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

Jian Zhu
Rui (Sammi) Tang

Eric Baron
Ming-Hui Chen

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**
Background
Real World Data

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

- 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

Current Trial Data \[\text{\rightarrow}\] External Data

Common Bayesian methods:
- power prior \cite{Ibrahim2015}
- commensurate power prior \cite{Hobbs2011}
- meta-analytic-predictive (MAP) prior \cite{Schmidli2014}
- elastic prior \cite{Jiang2020}

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

\[
\text{Overall Inconsistency (RWD vs Trial)} = \text{Inconsistency due to imbalanced covariates} + \text{Inconsistency due to other factors (incl. unmeasured)}
\]

- 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


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)

External Group

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

RWD
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

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

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

\[(\mu_1, \tau_1^2) \quad (\mu_2, \tau_2^2) \quad (\mu_3, \tau_3^2) \quad (\mu_4, \tau_4^2) \quad (\mu_5, \tau_5^2)\]

\[
\begin{align*}
(\mu_1, \tau_1^2) & \quad (\mu_2, \tau_2^2) & \quad (\mu_3, \tau_3^2) & \quad (\mu_4, \tau_4^2) & \quad (\mu_5, \tau_5^2) \\
\theta_{c1} & \quad \theta_{e1} & \quad \theta_{c2} & \quad \theta_{e2} & \quad \theta_{c3} & \quad \theta_{e3} & \quad \theta_{c4} & \quad \theta_{e4} & \quad \theta_{c5} & \quad \theta_{e5} \\
Y_{c1} & \quad Y_{e1} & \quad Y_{c2} & \quad Y_{e2} & \quad Y_{c3} & \quad Y_{e3} & \quad Y_{c4} & \quad Y_{e4} & \quad Y_{c5} & \quad Y_{e5} \\
\text{Stratum 1} & \quad \text{Stratum 2} & \quad \text{Stratum 3} & \quad \text{Stratum 4} & \quad \text{Stratum 5}
\end{align*}
\]
Double Hierarchical Approach

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

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

- We further assume

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

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

\[ \pi(\theta_{ck} | \bar{Y}_{ek}, \sigma_{ek}^2, \gamma_k) = (1 - \gamma_k)\pi_0(\theta_{ck} | \sigma_{ek}^2) + \gamma_k \pi(\theta_{ck} | \bar{Y}_{ek}, \sigma_{ek}^2) \]

RWD

Stratum-specific parameter of interest

Trial data

Stratum 1

Stratum 2

Stratum 3

Stratum 4

Stratum 5
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 $\theta_{ck}$ 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

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

RWD

Stratum-specific parameter of interest

Trials Data

Stratum 1 Stratum 2 Stratum 3 Stratum 4 Stratum 5
Power Prior

- Assume $\theta_{ek} = \theta_{ck}$
- Power Prior
  \[ \pi(\theta_{ck} | Y_{ek}, \sigma_{ek}^2, \alpha_k) \propto f(Y_{ek} | \theta_{ck}, \sigma_{ek}^2)^{\alpha_k} \pi_0(\theta_{ck} | \sigma_{ek}^2), \]
  where $0 \leq \alpha_k \leq 1$ and $\pi_0(\theta_{ck} | \sigma_{ek}^2)$ is the density of $N(0, 100 \cdot \sigma_{ek}^2)$
- Discounting Parameter: [Chen and Ibrahim, 2006, Jiang et al., 2020]
  \[ \alpha_k = \frac{1}{\frac{2\varphi_{0k} n_{ek}}{S_{ek}^2} + 1}, \]
  where $\varphi_{0k} = \max\{(Y_{ck} - Y_{ek})^2, 0.10 \cdot S_{ek}^2\}$ to avoid over-borrowing of the external data.
- $\sigma_{ck}^2 \sim IG(0.01, 0.01)$ and $\sigma_{ek}^2 \sim IG(0.01, 0.01)$ for $k = 1, \ldots, K$. 
Framework (Conquer): Combining inference from Different Strata

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

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

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

$w = (w_1, \ldots, w_K)'$ 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 $d$,

$$X_{ci} \sim N_d(M_c, \Sigma_c), \quad X_{ej} \sim N_d(M_e, \Sigma_e)$$

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

- $\Sigma_c = \Sigma_e = I_d$

- Generate outcome

$$Y_{ci} | X_{ci} = \beta_0 + \beta' X_{ci} + \epsilon_i \quad \text{and} \quad Y_{ej} | X_{ej} = \beta_0 + \beta' X_{ej} + \epsilon_j,$$

assuming $\epsilon_i \sim iid N(0, \eta_c^2)$ and $\epsilon_j \sim iid N(0, \eta_e^2)$ for $i = 1, \ldots, n_c$ and $j = 1, \ldots, n_e$.
Selected Simulation Scenarios

- Let $d = 3$, $K = 5$, $\beta_0 = 0$, $\beta' = 1_3$, $\mathcal{M}_e = (0.5, 1, 1)'$, 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 $p_5 = 1$

<table>
<thead>
<tr>
<th>Scenario</th>
<th>$\theta_{\text{true}}$</th>
<th>$n_c$</th>
<th>$n^*_e$</th>
<th>$n_e$</th>
<th>$\eta^2_e$</th>
<th>$\eta^2_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.45</td>
<td>100</td>
<td>1000</td>
<td>960</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3.45</td>
<td>100</td>
<td>1000</td>
<td>960</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>3.45</td>
<td>200</td>
<td>2000</td>
<td>1957</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1: Simulation Scenarios presented with $\theta_{\text{true}}$ 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

<table>
<thead>
<tr>
<th>Weight Scenario</th>
<th>$w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(0.20, 0.20, 0.20, 0.20, 0.20)</td>
</tr>
<tr>
<td>2</td>
<td>(0.10, 0.20, 0.40, 0.20, 0.10)</td>
</tr>
<tr>
<td>3</td>
<td>(0.05, 0.15, 0.60, 0.15, 0.05)</td>
</tr>
<tr>
<td>4</td>
<td>(0.05, 0.10, 0.70, 0.10, 0.05)</td>
</tr>
<tr>
<td>5</td>
<td>(0.00, 0.10, 0.80, 0.10, 0.00)</td>
</tr>
</tbody>
</table>

*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
- \[ \text{Bias} = \frac{1}{R} \sum_{r=1}^{R} (\hat{\theta}_r - \theta_r) \]
- Root mean squared error (RMSE) := \[ \sqrt{\frac{1}{R} \sum_{r=1}^{R} (\hat{\theta}_r - \theta_r)^2} \], where \( \hat{\theta}_r \) denotes the posterior mean of the \( r \)th replicated data set, and \( \theta_r \) denotes the true parameter in the current group in the \( r \)th replicate with \( r = 1, \ldots, R \) and \( R = 1000 \).
- Coverage probability (CP): the number of 95% HPD intervals of \( \theta_r \) that contain the true parameter and divide by the total number of replications.
- Sample standard deviation of the posterior mean over the replicates as
  \[ SD = \frac{1}{R} \sqrt{\sum_{r=1}^{R} \left( \frac{1}{M} \sum_{m=1}^{M} (\theta_r^{[m]} - \hat{\theta}_r)^2 \right) / (M - 1)} \]
  where \( \theta_r^{[m]} \) denotes the \( m \)th iteration of the MCMC sample for the parameter \( \theta_r \) in the \( r \)th replicate and \( m = 1, \ldots, M \).
- Simulation standard error: \[ SE = \sqrt{\frac{1}{R} \sum_{r=1}^{R} (\hat{\theta}_r - \frac{1}{R} \sum_{\ell=1}^{R} \hat{\theta}_\ell)^2} / (R - 1) \]
Hierarchical Approach: Scenario 1 and 2 (Bias and RMSE)
Hierarchical Approach: Scenario 1 and 2 (CP, SD, SE)
Mixture Approach: Scenario 1 and 2 (Bias and RMSE)

Scenario

1. Imbalance with small variance
2. Imbalance with large variance

Bias Plot Scenario 1

RMSE Plot Scenario 1

Bias Plot Scenario 2

RMSE Plot Scenario 2
Mixture Approach: Scenario 1 and 2 (CP, SD, SE)
Power Prior Approach: Scenario 1 and 2 (Bias and RMSE)
Power Prior Approach: Scenario 1 and 2 (CP, SD, SE)
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)