for survival outcomes utilizing external controls for RCTs.

Address EC covariate shift

- By semiparametric theory, we propose

$$\widehat{\tau}_{\mathrm{ACW}} = \frac{1}{N_{\mathcal{R}}} \sum_{i \in \mathcal{R} \cup \mathcal{E}} S_i \left[\widehat{\mu}_1(X_i) + \frac{A_i\{Y_i - \widehat{\mu}_1(X_i)\}}{\pi_A(X_i)} \right]$$

$$- \frac{1}{N_{\mathcal{R}}} \sum_{i \in \mathcal{R} \cup \mathcal{E}} \left[S_i \, \widehat{\mu}_0(X_i) + \left\{ \frac{\{ S_i \, (1 - A_i) + (1 - S_i) \, \widehat{r}_i(X_i) \}}{\widehat{q}(X_i)\{1 - \pi_A(X_i)\} + \widehat{r}(X_i)} \right\} \widehat{q}(X_i)\{Y_i - \widehat{\mu}_0(X_i)\} \right]$$

where

$$r(X) = \mathrm{var}\{Y(0) \mid X, S = 1\} / \mathrm{var}\{Y(0) \mid X, S = 0\},$$

$$q(X) = P(S = 1 \mid X) / P(S = 0 \mid X).$$

Address EC covariate shift



- By semiparametric theory, the ACW estimator is

$$\widehat{\tau}_{\mathrm{ACW}} = \frac{1}{N_{\mathcal{R}}} \sum_{i \in \mathcal{R} \cup \mathcal{E}} S_i \left[\widehat{\mu}_1(X_i) + \frac{A_i\{Y_i - \widehat{\mu}_1(X_i)\}}{\pi_A(X_i)} \right]$$

$$- \frac{1}{N_{\mathcal{R}}} \sum_{i \in \mathcal{R} \cup \mathcal{E}} \left[S_i \, \widehat{\mu}_0(X_i) + \left\{ \frac{\{ S_i \, (1 - A_i) + \, (1 - S_i) \, \widehat{r}_i(X_i)\}}{\widehat{q}(X_i)\{1 - \pi_A(X_i)\} + \widehat{r}(X_i)} \right\} \widehat{q}(X_i)\{Y_i - \widehat{\mu}_0(X_i)\} \right]$$

- Double robustness: $\widehat{\tau}_{\mathrm{ACW}}$ is consistent if either model for $\mu_0(X)$ or $q(X)$ is correct, not necessarily both.

- Local efficiency: if all are correct, $\widehat{\tau}_{\mathrm{ACW}}$ is semiparametrically efficient.

Address EC covariate shift

Super flexible.



- By semiparametric theory, the optimal estimator is

$$\widehat{\tau}_{\mathrm{ACW}} = \frac{1}{N_{\mathcal{R}}} \sum_{i \in \mathcal{R} \cup \mathcal{E}} S_i \left[\widehat{\mu}_1(X_i) + \frac{A_i\{Y_i - \widehat{\mu}_1(X_i)\}}{\pi_A(X_i)} \right]$$

$$- \frac{1}{N_{\mathcal{R}}} \sum_{i \in \mathcal{R} \cup \mathcal{E}} \left[S_i \, \widehat{\mu}_0(X_i) + \left\{ \frac{\{ S_i \, (1 - A_i) + \, (1 - S_i) \, \widehat{r}_i(X_i) \}}{\widehat{q}(X_i)\{1 - \pi_A(X_i)\} + \widehat{r}(X_i)} \right\} \widehat{q}(X_i)\{Y_i - \widehat{\mu}_0(X_i)\} \right]$$

- Flexible models: use AI/ML models for $\mu_0(X)$ or $q(X)$.

- Under relaxed conditions, $\widehat{\tau}_{\mathrm{ACW}}$ is root-n consistent, asymptotically Normal, and semiparametrically efficient.

NC STATE UNIVERSITY

Address outcome drift

- **The goal of Step 2** is to develop strategies to mitigate biases due to outcome drift.

- **Idea: selective borrowing**
  - Assign a bias parameter
    $$b_j = E\left(Y_j \mid X_j, S_j = 0\right) - E\left(Y_j \mid X_j, S_j = 1\right) \text{ for } j \in EC$$
  - Select only the ECs with $b_j \approx 0$ for Step 1 estimator.

Question: How to select the comparable EC set?

NC STATE UNIVERSITY

Translate selection to penalization

- 1. Construct an initial estimator $\widehat{\xi}$ for the bias parameter

$$\widehat{\xi}_i = \widehat{\mu}_{0,\mathcal{E}}(X_i) - \widehat{\mu}_{0,\mathcal{R}}(X_i), \quad i \in \mathcal{E}.$$

- 2. Use penalization to shrink $\tilde{b}$ toward zero

$$\tilde{\boldsymbol{b}} = \arg\min_{\boldsymbol{b}} \left\{ (\widehat{\boldsymbol{\xi}} - \boldsymbol{b})^{\top} \widehat{\Sigma}_{\xi}^{-1} (\widehat{\boldsymbol{\xi}} - \boldsymbol{b}) + \lambda_N \sum_{i \in \mathcal{E}} p(|b_i|) \right\},$$

where $\widehat{\Sigma}_{\xi}$ is the estimated variance of $\widehat{\boldsymbol{\xi}}$ and λ_N is the tuning parameter.

- 3. Borrow the selected comparable EC subset $\tilde{\mathcal{A}} = \{i : \tilde{b}_i = 0\}$.
- 4. Construct the ACW estimator based only on $\tilde{\mathcal{A}}$.

Properties

- **Adaptive ACW**

$$\widehat{\tau}_{\mathrm{ACW}}^{\mathrm{alasso}} = \frac{1}{N_{\mathcal{R}}} \sum_{i \in \mathcal{R} \cup \mathcal{E}} S_i \left[\widehat{\mu}_1(X_i) - \widehat{\mu}_0(X_i) + \frac{A_i\{Y_i - \widehat{\mu}_1(X_i)\}}{\pi_A(X_i)} \right]$$

$$- \frac{1}{N_{\mathbf{R}}} \sum_{i \in \mathbf{R} \cup \mathbf{E}} \left\{ \frac{S_i\ (1 - A_i) + \ (1 - S_i)\ \mathbf{1}(i \in \tilde{\mathcal{A}}) \widehat{r}_i(X_i)}{\widehat{q}(X_i)\{1 - \pi_A(X_i)\} + \widehat{r}(X_i)} \right\} \widehat{q}(X_i)\{Y_i - \widehat{\mu}_0(X_i)\},$$

- **Theorem (Adaptive ACW):**

Under some regularity conditions, we have $N^{1/2}(\widehat{\tau}_{\mathrm{ACW}}^{\mathrm{alasso}} - \tau) \to \mathcal{N}(0, V_\tau^{\mathrm{alasso}})$, where V_τ^{alasso} is strictly less than V_τ^{aipw} unless

$$\underbrace{r(X) = 0}_{Y|X,\ S=0 \text{ is extremely variant}} \quad \text{or} \quad \underbrace{P(b = 0 \mid X, S = 0) = 0}_{\text{no comparable subset available}}.$$

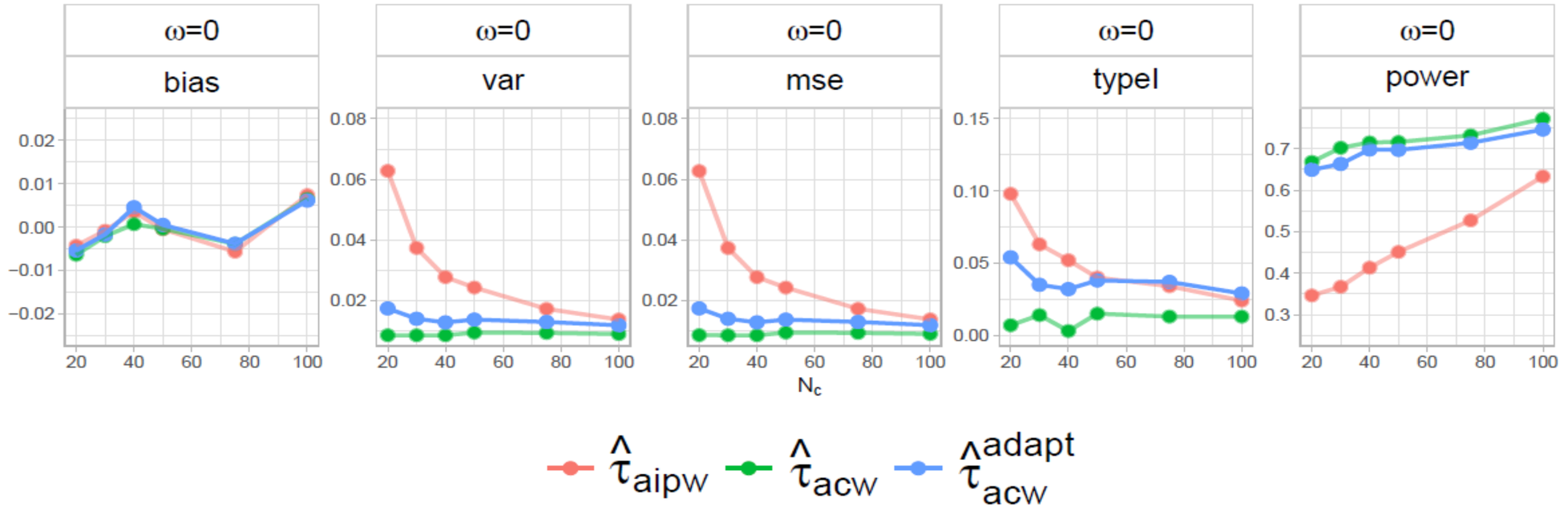
Designs

Goal: assess the RCT-only, ACW, and adaptive ACW estimators in terms of type-I error and power in various realistic settings.

Designs	
Covariates	12 baseline covariates X follow the correlation structure from an oncology RCT.
Sampling	$S \mid X, U \sim Bern\{\pi_S(X, U)\}$, where U is an unmeasured variable.
Outcome	$Y \mid X, A, U, S = s \sim N\{\mu_s(X, A, U), 1\}, s \in \{0,1\}$ *1. Assumption 2 holds* *2. Assumption 2 is violated due to unmeasured confounding*
Sample size	RCT control: $N_c = 20,30, \ldots, 100$ RCT treatment: $N_t = 200$ EC: $N_E = 3000$

Result I

Case 1: Assumption 2 holds.



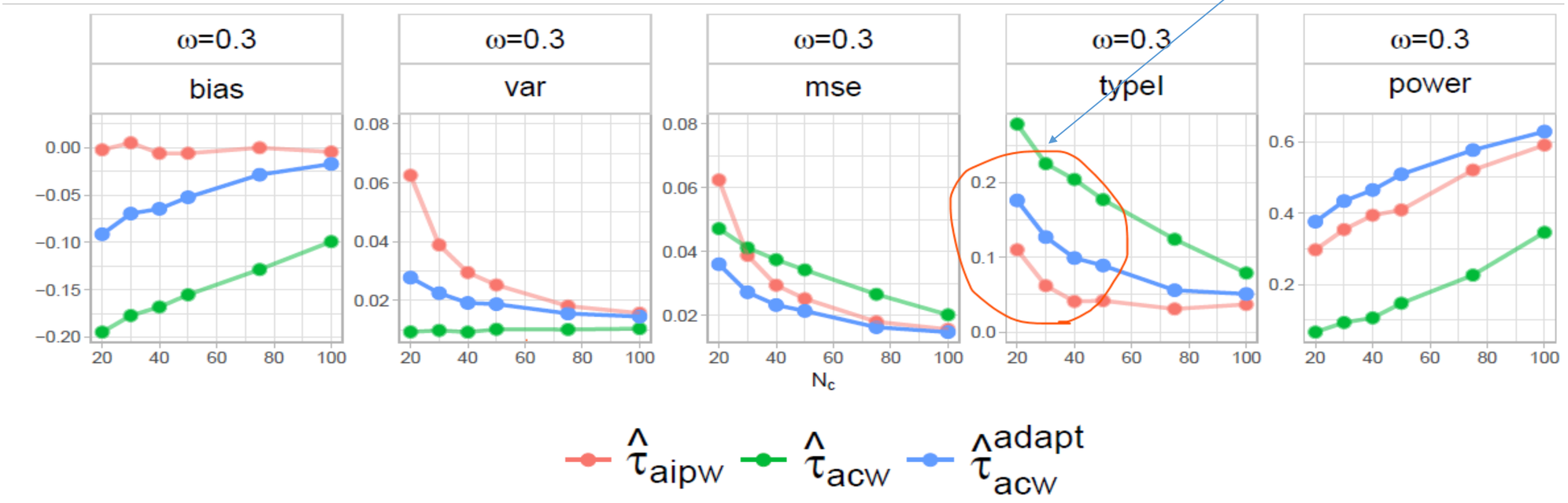
ACW is the best: small bias and var (gain eff over RCT-only, control type I error well, high power

Adaptive ACW is good: small bias and var (gain eff over RCT-only, control type I error well, high power

Result II

Type I error control is still poor for small samples!!

Case 2: Assumption 2 does not hold due to an unmeasured confounder.



ACW is the worst: **large** bias and small var, **cannot** control type I, low power

Adaptive ACW is still good: small bias and var (gain eff over RCT-only, control type I error well when Nc is large, high power

THE SMALL-SAMPLE CHALLENGE

“Can we **still control type-I error** and maintain **power gain** in **small samples**?”

Fisher Randomization Test (FRT)

Fisher Randomization Test (Fisher, 1935)

1. **Sharp Null**: $H_0 : Y_i(0) = Y_i(1), \forall i \in \mathcal{R}$, imputing all $Y_i(a)$.
2. **Test Statistic**: Compute $T(\boldsymbol{A}^{\text{obs}})$ for actual assignment $\boldsymbol{A}^{\text{obs}}$.
3. **Analyze as You Randomize**:
   - Generate A_i^b for RCT patients per the actual randomization procedure.
   - Keep $A_i^b \equiv 0$ for ECs, as they remain fixed during randomization in RCT.
4. **Compute p value**: Repeat for B iterations and compute:

$$\hat{p}^{\text{FRT}} = \frac{\sum_{b=1}^{B} \mathbb{I}\{T(\boldsymbol{A}^b) \geq T(\boldsymbol{A}^{\text{obs}})\} + 1}{B+1}.$$

- **Theorem 1 (Validity)**
  - FRT controls type I error rate for testing sharp null H_0
  - As long as A^b aligns with the true randomization A
  - Holds for any test statistic $T(A, X, Y)$, but which one should we use?

NC STATE UNIVERSITY

Fisher Randomization Test (FRT)

- **Theorem 2 (Model-based power analysis)**
  - Given that $T(A, X, Y)$ is consistent, smaller variance of $T(A, X, Y)$ → higher power

- With validity guaranteed, we need a consistent and efficient $T(A, X, Y)$ for higher power
  - No Borrowing AIPW (RCT-only)
  - Full Borrowing ACW (Addressing covariate shift)
  - Selective Borrowing ACW (Addressing covariate shift & outcome drift)
    - Penalization-based, IF-based, Conformal-based selection

Model-free

— and thus flexible to **fuse AI/ML** into testing and selection.

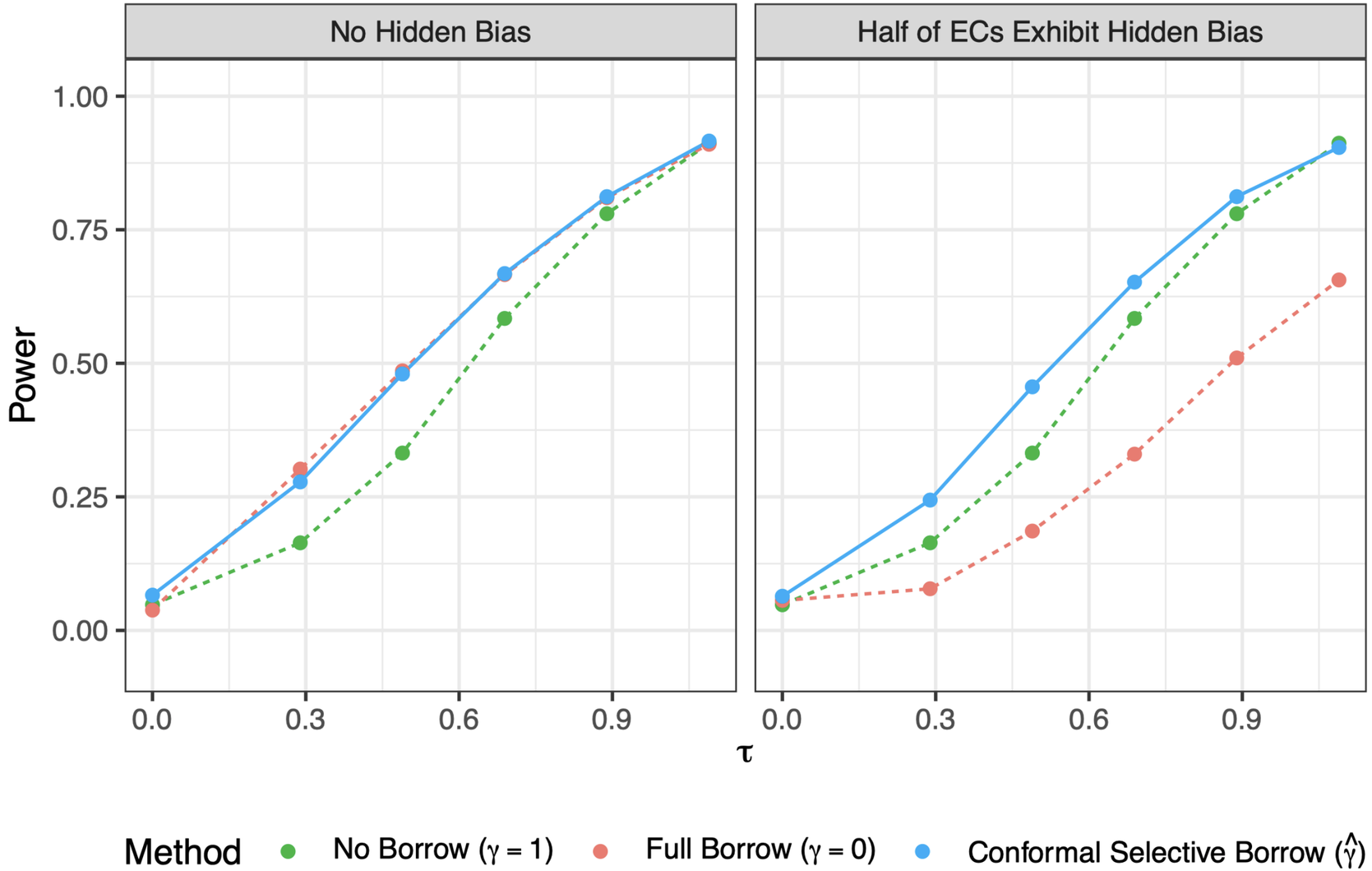
FRT + Selective Borrowing

- **Concern:** "Double use" of RCT controls for EC selection and treatment effect testing
- **Key action (FRT):** Selection is re-performed within each permutation, incorporating selection uncertainty into the reference distribution
- **Guarantee:** Validity of the FRT is preserved with selective borrowing

Setup

Sample size	RCT Treatment: $n_{\mathrm{IT}} = 50$ RCT Control: $n_{\mathrm{IC}} = 25$ External Control: $n_{\mathrm{EC}} = 50$
Covariates	2 baseline covariates $X \sim N(0,1)$
Sampling	$S \sim \mathrm{Bernoulli}\big(\pi(X)\big)$ where $\pi(X) = \big\{1 + \exp\big(\eta_0 + X^{\mathrm{T}}\eta\big)\big\}^{-1}$
Assignment	$A \sim \mathrm{Bernoulli}(n_{\mathrm{IT}}/n_{\mathrm{RCT}})$ for $S = 1$; $A = 0$ for $S = 0$
Potential Outcome	RCT Treatment: $Y(1) = 0.4 + X^{\mathrm{T}}\beta_1 + \varepsilon$ RCT Control: $Y(0) = X^{\mathrm{T}}\beta_0 + \varepsilon$ Half of External Control: $Y(0) = X^{\mathrm{T}}\beta_0 + 0.5\varepsilon$ Half of External Control: $Y(0) = -b + X^{\mathrm{T}}\beta_0 + 0.5\varepsilon,\ b = 0,1,2,\ldots,8$

Power Curve


No Hidden Bias
Half of ECs Exhibit Hidden Bias
Power
1.00
0.75
0.50
0.25
0.00
0.0
0.3
0.6
0.9
τ
Method
No Borrow (γ = 1)
Full Borrow (γ = 0)
Conformal Selective Borrow (γ̂)

A Bias-Aware Frequentist toolkit for RWD/DT borrowing

METHODS TO HANDLE	OBSERVED BIAS	HIDDEN BIAS	PROS	CONS
Test then Pool	Preliminary test		• Intuitive • Consistent	• Conservative inference • Possibly wasteful: study borrowing
Bias function modeling	Model the bias		• Simple modeling • Information borrowing • Consistent and gain efficiency	• May not be statistical optimal
Semiparametric efficient estimator	Calibration weighting	Penalty-based selective borrowing	• Individual borrowing • Doubly robust, local efficient, flexible to fuse AI&ML	• Based on large-sample theory
Fisher Randomization Test	Calibration weighting	Conformal selection	• Individual borrowing • Strictly control type-I error, even in small samples	• Sharp null hypothesis

NC STATE UNIVERSITY

Takeaways

01 The frontier is **integration** — not RCT-vs-RWD or statistics-vs-AI.

02 A **causal roadmap** disciplines every integrative analysis.

03 **Hybrid controlled trials** show the payoff:
randomization anchors, borrowing extends, selection guards.

04 Our **FRT + selective borrowing** controls type-I error even in small samples.

05 AI **empowers** prediction; statistics determines what is **credible**.

UNIVERSITY OF CLEMSON

NC STATE UNIVERSITY

Thank you!

Questions?

Shu Yang
Department of Statistics, North Carolina State University

syang24@ncsu.edu
github.com/IntegrativeStats · shuyang.wordpress.ncsu.edu