Bayesian Divide-and-Conquer Propensity Score Based Approaches

for Leveraging Real World Data in Single Arm Clinical Trials

Jian Zhu, PhD — Servier Pharmaceuticals

Eric Baron — Univ. of Connecticut

DIA BSWG KOL LECTURE SERIES

MAY 20, 2022

Acknowledgement

Servier and UConn Stats Co-op

SERVIER

UCONN | UNIVERSITY OF CONNECTICUT

Department of Statistics

Jian Zhu

Rui (Sammi) Tang

Eric Baron

Ming-Hui Chen

Baron, E., Zhu, J., Tang, R., & Chen, M. H. (2022). Bayesian divide-and-conquer propensity score based approaches for leveraging real world data in single arm clinical trials. Journal of Biopharmaceutical Statistics, 1-15.

Outline for this talk

- **Background**
  - Real-World Data (RWD)
  - Bayesian borrowing
  - Individual Patient Data (IPD)
- **Framework**
  - Divide:
    - Propensity score stratification
    - Bayesian borrowing within each stratum
  - Conquer: combine inference across strata
  - Illustrative methods (power prior, mixture prior, double hierarchical prior)
- **Simulation**
- **Takeaway messages**

SERVIER

Background

Real World Data

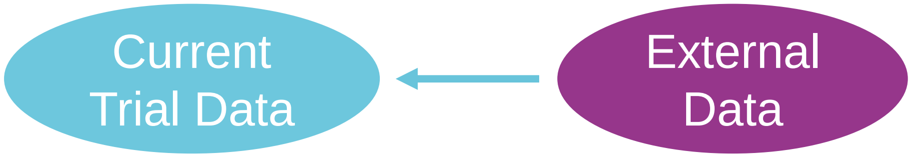
- Using RWD to supplement trial data is particularly relevant in rare diseases

Literature | Regulatory Guidance | Case Studies | ...

- This presentation focuses on leveraging RWD to estimate parameter of interest (e.g. treatment effect) in single arm trials

Bayesian Borrowing

- Naturally used with prior elicitation
- Data inconsistency between sources



Inconsistency

- different study conduct (incl./excl., supportive care …)
- different distribution of baseline prognostic factors (age, ethnicity, BMI, prognostic biomarker …)
- and more

- Common Bayesian methods:
  - power prior [Ibrahim et al., 2015]
  - commensurate power prior [Hobbs et al., 2011]
  - meta-analytic-predictive (MAP) prior [Schmidli et al., 2014]
  - elastic prior [Jiang et al., 2020]

SERVIER

Individual Patient Data

- Patient level baseline characteristics and prognostic factors data
- Some inconsistency can be mitigated by balancing the baseline covariates
  - e.g. Propensity Score (PS) matching, weighting, stratification*
- Separate inconsistency into two parts

Overall Inconsistency (RWD vs Trial) = Inconsistency due to **imbalanced covariates** + Inconsistency due to other factors (incl. unmeasured)

PS stratification

Bayesian Borrowing

* Focus in this talk

Existing methods

- Propensity Score Integrated Methods:
  - PS power prior [Wang et al., 2019]
  - PS MAP prior (multiple external data sources) [Liu et al., 2021]

- We focus on one external data source and explore a general framework

Wang, C., Li, H., Chen, W.-C., Lu, N., Tiwari, R., Xu, Y., and Yue, L. Q. 2019. Propensity score-integrated power prior approach for incorporating real-world evidence in single-arm clinical studies. Journal of biopharmaceutical statistics, 29(5):731–748.

Liu, M., Bunn, V., Hupf, B., Lin, J., and Lin, J. 2021. Propensity-score-based meta-analytic predictive prior for incorporating real-world and historical data. Statistics in medicine.

Framework

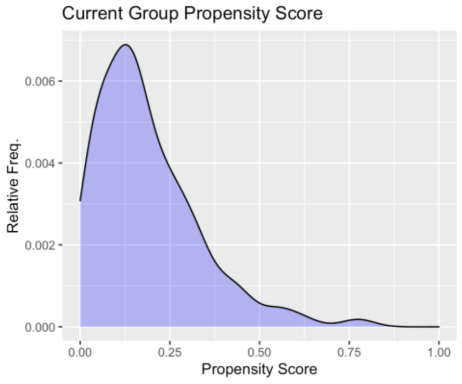
Our Objectives

- Explore generalized framework for **normal endpoint**
- Propose three additional propensity score stratified methods to compare with their non-stratified counterparts
  - Extension of the PS power prior approach
  - Mixture prior approach (MP)
  - Double hierarchical prior approach (HB)
- Properly consider nuisance parameters

Framework (Divide): PS-stratification

Current (target) Trial

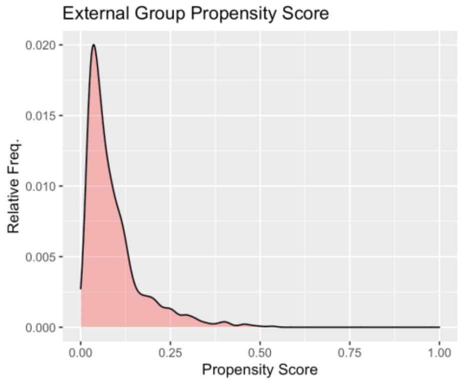
Current Group



PS(Xc)=Pr(included in trial | Xc)

RWD

External Group

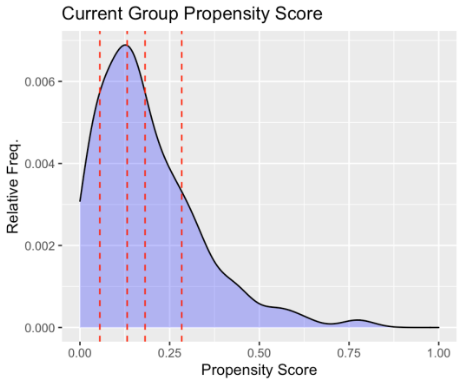


PS(Xe)=Pr(included in trial | Xe)

Framework (Divide): PS-stratification

Current (target) Trial

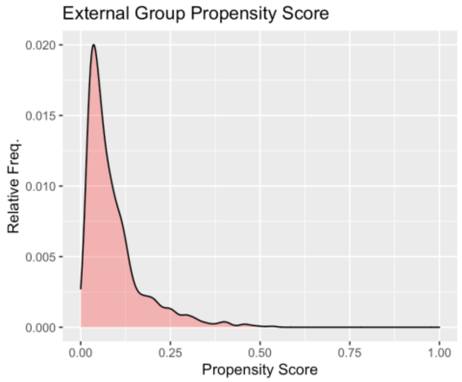
Current Group



Create K strata w/ thresholds based on quantiles of trial PS (e.g. K=5)

RWD

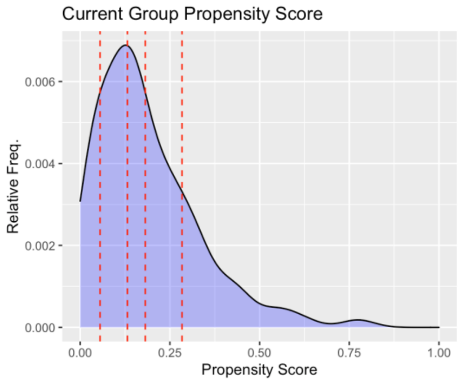
External Group



Framework (Divide): PS-stratification

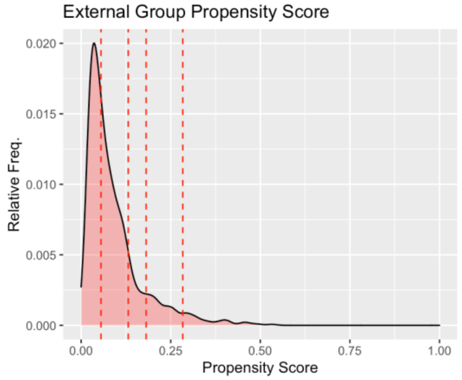
Current (target) Trial

Current Group



RWD

External Group

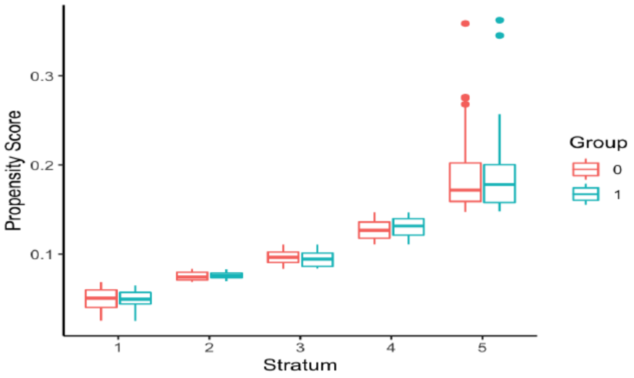


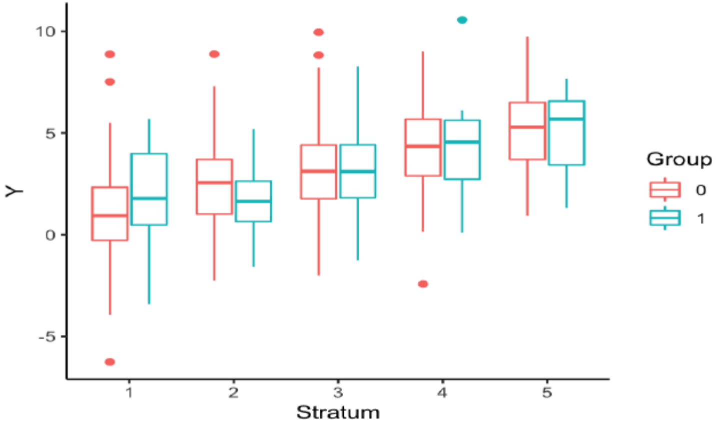
Allocate external pts into corresponding strata

What do we expect to see after stratification

- Balance the prognostic factors (at least for those included in the PS model)

- What about PS distribution and outcome distribution?

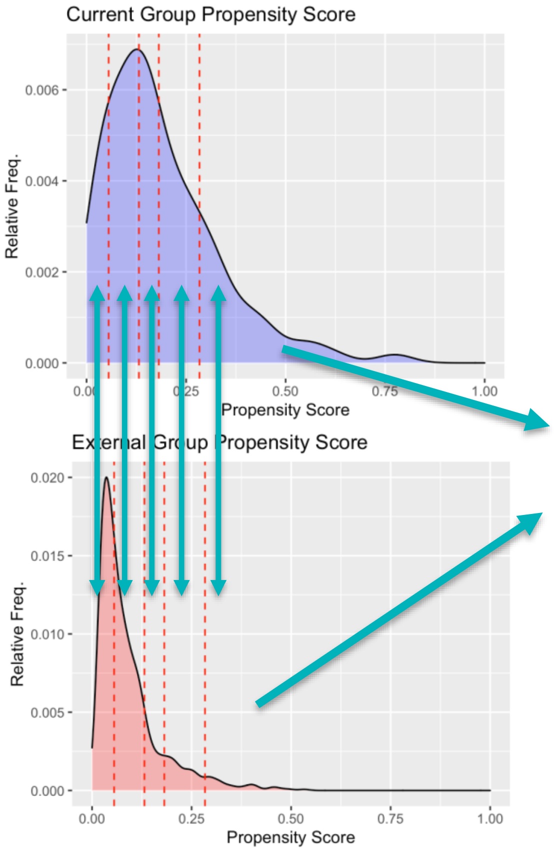




Framework (Divide): Bayesian borrowing within each stratum

Current Group

External Group



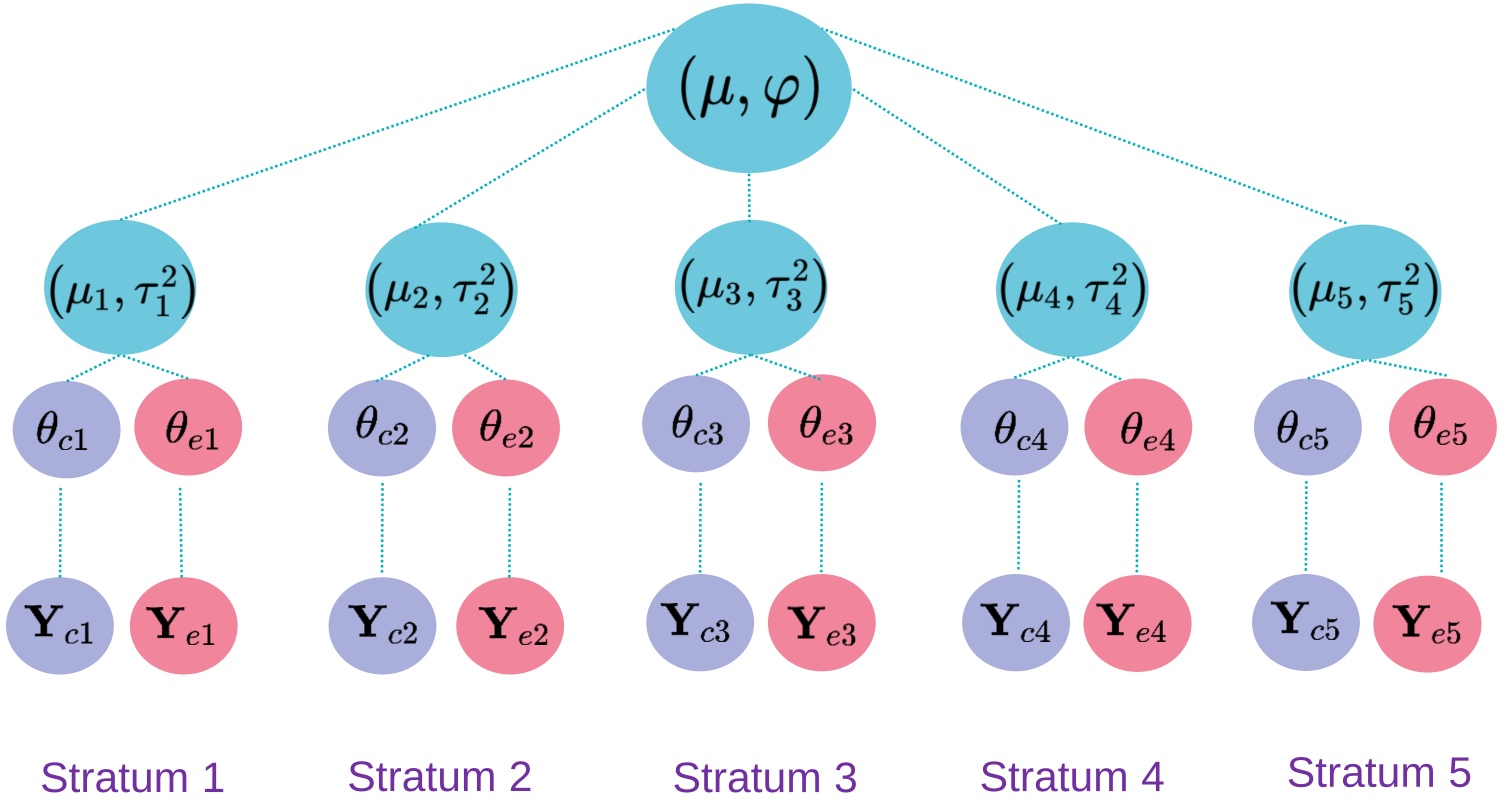
Within each stratum, apply prior to estimate stratum-specific parameter of interest, accounting for heterogeneity among strata

Goal: Estimate overall parameter of interest for the target trial

Remarks

- Trimming
  - External group subjects are omitted if their PS is outside the range of the PS of the current group
- Not guaranteed that stratum-specific sample size for each group is large enough
  - In this case, will not leverage any information from RWD and will use non-informative prior

Bayesian Borrowing Option 1: Double Hierarchical



Double Hierarchical Approach

- Assume that $n_{ek} \geq 2$ and $n_{ck} \geq 2$ for $k = 1, \ldots, K$
- First stage in the kth stratum

$$\theta_{ck} \sim N\left(\mu_k, \tau_k^2\right) \text{ and } \theta_{ek} \sim N\left(\mu_k, \tau_k^2\right).$$

- Second stage in the kth stratum

$$\mu_k \sim N(\mu, \varphi^2) \text{ and } \tau_k^2 \sim \text{TN}\Big(b_{01}, b_{02}, b_{03}, b_{04}\Big),$$

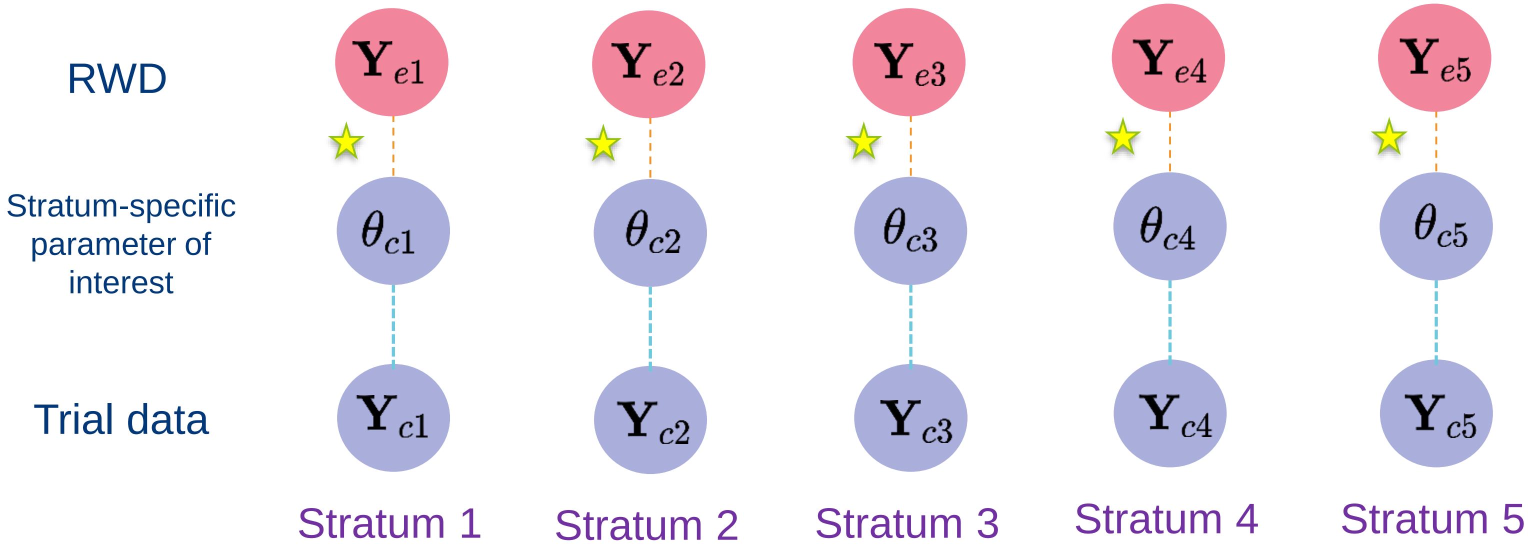
- We further assume

$$\mu \sim N(0, \kappa_0 \varphi^2),\ \frac{1}{\varphi^2} \sim \text{Gamma}\Big(b_{05}, b_{06}\Big),$$

- $\sigma_{ck}^2 \sim IG(0.01, 0.01)$ and $\sigma_{ek}^2 \sim IG(0.01, 0.01)$ for $k = 1, \ldots, K$.

Bayesian Borrowing Option 2: Mixture Prior

$$\pi(\theta_{ck}|\overline{Y}_{ek},\sigma^2_{ek},\gamma_k) = (1-\gamma_k)\pi_0(\theta_{ck}|\sigma^2_{ek}) + \gamma_k\pi(\theta_{ck}|\overline{Y}_{ek},\sigma^2_{ek})$$



Mixture Prior (MP)

- To construct the mixture prior, [Yuan et al., 2021] assume $\theta_{ek} = \theta_{ck}$.
- Then use $\overline{Y}_{ek}$ to construct the MP for θ_{ck} as follows

$$\pi(\theta_{ck}|\overline{Y}_{ek}, \sigma^2_{ek}, \gamma_k) = (1 - \gamma_k)\pi_0(\theta_{ck}|\sigma^2_{ek}) + \gamma_k \pi(\theta_{ck}|\overline{Y}_{ek}, \sigma^2_{ek}),$$

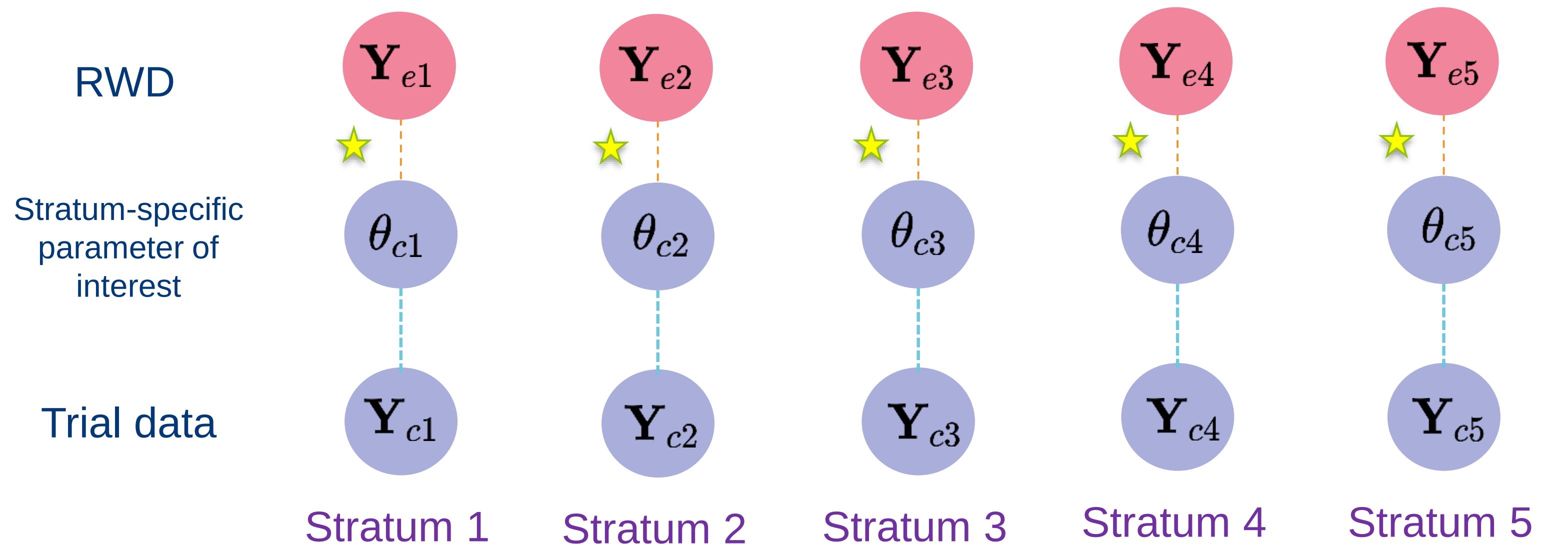
- $0 < \gamma_k < 1$ is the weight, $\pi_0(\theta_{ck}|\sigma^2_{ek})$ is the density of $N(0, 100 \cdot \sigma^2_{ek})$
- Mixing parameter for $k = 1, \ldots K$

$$\gamma_k = \frac{n_{ck}}{2 \cdot n_{ek}}$$

- $\sigma^2_{ck} \sim IG(0.01, 0.01)$ and $\sigma^2_{ek} \sim IG(0.01, 0.01)$ for $k = 1, \ldots, K$.

Bayesian Borrowing Option 3: Power Prior

$$\pi(\theta_{ck}|\overline{Y}_{ek}, \sigma^2_{ek}, \alpha_k) \propto f(\overline{Y}_{ek}|\theta_{ck}, \sigma^2_{ek})^{\alpha_k}\pi_0(\theta_{ck}|\sigma^2_{ek})$$



Power Prior

- Assume $\theta_{ek} = \theta_{ck}$
- Power Prior

$$\pi(\theta_{ck}|\overline{Y}_{ek}, \sigma^2_{ek}, \alpha_k) \propto f(\overline{Y}_{ek}|\theta_{ck}, \sigma^2_{ek})^{\alpha_k}\pi_0(\theta_{ck}|\sigma^2_{ek}),$$

  where $0 \le \alpha_k \le 1$ and $\pi_0(\theta_{ck}|\sigma^2_{ek})$ is the density of $N(0, 100 \cdot \sigma^2_{ek})$
- Discounting Parameter: [Chen and Ibrahim, 2006, Jiang et al., 2020]

$$\alpha_k = \frac{1}{\frac{2\varphi_{0k} n_{ek}}{S^2_{ek}} + 1},$$

  where $\varphi_{0k} = \max\{(\overline{Y}_{ck} - \overline{Y}_{ek})^2, 0.10 \cdot S^2_{ek}\}$ to avoid over-borrowing of the external data.
- $\sigma^2_{ck} \sim IG(0.01, 0.01)$ and $\sigma^2_{ek} \sim IG(0.01, 0.01)$ for $k = 1, \ldots, K$.

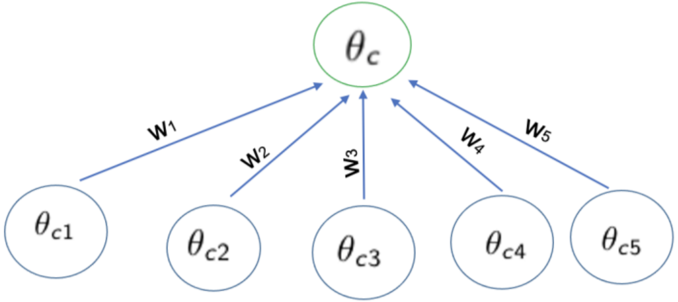
Framework (Conquer):
Combining inference from Different Strata

To estimate θ_c, we combine $\theta_{c1}, \ldots, \theta_{cK}$

Use draws from the posterior distribution of θ_c made using the draws from the posterior distributions of $\theta_{c1}, \ldots, \theta_{cK}$

$$\theta_c = \sum_{k=1}^{K} w_k \theta_{ck},$$

$\boldsymbol{w} = (w_1, \ldots, w_K)'$ are the weights such that $\sum_{k=1}^{K} w_k = 1$



Simulation

Simulation Objectives

- Examine the performance of the proposed methods when there is imbalance in the current and external groups
  - Difference in means of baseline characteristics
- Examine effect of sample size on performance
- Pairwise comparison of stratified and non-stratified approaches
- Investigate advantages of using different combining weights

Simulation Setting

- Simulate covariates for both the current group and external group using a multivariate normal distribution with dimension d,

$$\boldsymbol{X}_{ci} \sim N_d(\mathcal{M}_c, \Sigma_c), \; \boldsymbol{X}_{ej} \sim N_d(\mathcal{M}_e, \Sigma_e)$$

  for $i = 1, \ldots, n_c$ and $j = 1, \ldots, n_e$.

- $\Sigma_c = \Sigma_e = \boldsymbol{I}_d$
- Generate outcome

$$\boldsymbol{Y}_{ci} | X_{ci} = \beta_0 + \boldsymbol{\beta}' \boldsymbol{X}_{ci} + \epsilon_i \text{ and } \boldsymbol{Y}_{ej} | X_{ej} = \beta_0 + \boldsymbol{\beta}' \boldsymbol{X}_{ej} + \epsilon_j,$$

  assuming $\epsilon_i \overset{iid}{\sim} N(0, \eta_c^2)$ and $\epsilon_j \overset{iid}{\sim} N(0, \eta_e^2)$ for $i = 1, \ldots, n_c$ and $j = 1, \ldots, n_e$

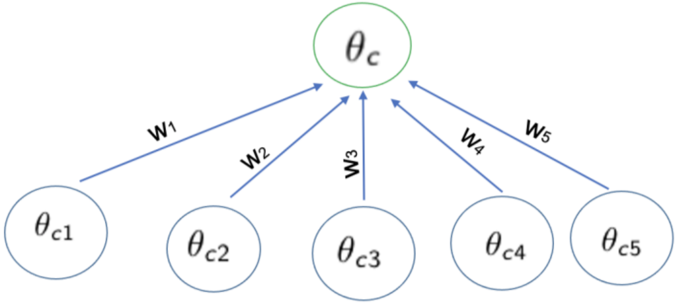
Selected Simulation Scenarios

- Let $d = 3$, $K = 5$, $\beta_0 = 0$, $\boldsymbol{\beta}' = \mathbf{1}_3$, $\mathcal{M}_e = (0.5, 1, 1)'$, and $\mathcal{M}_c = (1, 1.2, 1.25)'$.
- Cutoffs: $p = (p_1, p_2, p_3, p_4) = (0.2, 0.4, 0.6, 0.8)$ and $p_5 = 1$

	Scenario	θ_{true}	n_c	n_e^*	n_e	η_e^2	η_c^2
Imbalance with **small** variance	1	3.45	100	1000	960	1	1
Imbalance with large variance	2	3.45	100	1000	960	3	3
Imbalance with double sample	3	3.45	200	2000	1957	1	1

Table 1: Simulation Scenarios presented with θ_{true} calculated as the mean response in the current group and n_e as the mean number of observations in the external group post-trimming averaged over 1000 simulated data sets.

Exploring different weight



Rely more on middle strata

Weight Scenario	$\boldsymbol{w}$
1	(0.20, 0.20, 0.20, 0.20, 0.20)
2	(0.10, 0.20, 0.40, 0.20, 0.10)
3	(0.05, 0.15, 0.60, 0.15, 0.05)
4	(0.05, 0.10, 0.70, 0.10, 0.05)
5	(0.00, 0.10, 0.80, 0.10, 0.00)

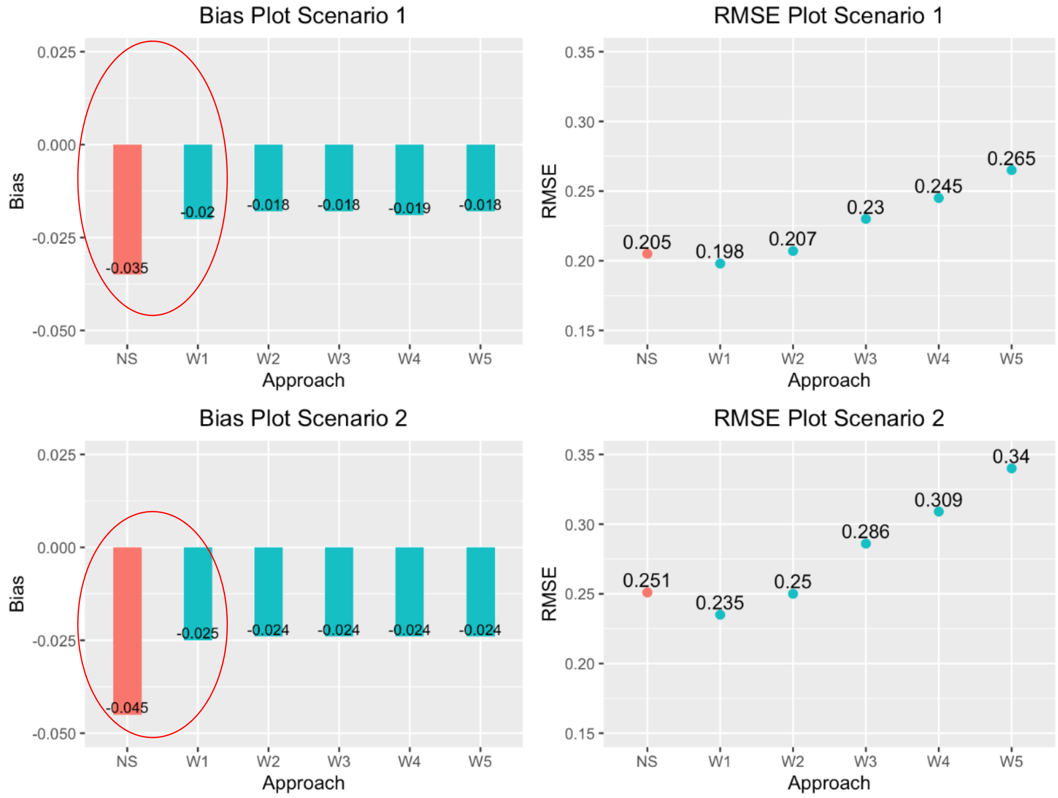
Table 2: Weighting scenarios for conquering weights.

Model Evaluation

- Use 1000 replications of the data for each scenario
- For each replicate, generate a MCMC sample of 10000 iterations with first 1000 discarded
- Bias$=\sum_{r=1}^{R}(\hat{\theta}_r - \theta_r)/R$
- Root mean squared error(RMSE):$= \sqrt{\sum_{r=1}^{R}(\hat{\theta}_r - \theta_r)^2/R}$, where $\hat{\theta}_r$ denotes the posterior mean of the rth replicated data set, and θ_r denotes the true parameter in the current group in the rth replicate with $r = 1, \ldots, R$ and $R = 1000$.
- Coverage probability (CP): the number of 95% HPD intervals of θ_r that contain the true parameter and divide by the total number of replicates.
- Sample standard deviation of the posterior mean over the replicates as $SD = \sum_{r=1}^{R}\sqrt{\sum_{m=1}^{M}(\theta_r^{[m]} - \hat{\theta}_r)^2/(M-1)}/R$,
- $\theta_r^{[m]}$ denotes the mth iteration of the MCMC sample for the parameter θ_r in the rth replicate and $m = 1, \ldots, M$.
- Simulation standard error: $\mathrm{SE} = \sqrt{\sum_{r=1}^{R}(\hat{\theta}_r - \frac{1}{R}\sum_{\ell=1}^{R}\hat{\theta}_\ell)^2/(R-1)}$

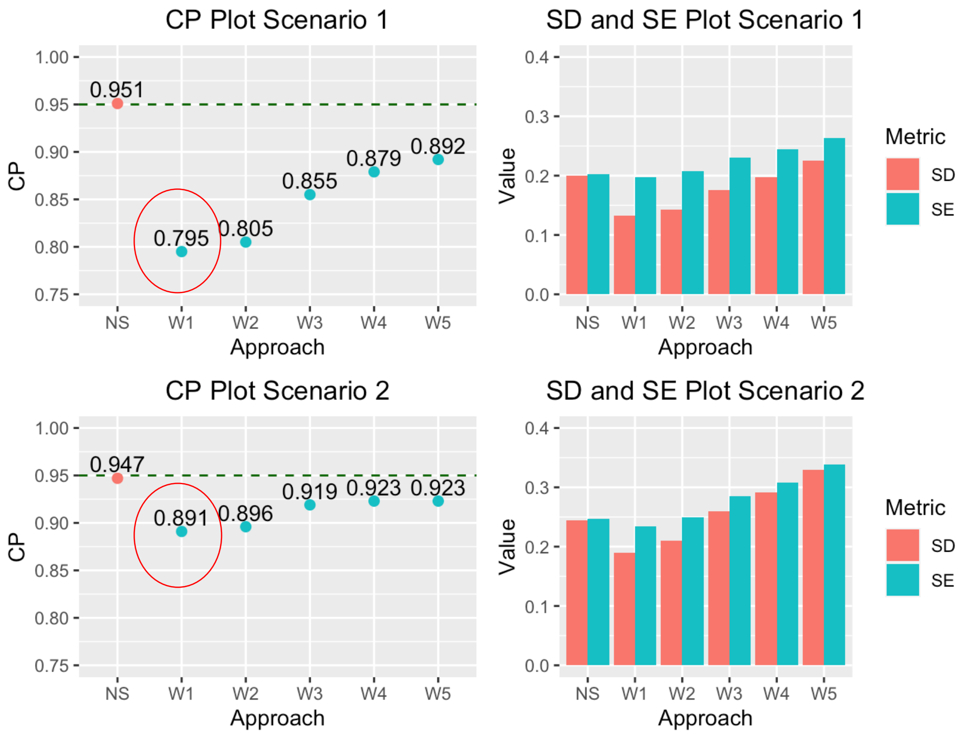
Hierarchical Approach: Scenario 1 and 2 (Bias and RMSE)

	Scenario
Imbalance with small variance	1
Imbalance with large variance	2

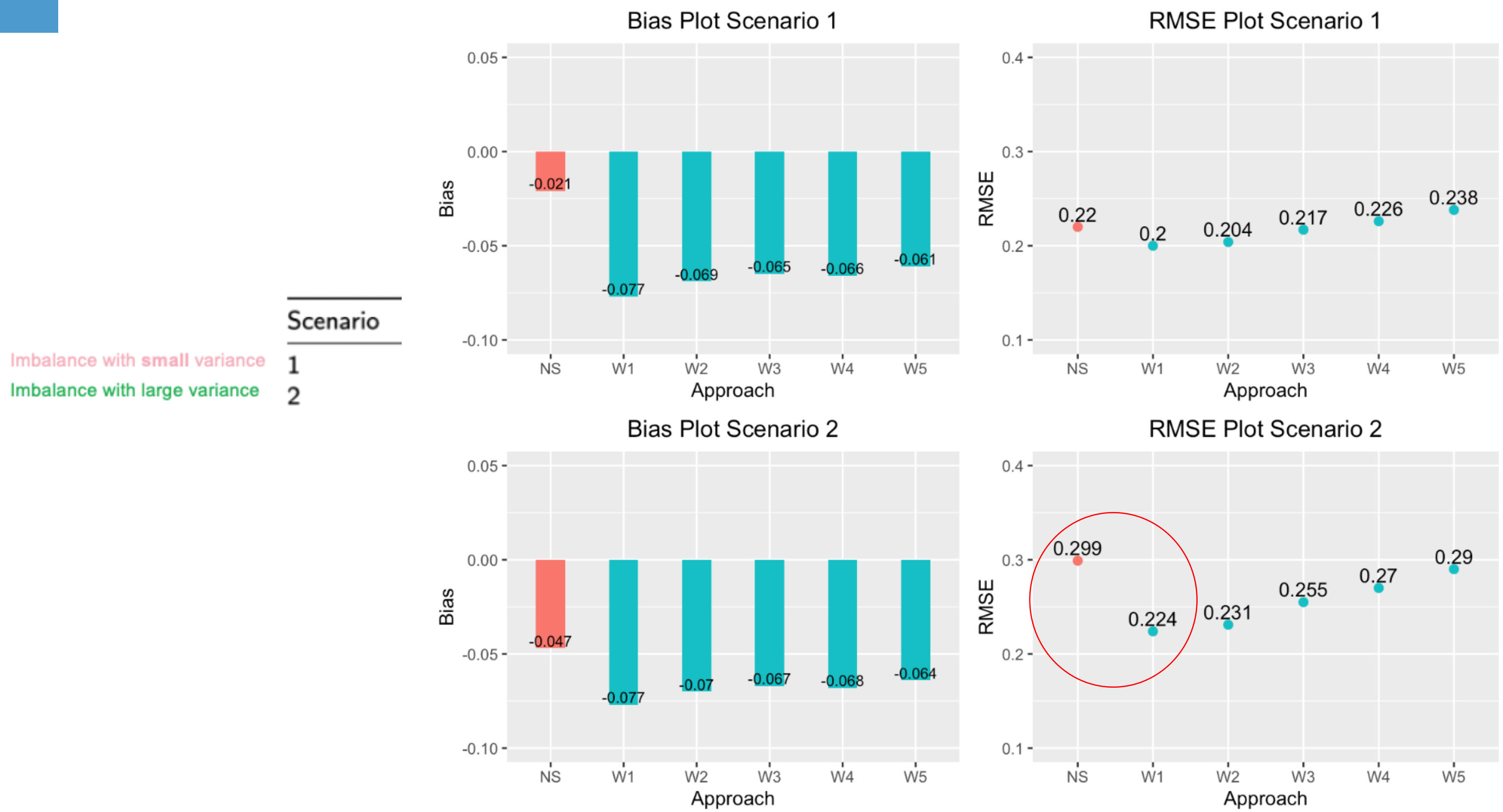


Hierarchical Approach: Scenario 1 and 2 (CP, SD, SE)

	Scenario
Imbalance with small variance	1
Imbalance with large variance	2

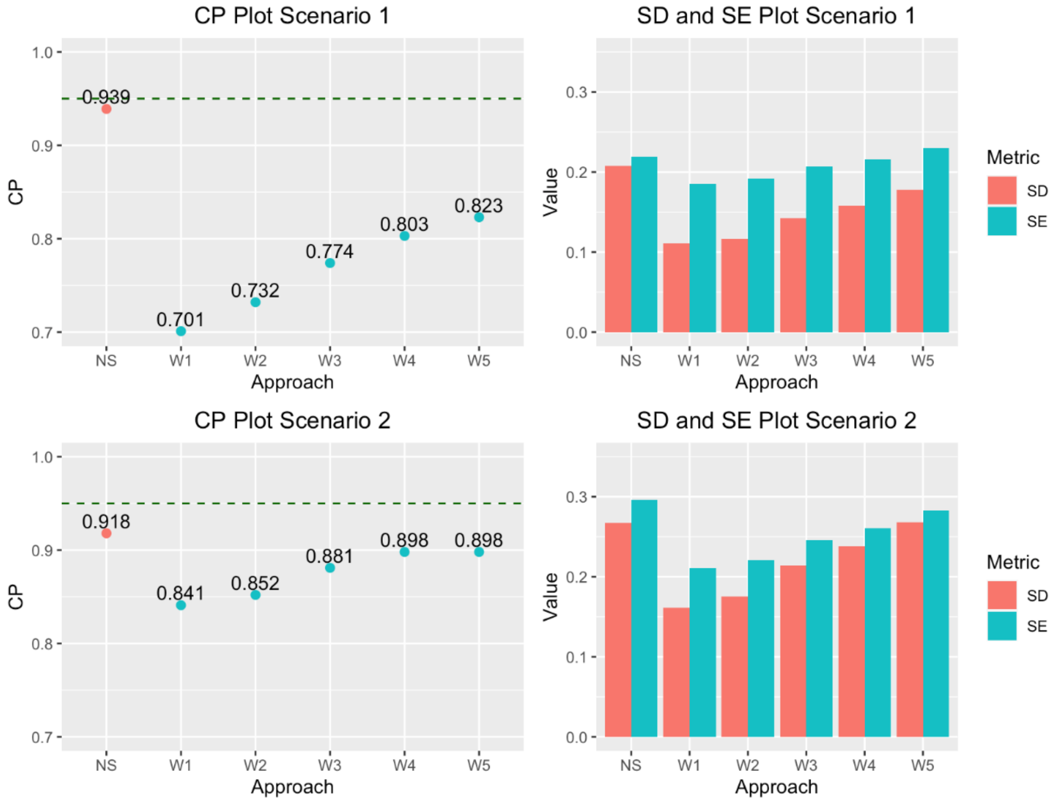


Mixture Approach: Scenario 1 and 2 (Bias and RMSE)



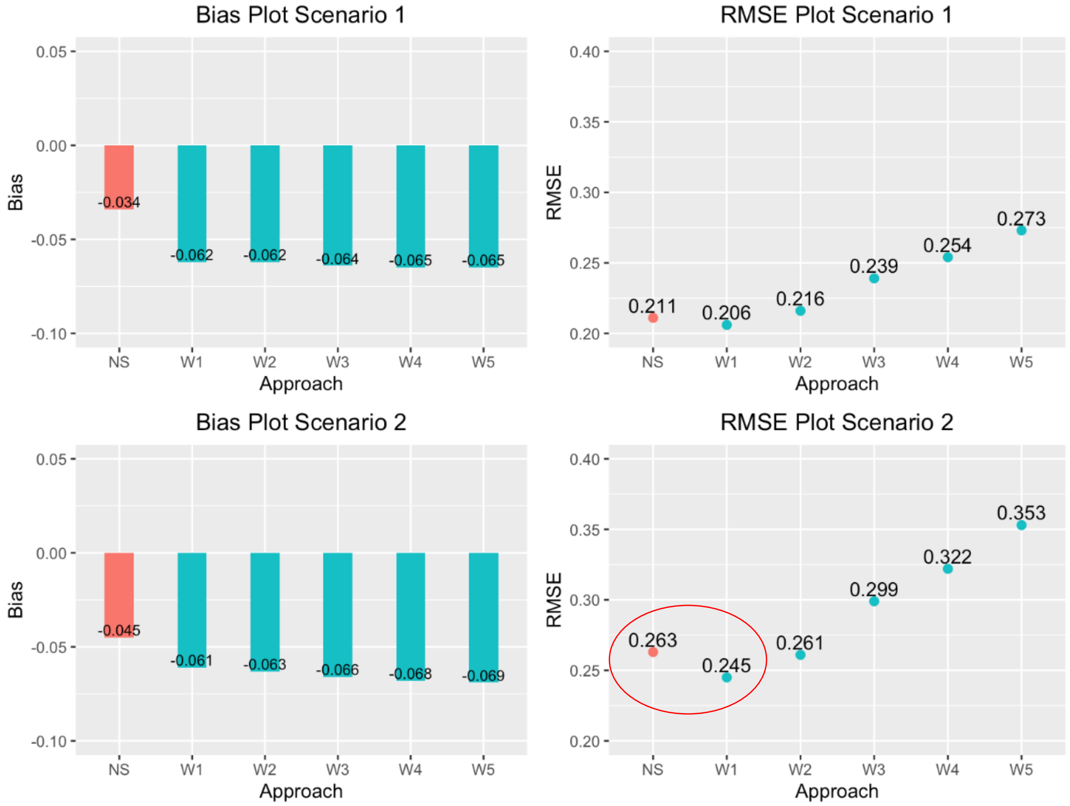
Mixture Approach: Scenario 1 and 2 (CP, SD, SE)

	Scenario
Imbalance with **small** variance	1
Imbalance with large variance	2



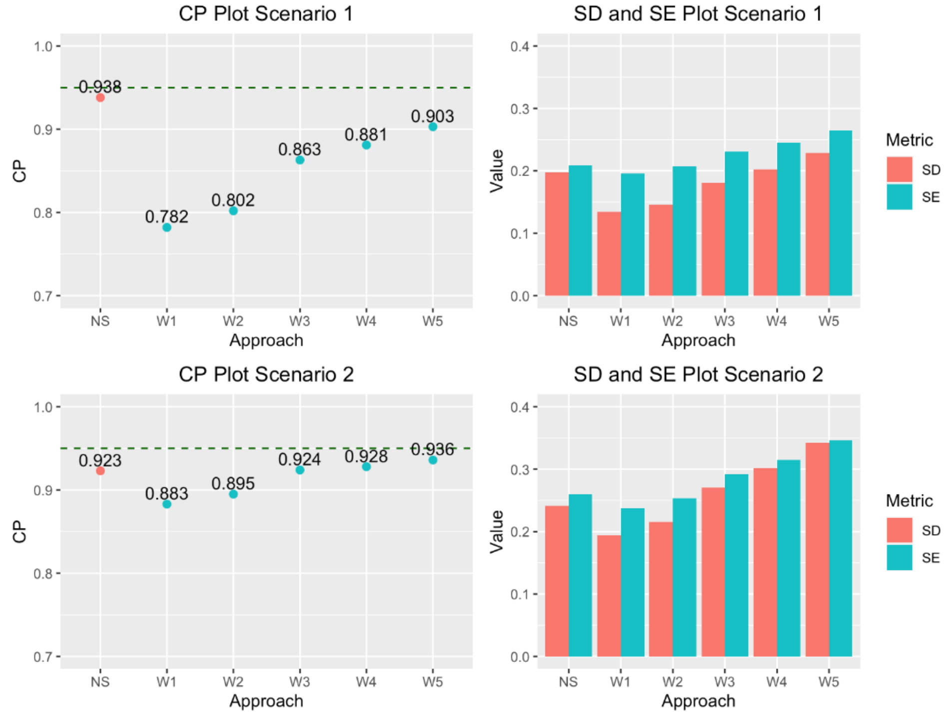
Power Prior Approach: Scenario 1 and 2 (Bias and RMSE)

Scenario	
Imbalance with **small** variance	1
Imbalance with large variance	2

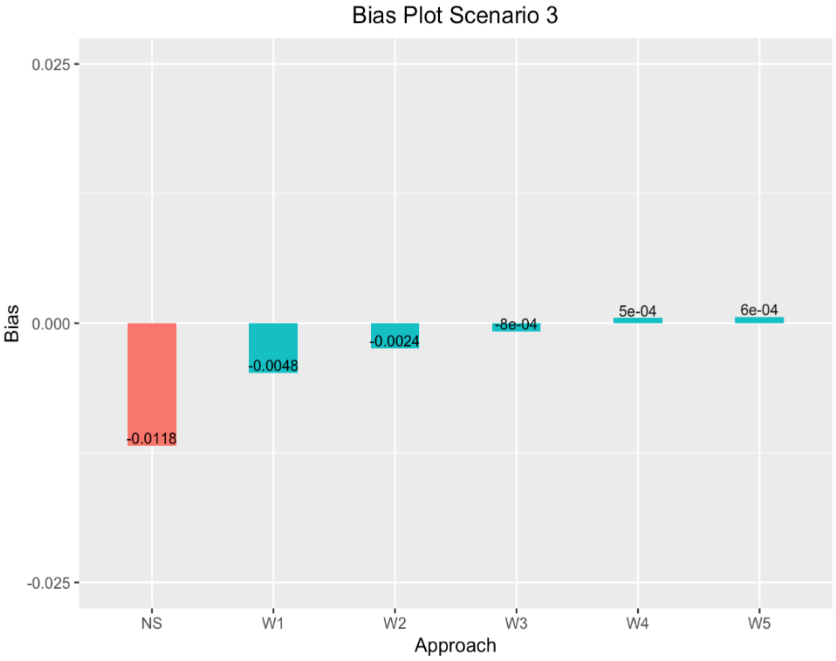


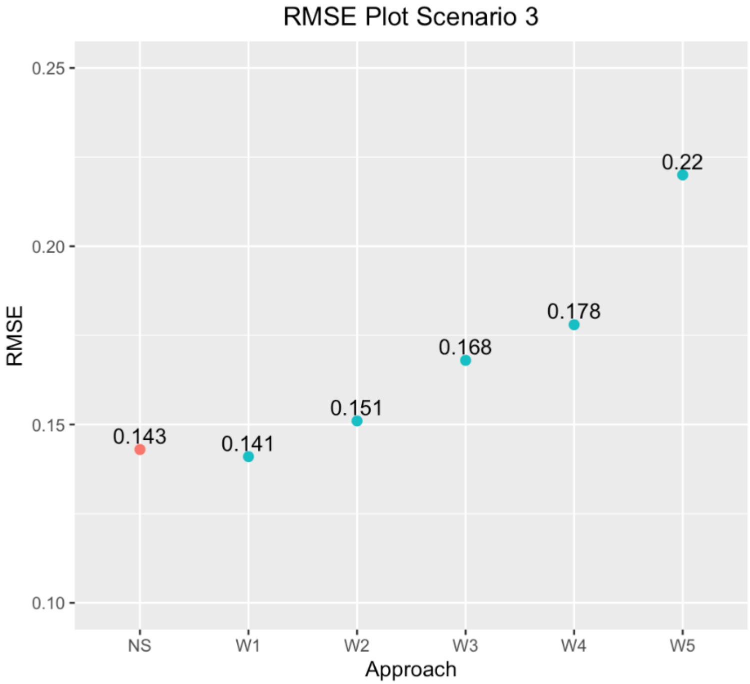
Power Prior Approach: Scenario 1 and 2 (CP, SD, SE)

	Scenario
Imbalance with **small** variance	1
Imbalance with large variance	2

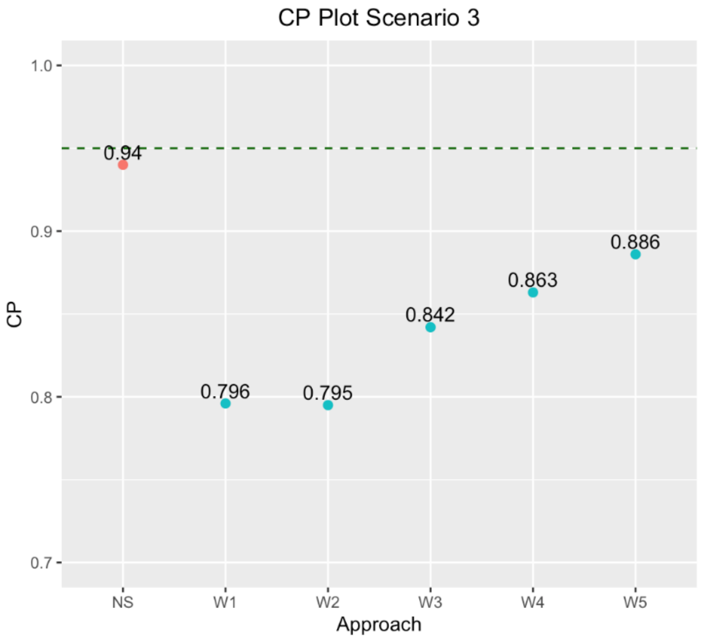


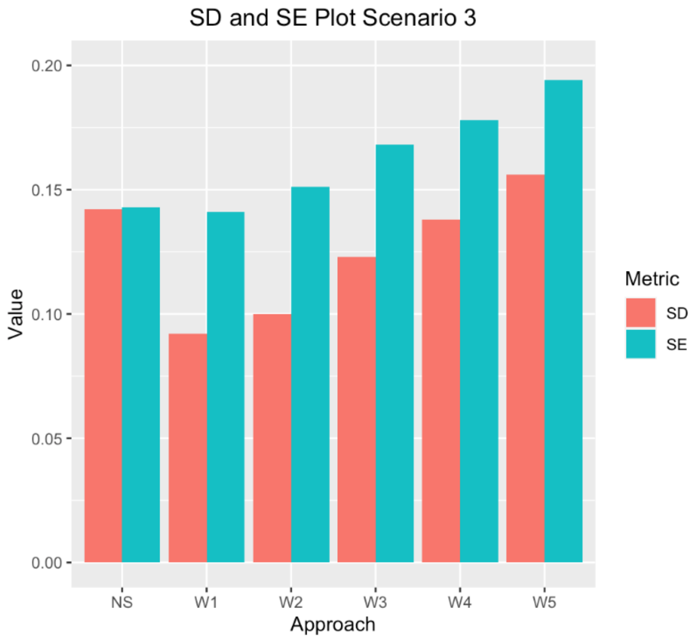
Hierarchical Approach: Scenario 3 (Double Sample Size)





Hierarchical Approach: Scenario 3 (Double Sample Size)





Next Steps

- Explore additional scenarios
  - Time trend effect
  - Misspecification of propensity score model
- Determine appropriate variance estimation when combining stratum-specific estimates
- Extend to design setting
- Generalize to randomized control trial and compare with existing methods

Takeaway Messages

"Divide and conquer" allows more intuitive handling of inconsistency

Certain improvement has been observed

Additional research is needed for further improvement

Thank You!

Eric Baron (eric.baron@uconn.edu)
Jian Zhu (jian.zhu@servier.com)