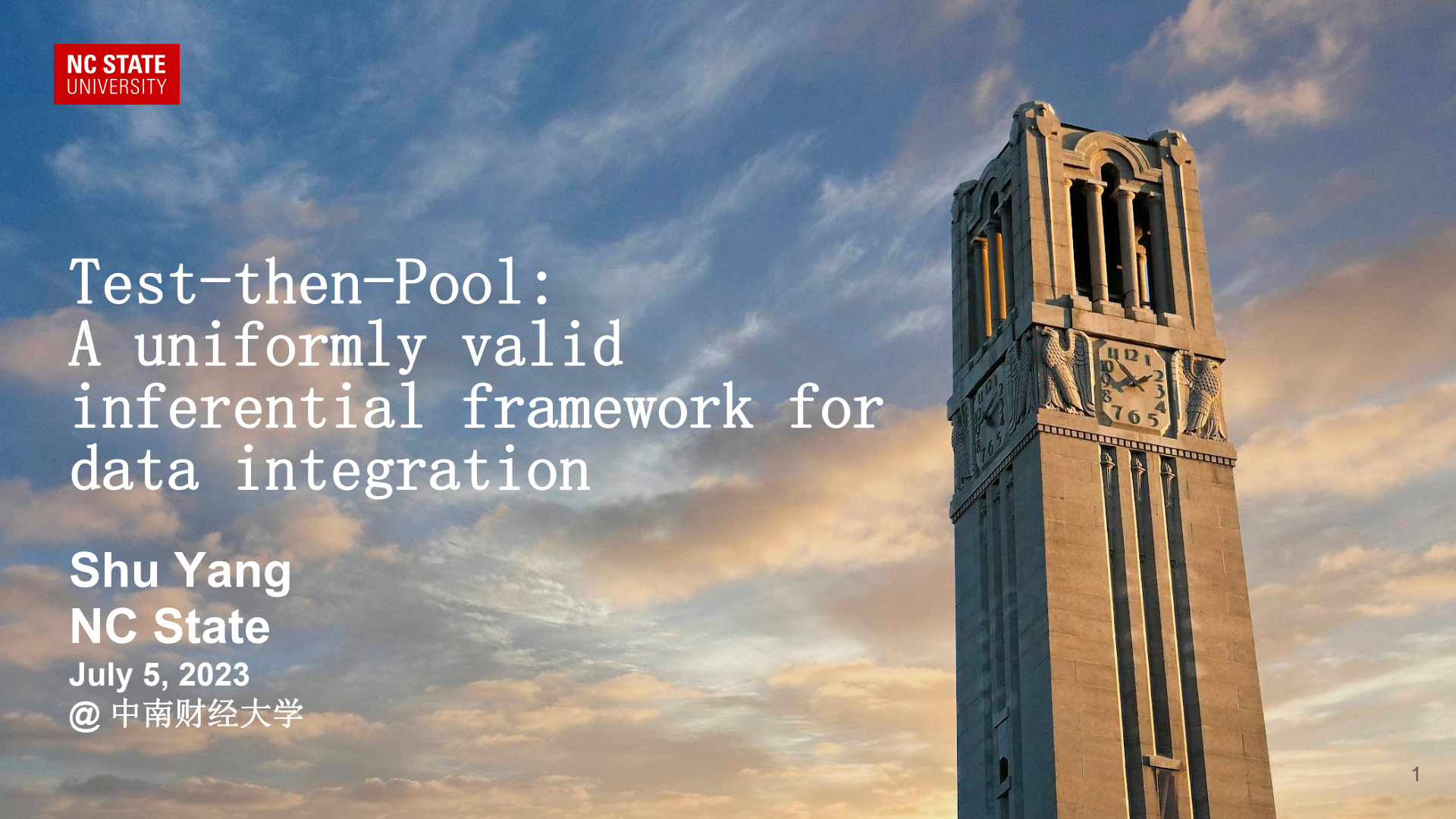


Test-then-Pool: A uniformly valid inferential framework for data integration

Shu Yang
NC State

July 5, 2023
@ 中南财经大学

Acknowledgements

Chenyin Gao
(NC State)

Xiaofei Wang
(Duke University)

Donglin Zeng
(UNC)

NIH National Institute on Aging

NIH-NIA grant 1R01AG066883

"Empower Treatment Effects Evaluation of Randomized Clinical Trials for Elderly Patients with Integrated Real-World Data."

Based on:

C. Gao and S. Yang (2023). Pretest estimation in combining probability and non-probability samples. *EJS*

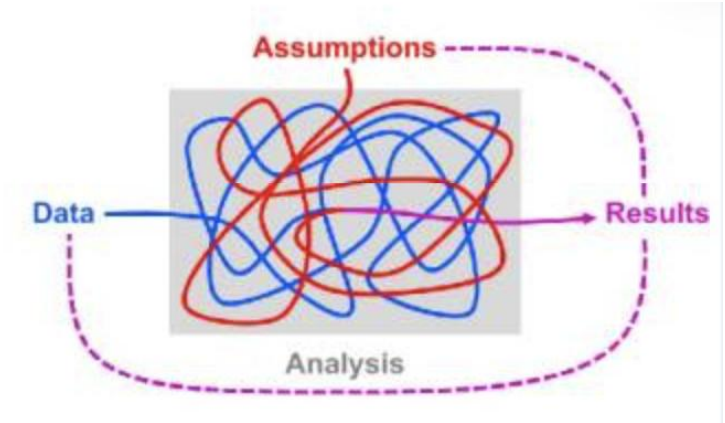
S. Yang, C. Gao, D. Zeng and X. Wang (2023). Elastic integrative analysis of randomized trial and real-world data for treatment heterogeneity estimation. *JRSSB*

NC STATE UNIVERSITY

Circular Analysis

Selective analysis/double dipping

- The selection of the details of a data analysis using the data that is being analyzed.
- Use the same data twice.
- Assumptions can interact with data to shape the results.



https://en.wikipedia.org/wiki/Circular_anal

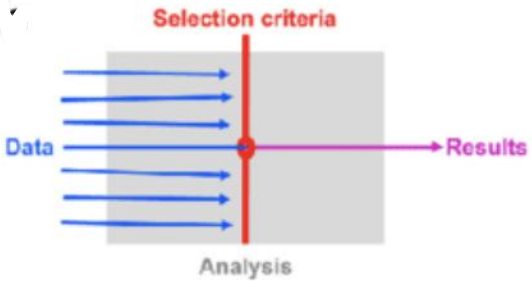
NC STATE UNIVERSITY

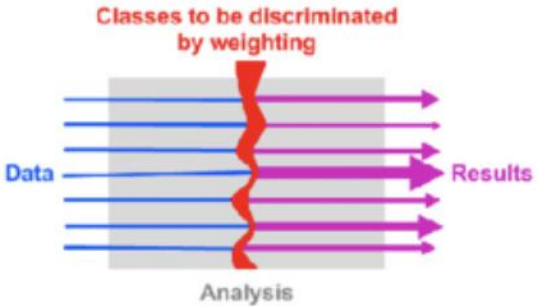
Circular Analysis

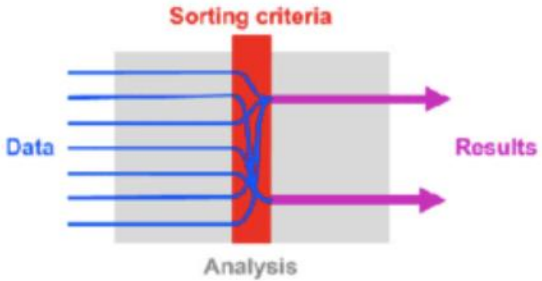
Selective analysis/double dipping

- Common causes of circularity: selection, weighting, and sorting, *consciously or unconsciously*.

- **Example 1.** Post-model or feature selection analysis
- **Example 2.** Model weighting or averaging
- **Example 3.** Pre-testing estimation







NC STATE UNIVERSITY

Circular Analysis
The danger of double dipping

- However, each of the three can tinge the results, distorting the estimate or invalidating statistical tests.

Yes, but really?

- Today, showcase an example that double dipping might be beneficial!

NC STATE UNIVERSITY

Circular Analysis in Data Integration

Multiple data sources

Multiple data sources are popular for research purposes.

Example 1. Causal analysis for treatment effect evaluation

- Conventional gold-standard randomized clinical trials (RCT)
- Newly emerging real-world observational data (RWD) (e.g., electronic health records)

Example 2. Survey sampling for population quantity analysis

- Conventional gold-standard probability samples
- Newly emerging non-probability samples (e.g., large web-panel data)

NC STATE UNIVERSITY

Data Integration
Complementarity

Data sources	Features 
RCT	Randomization of treatment; however, small sample size, less patient diversity, costly and time-consuming.
RWD	Large sample size, more patient diversity, cheaper and fast data collection; however, lack randomization of treatment.
Prob samples	Known sampling weights; however, costly and time-consuming, increasing nonresponses
Non-prob samples	Large sample size, cheaper and fast turn-around data collection; however, unknown sampling mechanism

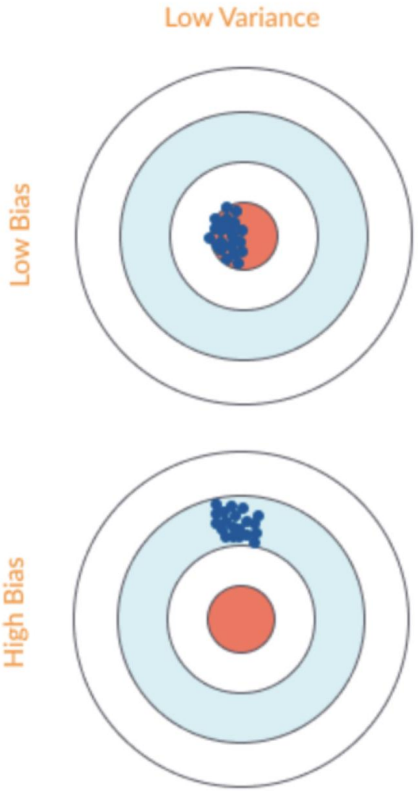
NC STATE UNIVERSITY

Data Integration

Test-then-Pool is a natural idea

It is appealing to integrate the small gold-standard sample with the large external sample to reduce variance.

Pretesting sample comparability is crucial to avoid biases from the external sample.



NC STATE UNIVERSITY

Data Integration

Test-then-Pool is a natural idea

"Test-then-Pool" is a natural idea, but double dipping!

We will need

1. an appropriate asymptotics framework to reveal its non-regularity;
2. an appropriate inference framework for uniformly valid inference.

NC STATE UNIVERSITY

Data Integration

General statistical setup

- V: the generic data variable
  - $V = (X, A, Y)$: the pretreatment covariates, treatment assignment, outcome
  - $V = (X, Y)$: the auxiliary variables and study variable of interest
- δ: $\delta = 1$ Sample A (the gold-standard study) and 0 in Sample B (the auxiliary study)

- Parameter of interest $\psi_0 \in \mathbb{R}^p$: defined through $\mathbb{E}\{S_{\psi_0}(V)\} = 0$, where $S \in \mathbb{R}^p$ is a score function, including
  - Average treatment effect, Marginal Structural Model parameters, etc
  - Population mean of Y

NC STATE UNIVERSITY

Data Integration

Two simple analysis strategies

- Sample A data score: $S_{\mathrm{A},\psi}(V) = \delta S_\psi(V) \in \mathbb{R}^p$
- Sample B data score: $S_{\mathrm{B},\psi}(V) = (1-\delta) S_\psi(V) \in \mathbb{R}^p$

- $\widehat{\psi}_{\mathrm{A}}$ solves $\mathbb{P}_n S_{\mathrm{A},\psi}(V) = 0$: small bias but large variance
- $\widehat{\psi}_{\mathrm{AB}}$ solves $\mathbb{P}_n\{S_{\mathrm{A},\psi}(V) + S_{\mathrm{B},\psi}(V)\} = 0$: small variance but possibly large bias
- $\mathbb{P}_n f(V) = n^{-1}\sum_{i=1}^n f(V_i)$

NC STATE UNIVERSITY

Circular Analysis in Data Integration

Test-then-Pool

Test for Comparability of Sample A and Sample B

Sample A is gold standard: $E\left[S_{\mathrm{A},\psi_0}(V)\right] = 0$ by design.

H_0: $E\left[S_{\mathrm{B},\psi_0}(V)\right] = 0$ versus H_a: $E\left[S_{\mathrm{B},\psi_0}(V)\right] = \eta_{\mathrm{fix}} \neq 0$

If H_0 is not rejected

Conduct joint analysis using pooled Sample A and Sample B

If H_0 is rejected

Conduct analysis using only Sample A

NC STATE UNIVERSITY

Test-then-Pool
Test statistics

Key insights

1. $\widehat{\psi}_{\mathrm{A}}$ is always consistent by design
2. $\widehat{S}_{\mathrm{B},\widehat{\psi}_{\mathrm{A}}}(V_i)$ is small under H_0 and large under H_a

The test statistic

$$T = \left\{ \sum_{i \in \mathscr{B}} \widehat{S}_{\mathrm{B},\widehat{\psi}_{\mathrm{A}}}(V_i) - 0 \right\}^{\mathrm{T}} \widehat{\Sigma}_{SS}^{-1} \left\{ \sum_{i \in \mathscr{B}} \widehat{S}_{\mathrm{B},\widehat{\psi}_{\mathrm{A}}}(V_i) - 0 \right\}$$

- $\widehat{\Sigma}_{SS}$ is a variance estimator for $\sum_{i \in \mathscr{B}} \widehat{S}_{\mathrm{B},\widehat{\psi}_{\mathrm{A}}}(V_i)$, e.g., obtained by the re-sampling method.

NC STATE UNIVERSITY

Test-then-Pool

Test-based integrative estimator

- $c_\gamma = \chi^2_{p,\gamma}$ the $(1-\gamma)$th percentile of χ^2_p

The test-then-pool integrative estimator $\widehat{\psi}_{\text{ttp}}$ solves

$$\sum_{i \in \mathscr{A} \cup \mathscr{B}} \left\{ \underbrace{\delta_i \widehat{S}_\psi(V_i)}_{\text{Sample A EE}} + \underbrace{I(T < c_\gamma)}_{\text{Combine or not}} \underbrace{(1-\delta_i)\widehat{S}_\psi(V_i)}_{\text{Sample B EE}} \right\} = 0$$

- $\widehat{\psi}_{\text{ttp}}$ mixes $\widehat{\psi}_{\text{A}} \mid (T \geq c_\gamma)$ and $\widehat{\psi}_{\text{AB}} \mid (T < c_\gamma)$

NC STATE UNIVERSITY

Test-then-Pool

Asymptotic distribution?

- $\widehat{\psi}_{\text{ttp}}$ belongs to the class of pre-test estimators
- Challenge
  - i) estimator depends on the random outcome of pre-testing
  - ii) test and estimator are constructed based on the same data

- What is the asymptotic distribution of $\widehat{\psi}_{\text{ttp}}$?

A. Normal

B. Mixture of Normals

C. Mixture of non-Normals

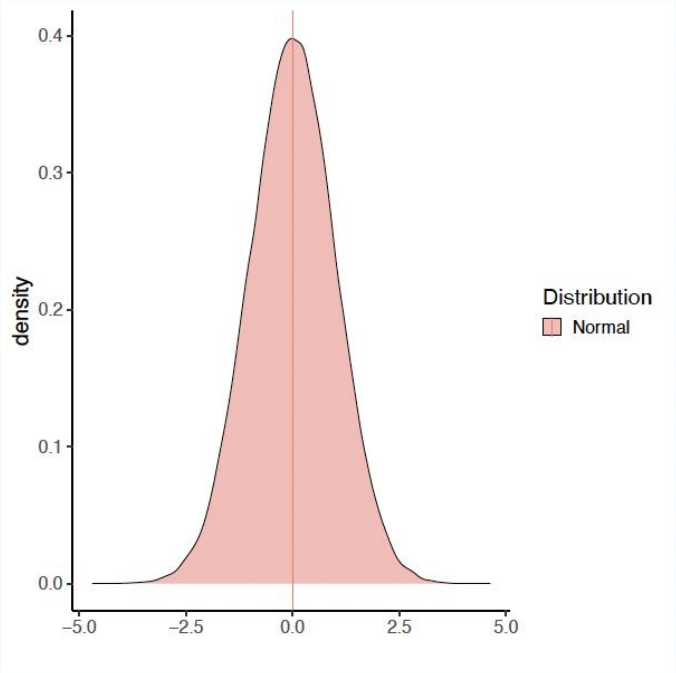
D. Mixture of Normal and non-Normal

NC STATE UNIVERSITY

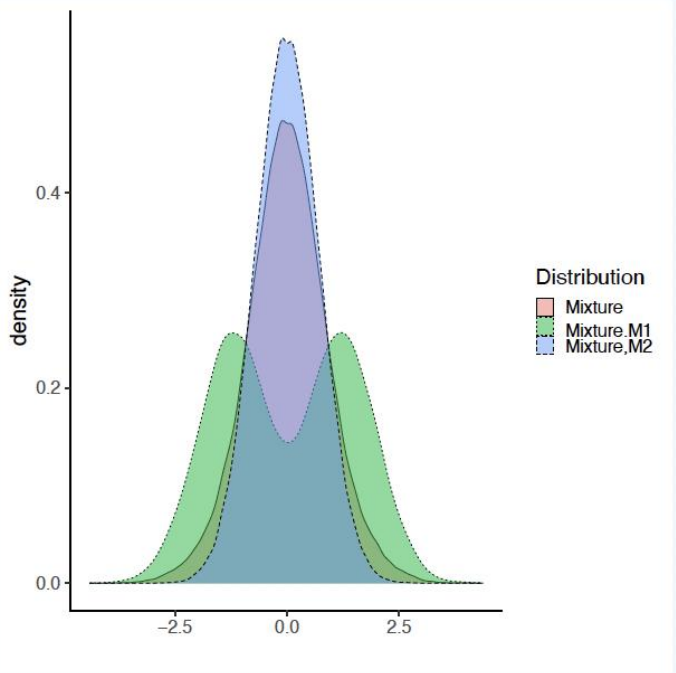
Test-then-Pool

Correct choice: A & D

- Under H_a: **A.** Normal (N)



- Under H_0: **D.** Mixture of N and non-N



NC STATE UNIVERSITY

Test-then-Pool
Asymptotic distribution

Theorem 1. *Assume the regularity conditions hold.*

1. *Under* H_0,

$$n^{1/2}(\widehat{\psi}_{\text{ttp}} - \psi_0) \xrightarrow{\mathscr{D}} \mathscr{M}(\gamma) = \begin{cases} \mathscr{M}_1(\gamma) = V_{\text{rt-eff}}^{1/2} \mathscr{Z}_{c_\gamma}^{\text{t}} - V_{\text{AB}}^{1/2} \mathscr{Z}_2, & w.p.\ \xi \\ \mathscr{M}_2 = -V_{\text{AB}}^{1/2} \mathscr{Z}_2, & w.p.\ 1-\xi \end{cases}$$

with the mixing probability $\xi = \lim_{n\to\infty} \mathbb{P}(T \geq c_\gamma)$.

2. *Under* H_a,

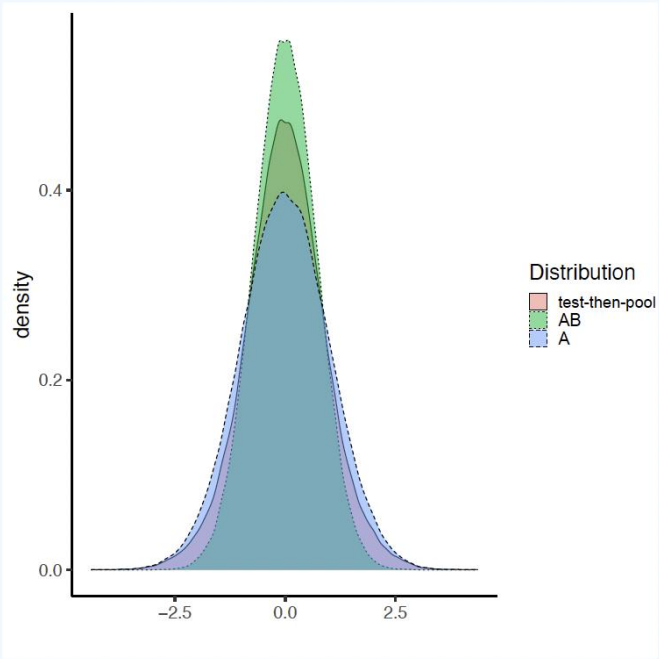
$$n^{1/2}(\widehat{\psi}_{\text{ttp}} - \psi_0) \xrightarrow{\mathscr{D}} \text{Normal}(0, V_{\text{A}}).$$

NC STATE UNIVERSITY

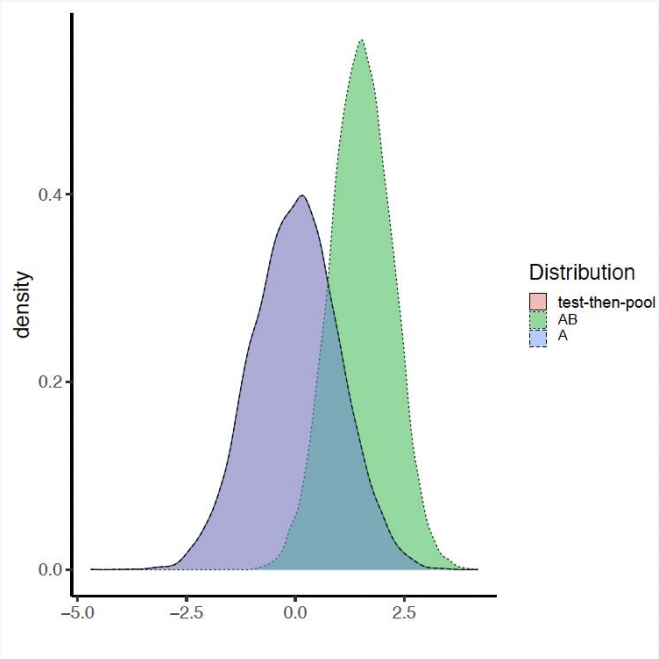
Test-then-Pool

Comparison among $\widehat{\psi}_A$, $\widehat{\psi}_{ttp}$, $\widehat{\psi}_{AB}$

- Under H_0



- Under H_a

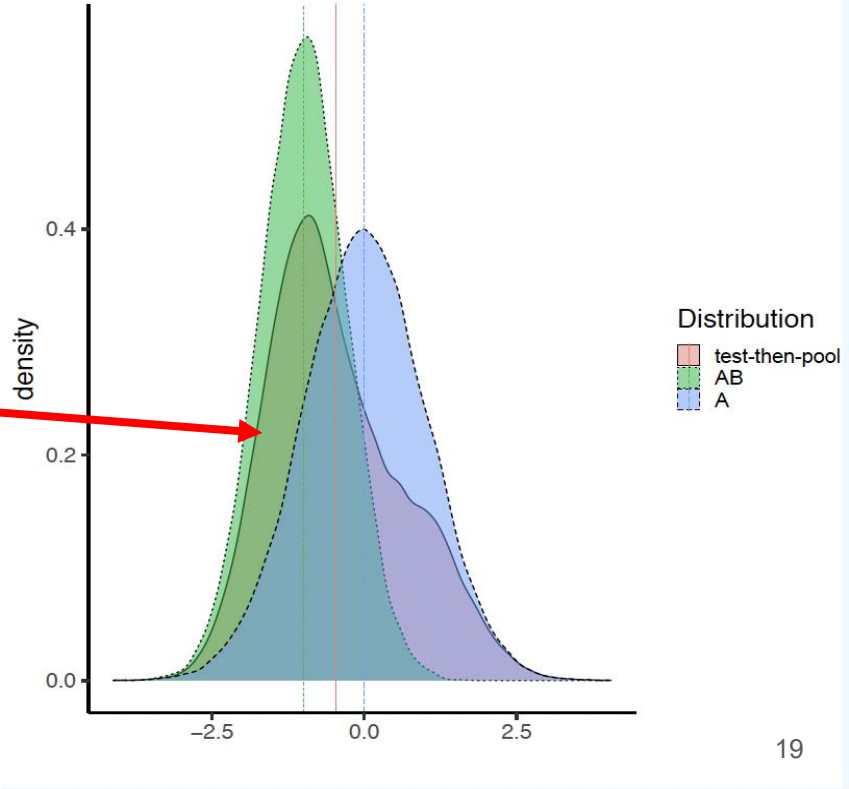


NC STATE UNIVERSITY

Test-then-Pool

Finite-sample behavior of $n^{1/2}(\hat{\psi}_{\mathbf{ttp}} - \psi_0)$

- A toy example with a weak violation of the comparability between Samples A and B
- The finite-sample distribution of test-then-pool can be far from the asymptotic normality.



NC STATE UNIVERSITY

Test-then-Pool

Problems with fixed althernative

- The asymptotic approximation to $n^{1/2}(\widehat{\psi}_{\text{ttp}} - \psi_0)$ is poor as it fails to account for the uncertainty in $I(T < c_\gamma)$
  - No uncertainty about $I(T < c_\gamma)$ as $n \to \infty$; i.e., the power of rejecting H_0 goes to 1
  - Under weak violation (e.g., existance of a weak unmeasured confounder), difficulty in making a definitive decision even with a large sample size
  - Mismatch between the finite-sample behavior and the fixed-parameter asymptotics

- We need an asymptotic framework that retains uncertainty about $I(T < c_\gamma)$ as $n \to \infty$

NC STATE UNIVERSITY

Test-then-Pool
Local alternative

- Idea: consider a sequence of alternative hypotheses, under which the key features of sampling distribution are retained asymptotically

- **Sample comparability may be violated**
  - (Local alternative) $\mathrm{H}_{a,n}$: $\mathbb{E}\{S_{\mathrm{B},\psi_0}(V)\} = n^{-1/2}\eta$, where η is the local parameter

- $n^{-1/2}$ perturbations of H_0
- Commonly used for comparing power of hypothesis tests
- $\mathrm{H}_{a,n}$ is useful to study finite-sample properties (e.g. weak violation)
  - Staiger and Stock (1997) IV regression with weak instruments
- $\mathrm{H}_{a,n}$ compasses H_0 (i.e., $\eta = 0$) and H_a (i.e., $\eta = \pm\infty$)

NC STATE UNIVERSITY

Test-then-Pool

Moving-parameter asymptotics

- $\widehat{\psi}_{\text{ttp}}$ follows a limiting "mixture" distribution

Theorem 2. *Assume that the regularity conditions hold. Under* $\mathrm{H}_{a,n}$,

$$n^{1/2}(\widehat{\psi}_{\text{ttp}} - \psi_0) \xrightarrow{\mathscr{D}} \mathscr{M}(\gamma;\eta) = \begin{cases} \mathscr{M}_1(\gamma;\eta) = V_{\text{rt-eff}}^{1/2}\mathscr{Z}_{c_\gamma}^{\text{t}} - V_{\text{AB}}^{1/2}\mathscr{Z}_2(\eta), & w.p.\ \xi \\ \mathscr{M}_2(\eta) = -V_{\text{AB}}^{1/2}\mathscr{Z}_2(\eta), & w.p.\ 1-\xi \end{cases}$$

with the mixing probability $\xi = \lim_{n\to\infty}\mathbb{P}(T \geq c_\gamma)$.

1. H_a : $\mathscr{M}(\gamma;\eta = \pm\infty)$ reduces to $\text{Normal}(0, V_{\text{A}})$, regular

2. H_0 and $\mathrm{H}_{a,n}$: $\mathscr{M}(\gamma;\eta)$ indexed by η, non-regular

NC STATE UNIVERSITY

Test-then-Pool

Adaptive selection of γ

Insight for selecting γ (recall $c_\gamma = \chi^2_{p,1-\gamma}$)

$$\mathrm{MSE}(\gamma, \eta) = \underbrace{V_{\mathrm{AB}} + V_{\text{rt-eff}} g_1(c_\gamma; \lambda)}_{\text{Term 1}} + \underbrace{(V_{\mathrm{AB}}\eta)^{\otimes 2} g_2(c_\gamma; \lambda)}_{\text{Term 2}}$$

- if η is small, the MSE is dominated by Term 1, can be made small for a small γ (more likely to accept RWD)
- if η is large, the MSE is dominated by Term 2, can be made small for a large γ (more likely to reject RWD)

Adaptive selection

- select γ that minimizes $\mathrm{MSE}(\gamma; \widehat{\eta})$, where $\widehat{\eta} = n^{-1/2} \sum_{i \in \mathscr{B}} \widehat{S}_{\mathrm{B}, \widehat{\psi}_{\mathrm{A}}}(V_i)$

NC STATE UNIVERSITY

Test-then-Pool

Non-regularity and inference

- Limit depends on local parameter η → non-regularity

- We need inference procedures that are valid under local alternatives

NC STATE UNIVERSITY

Our journey for solving this problem

Several failed attempts

NC STATE UNIVERSITY

Q1
Naïve bootstrap?

- Can we use the nonparametric bootstrap? 

- $\widehat{\psi}_{\text{ttp}}$ is a non-smooth test-based estimator due to the indicator of the test

Theorem 3. *The nonparametric bootstrap is inconsistent for* $n^{1/2}(\widehat{\psi}_{\text{ttp}} - \psi_0)$.

NC STATE UNIVERSITY

Q2
m-out-n bootstrap or subsampling?

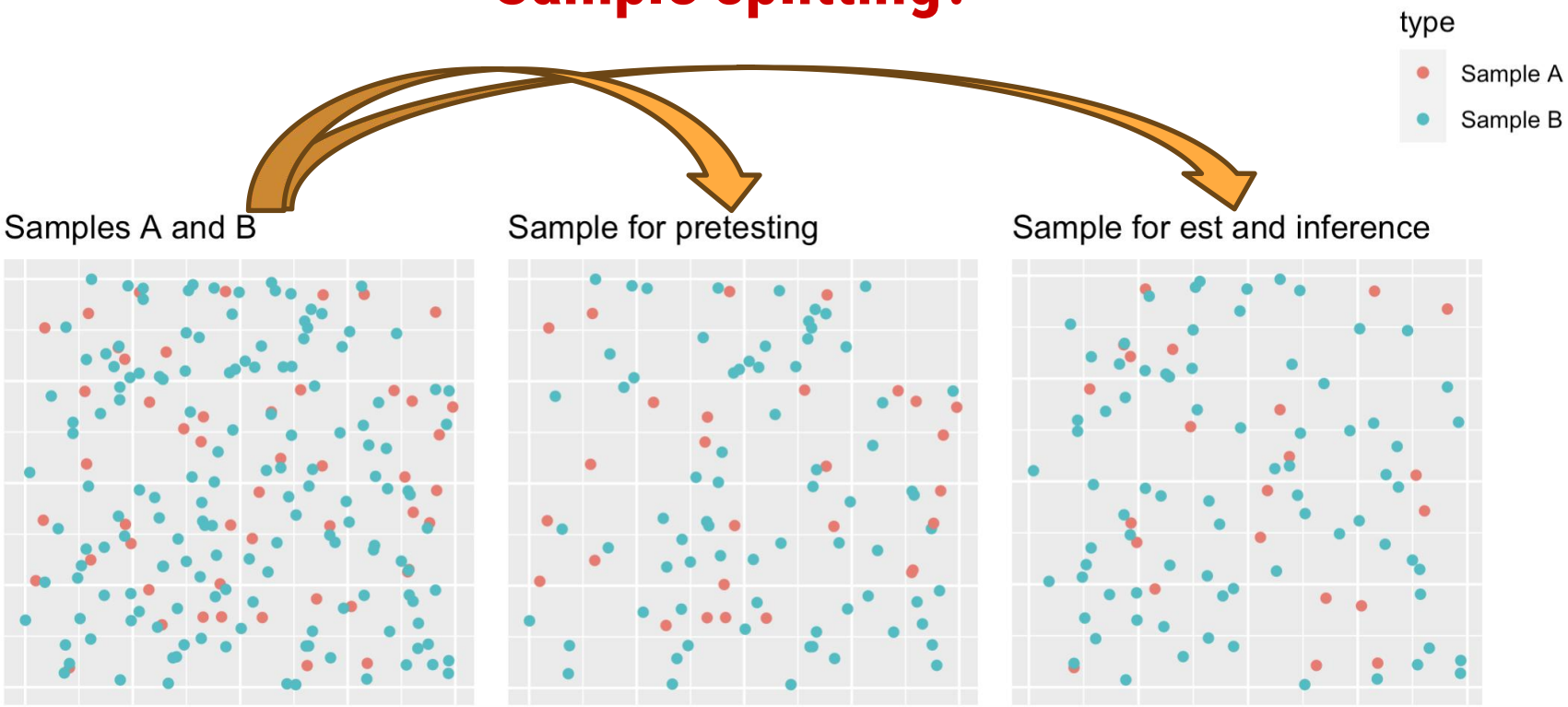
- Yes! There are potential solutions:
  - m-out-of-n bootstrap (Bickel et al 1997)
  - subsampling (Politis et al 1999)

However, these can be difficult to tune and can induce significant finite-sample bias.

NC STATE UNIVERSITY

Q3

Sample splitting?



NC STATE UNIVERSITY

Q3
Sample splitting?

- 

sample splitting cannot solve the problem.
- The pretesting procedure has errors.
- While classical inference (even based on the inference sample) does not take this into account.

		Pre-testing	
	b	$\psi_1 = 1$	$\psi_2 = 1$
Comparable	0	94.0	95.2
	0.11	93.0	95.2
	0.23	91.6	92.2
	0.34	90.6	88.8
Weak violation	0.46	89.4	90.0
	0.57	88.6	89.6
	0.69	90.8	91.2
	0.8	92.8	92.8
	1	93.6	94.2
Strong violation	2	94.4	94.6

NC STATE UNIVERSITY

Q4

Soft thresholding mitigates non-regularity?

- Can we replace $I(T < c_\gamma)$ by a smooth approximation?
- **Example 1.** A smooth weight function $\Phi_\epsilon(c_\gamma - T)$, where $\Phi_\epsilon(z)$ is the normal CDF
- **Example 2.** The p-value from the test $1 - F_{\chi_p^2}(T)$, where $F_{\chi_p^2}(z)$ is the χ_p^2 CDF
- **Example 3.** Bagging: average the bootstrap replications

NC STATE UNIVERSITY

Q4

Soft thresholding mitigates non-regularity?

- Yes. But really, no.
- The bias induced by this smoothing is non-trivial and driving the bias down either inflates the variance or destroys local uniform convergence.

NC STATE UNIVERSITY

Test-then-Pool

An adaptive confidence interval

- Consider $e_k^{\mathrm{T}}\psi_0$ the kth component of ψ_0

Adaptive CI for $n^{1/2}e_k^{\mathrm{T}}(\widehat{\psi}_{\mathrm{ttp}} - \psi_0)$

1. H_a, standard asymptotics, usual CI based on normal quantiles: $[\widehat{Q}_{\alpha/2}, \widehat{Q}_{1-\alpha/2}]$
2. H_0, $\mathrm{H}_{a,n}$, non-standard asymptotics, least favorable CI

- Use T to distinguish between H_a and $\mathrm{H}_0, \mathrm{H}_{a,n}$: generalized moment selection (Andrews and Soares, 2007)

T	Compare T and $\kappa_n = \{\log(n)\}^{1/2}$
H_a: $T \to \infty$	if $T > \kappa_n$ then H_a
$\mathrm{H}_0, \mathrm{H}_{a,n}$: $T = O_{\mathbb{P}}(1)$	if $T \leq \kappa_n$ then $\mathrm{H}_0, \mathrm{H}_{a,n}$

NC STATE UNIVERSITY

Test-then-Pool

An adaptive confidence interval

- Least favorable CI under H_0 and $H_{a,n}$
  - For a fixed η, a $(1-\alpha)$ CI is $[\widehat{Q}_{\alpha/2}(\eta), \widehat{Q}_{1-\alpha/2}(\eta)]$, where $\widehat{Q}_{\alpha}(\eta)$ approximate α-th quantile of $\mathscr{M}(\gamma;\eta)$
  - $\mathscr{B}_{1-\alpha}$: a bounded $1-\alpha$ region of η from Normal$\{n^{-1/2}\sum_{i\in\mathscr{B}}\widehat{S}_{\mathrm{B},\widehat{\psi}_{\mathrm{A}}}(V_i), \widehat{\Sigma}_{SS}\}$
  - $(1-\alpha)$ least favorable CI

$$\left[\inf_{\eta\in\mathscr{B}_{1-\tilde{\alpha}}} \widehat{Q}_{\tilde{\alpha}/2}(\eta), \sup_{\eta\in\mathscr{B}_{1-\tilde{\alpha}}} \widehat{Q}_{1-\tilde{\alpha}/2}(\eta)\right]$$

with $(1-\tilde{\alpha})^2 = 1-\alpha$

NC STATE UNIVERSITY

Test-then-Pool

An adaptive confidence interval

$$\mathrm{ACI}_{1-\alpha} = \begin{cases} \left[\inf_{\eta \in \mathscr{B}_{1-\tilde{\alpha}}} \widehat{Q}_{\tilde{\alpha}/2}(\eta), \sup_{\eta \in \mathscr{B}_{1-\tilde{\alpha}}} \widehat{Q}_{1-\tilde{\alpha}/2}(\eta)\right], & \text{if } T \le \kappa_n, \\ [\widehat{Q}_{\alpha/2}, \widehat{Q}_{1-\alpha/2}], & \text{if } T > \kappa_n. \end{cases}$$

Theorem 4.

$$\lim_{n\to\infty} \mathbb{P}\left\{ n^{1/2} e_k^{\mathrm{T}} (\widehat{\psi}_{\mathrm{ttp}} - \psi_0) \in \mathrm{ACI}_{1-\alpha} \right\} \ge 1 - \alpha,$$

- *the equality holds under* H_a.

NC STATE UNIVERSITY

Simulation – setup in causal inference

Goal: assess the robustness against RWD incomparability and the adaptive inference

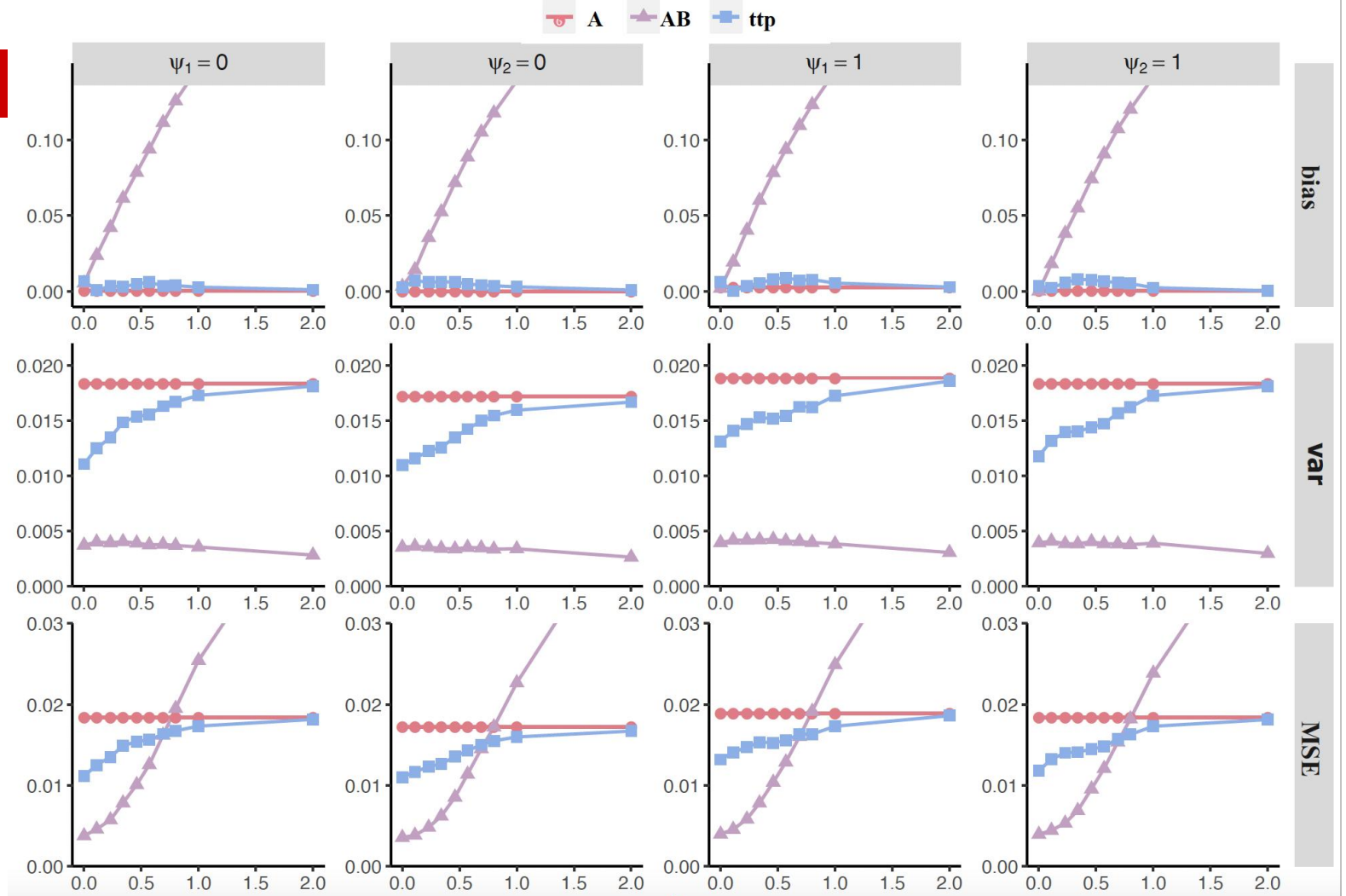
- Observed $X = Z = (1, X_1, X_2)^{\mathrm{T}}$, $X_j \sim \text{Normal}(1,1)$ for $j = 1,2$; Unmeasured $\mathbf{U} \sim \text{Normal}(1,1)$
- Potential outcome $Y(a) \mid X = \mu(X) + a \times \tau(Z) + \varepsilon(a)$,

$$\mu(X) = X_1 + X_2 + \mathbf{U}, \quad \text{HTE: } \tau(Z) = \psi_0 + \psi_1 X_1 + \psi_2 X_2$$

- Parameter of interest
  - $(\psi_1, \psi_2) = (0,0)$: zero effect modification; $(\psi_1, \psi_2) = (1,1)$: non-zero effect modification
- RCT ($n \approx 600$) $A \sim \text{Bernoulli}(0.5)$
- RWD ($m = 2000$) $A \sim \text{Bernoulli}\{e_0(X)\}$, $\text{logit}\{e_0(X)\} = 1 - 2X_1 - b\mathbf{U}$

$$b \in \underbrace{\{0, 0.11, 0.23, 0.34, 0.46, 0.57, 0.69, 0.80, 1, 2\}}_{\text{zero} \; - - - - - - \rightarrow \; \text{weak} \; - - - \rightarrow \; \text{strong violation of UNC}}$$

NC STATE UNIVERSITY



Result: coverage properties

A and ttp have good coverage rates

Case 1: Zero HTE

	RCT (A)		Eff (AB)		ttp		RCT (A)		Eff (AB)		ttp	
	$\psi_1 = 0$	$\psi_2 = 0$	$\psi_1 = 0$	$\psi_2 = 0$	$\psi_1 = 0$	$\psi_2 = 0$	$\psi_1 = 0$	$\psi_2 = 0$	$\psi_1 = 0$	$\psi_2 = 0$	$\psi_1 = 0$	$\psi_2 = 0$
b	Coverage Rate (%)						Width ($\times 10^{-3}$)					
0	94.1	94.1	93.8	93.7	92.7	92.5	528	528	243	242	472	473
0.11	94.1	94.1	92.2	92.7	93.2	92.8	527	528	242	242	488	487
0.34	94.1	94.0	83.2	84.5	94.0	93.8	528	528	241	241	516	516
0.46	94.1	94.0	74.7	76.3	94.5	94.5	528	528	239	240	530	530
0.57	94.1	94.0	66.4	66.1	95.5	95.2	528	528	238	238	535	535
0.69	94.1	94.1	56.1	56.3	95.5	95.8	528	528	235	236	534	534
0.8	94.1	94.0	46.3	46.8	95.5	95.6	528	528	233	234	532	532
1	94.1	94.0	31.5	31.1	95.5	95.0	528	528	229	230	529	529
2	94.1	94.0	2.9	3.6	94.3	94.4	528	528	207	208	527	527

AB cannot control type-I error

ttp is more precise than A

Result: coverage properties

A and ttp have good coverage rates

Case 2: Non-zero HTE

	RCT (A)		Eff (AB)		ttp		RCT (A)		Eff (AB)		ttp	
	$\psi_1 = 1$	$\psi_2 = 1$	$\psi_1 = 1$	$\psi_2 = 1$	$\psi_1 = 1$	$\psi_2 = 1$	$\psi_1 = 1$	$\psi_2 = 1$	$\psi_1 = 1$	$\psi_2 = 1$	$\psi_1 = 1$	$\psi_2 = 1$
b	Coverage Rate (%)						Width ($\times 10^{-3}$)					
0	94.3	93.8	95.0	94.2	92.7	92.5	529	530	243	243	472	474
0.11	94.3	93.8	93.3	92.9	92.9	92.7	529	530	242	243	479	480
0.34	94.3	93.8	84.9	83.5	94.4	93.5	529	530	241	242	511	514
0.46	94.3	93.8	76.8	75.8	94.5	94.4	529	530	240	240	524	526
0.57	94.3	93.8	67.2	66.8	95.5	94.8	529	530	238	239	530	532
0.69	94.3	93.8	56.8	55.9	95.3	94.6	529	530	236	236	529	531
0.8	94.3	93.8	46.5	45.2	95.3	95.0	529	530	233	234	530	532
1	94.3	93.8	30.9	29.4	95.5	94.9	529	530	229	230	530	532
2	94.3	93.8	2.6	3.0	94.7	94.2	529	530	208	209	528	530

ttp is more precise/powerful than A

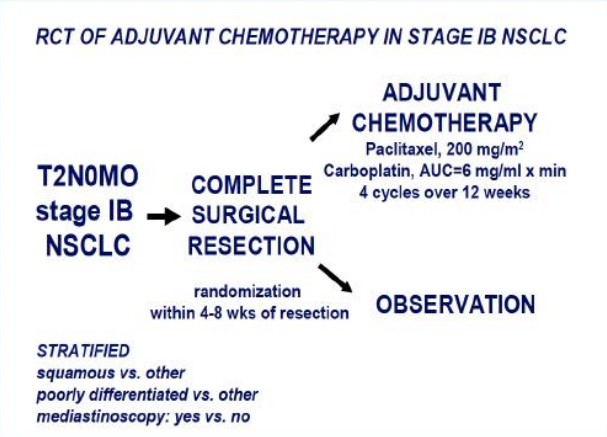
NC STATE UNIVERSITY

CALGB 9633

Adjuvant chemotherapy for early-stage NSCLC

Goal: evaluate the effect of adjuvant chemotherapy for early-stage non-small-cell lung cancer (NSCLC)

- RCT: CALGB 9633 is the only trial designed for early-stage NSCLC[a]
- RCT was undersized to detect clinically meaningful improvements.
- Exploratory analysis: patients with tumor size ≥ 4.0 cm may benefit



[a]

- Strauss, G. M., et al. (2008). Adjuvant paclitaxel plus carboplatin compared with observation in stage IB non-small-cell lung cancer: C9633 with the Cancer and Leukemia Group B, Radiation Therapy Oncology Group, and North Central Cancer Treatment Group Study Groups. *Journal of Clinical Oncology*.
- Strauss, G. M., et al. (2011). Adjuvant chemotherapy (AC) in stage IB non-small cell lung cancer (NSCLC): Long-term follow-up of CALGB 9633. *Journal of Clinical Oncology*.

NC STATE UNIVERSITY

NCDB
National Cancer Database

- NCDB is an oncology outcomes database that collects information on 70% of all new invasive cancer diagnoses in U.S.
- 1,5207 patients diagnosed with NSCLC between years 2004 – 2016 with stage IB disease who first had surgery and then received either adjuvant chemotherapy or on observation (i.e. no chemotherapy).

NC STATE UNIVERSITY

Summary of two data sources

- **Treatment**: On Observation (Obs), Adjuvant chemotherapy (Adj chemo)
- **Covariate**: gender (1 = male, 0 = female), age, indicator for histology (1 = squamous, 0 = non-squamous), tumor size in cm
- **Outcome**: whether died by year 3 after surgery

Covariate means by treatment group in CALGB 9633 and NCDB.

	A	Sample size	Gender	Age	Histology	Tumor size
RCT: CALGB 9633	$A = 1$	156	64.1%	60.57	40.4%	4.62
$n = 319$	$A = 0$	163	63.8%	61.08	39.3%	4.57
RWD: NCDB	$A = 1$	4271	54.2%	**63.93**	**35.5%**	**5.19**
$m = 15207$	$A = 0$	10936	54.8%	**69.42**	**40.4%**	**4.68**

Clinical questions of interest:
How the effect of adjuvant chemotherapy varies over patients with different tumor sizes?

NC STATE UNIVERSITY

Results: HTE

- Causal risk difference:

$$\tau_{\psi_0}(X) = \psi_{0,1} + \psi_{0,2} \times \text{tumor size}^*$$

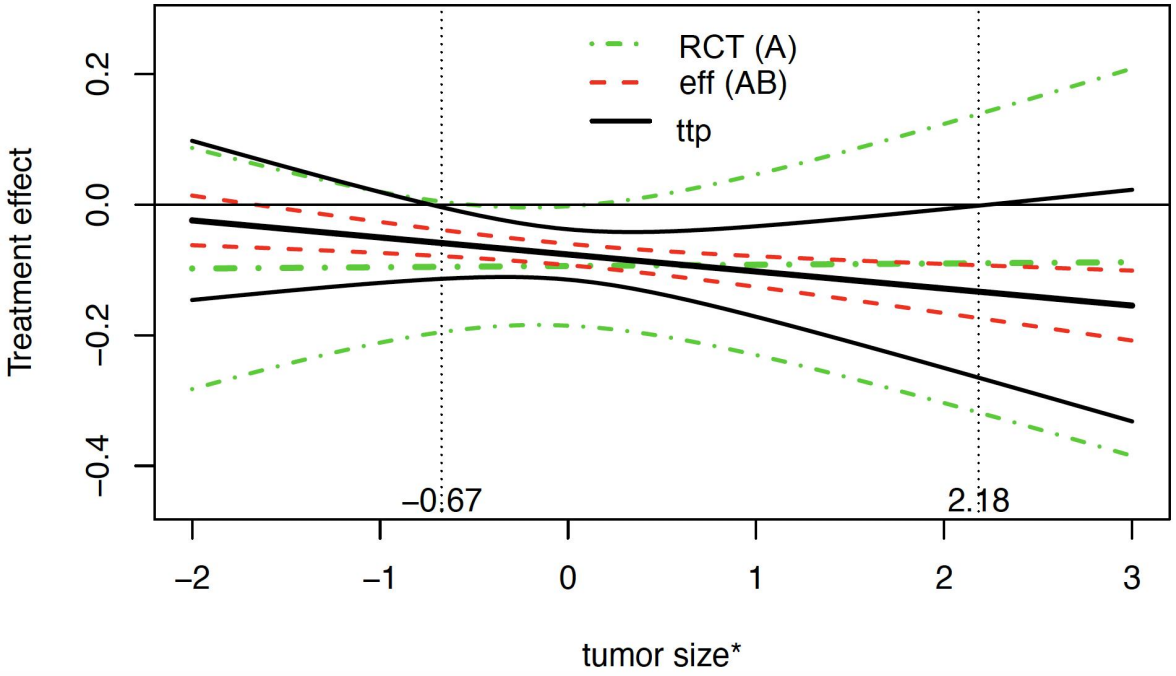
tumor size*= (tumor size−4.82)/1.72

	Intercept ($\psi_{0,1}$)			tumor size* ($\psi_{0,2}$)		
	Est.	S.E.	C.I.	Est.	S.E.	C.I.
RCT	-0.094	0.054	(-0.202, 0.015)	0.002	0.055	(-0.107, 0.111)
Eff	-0.076	0.0083	(-0.093, -0.059)	-0.026	0.009	(-0.043, -0.009)
Test-based	-0.076	0.0196	(-0.115, -0.037)	-0.026	0.029	(-0.084, 0.032)

- $T = 1.9$: no strong evidence of incomparability of the matched NCDB
- $\widehat{\psi}_{\text{tbi}} = \widehat{\psi}_{\text{eff}}$
- SE of $\widehat{\psi}_{\text{tbi}}$ is larger than that of $\widehat{\psi}_{\text{eff}}$ due to the pre-test

NC STATE UNIVERSITY

Results: HTE



Result: Patients with tumor sizes in [3.67;8.57] cm significantly benefit from adj. chomo. in reducing death rates within 3 years after the surgery.

NC STATE UNIVERSITY

Summary

Say “No” to double dipping. Really?

- Pretesting helps borrowing comparable information.

Pretesting causes non-regularity

- Fixed-parameter asymptotics provide poor approximation to the finite sample distribution → standard inference unreliable
- Local-parameter asymptotics faithfully capture non-regularity as sample size grows large → reliable for inference
- Adaptive CIs remain valid under local parameter asymptotics

NC STATE UNIVERSITY

Thank You for Your Attention!

https://shuyang.wordpress.ncsu.edu/

syang24@ncsu.edu