A Bayesian Non-Parametric Causal Inference Model for Comparative Effectiveness Research

CHENGUANG WANG GARY ROSNER

SRCOS 2014

JOHNS HOPKINS
UNIVERSITY

Outline

Comparative Effectiveness Research

- Designed to inform health-care decisions by providing for different treatment options evidence on
  - effectiveness
  - benefits
  - harms
- Synthesize evidence from disparate sources
  - interventional clinical trials
  - observational studies
  - etc.
- *Challenge* is to address study-specific heterogeneities to increase the scientific validity and efficiency of such synthesis analysis

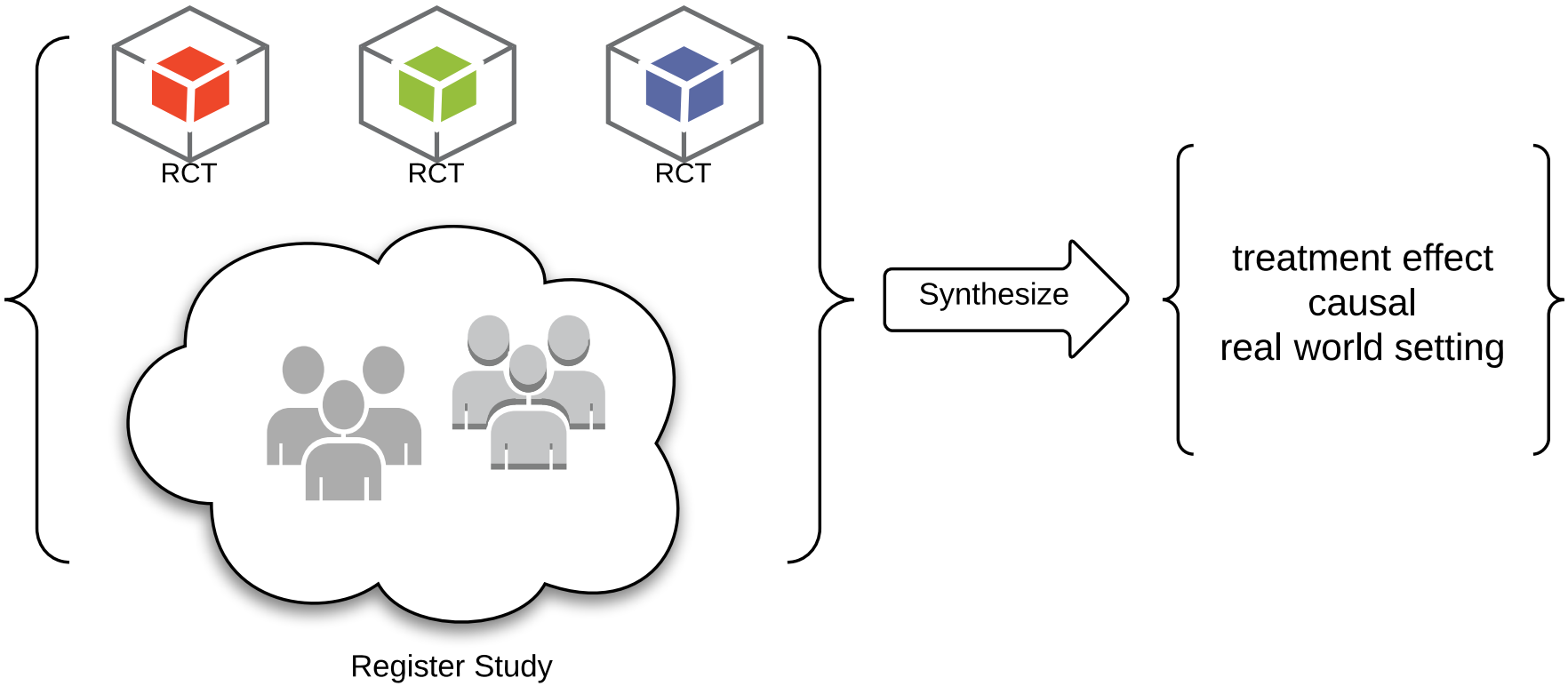
Randomized Clinical Trial (RCT)

- *Gold standard*
- Strict entry criteria
- Participants *randomly* assigned by study investigators to different interventions
- Balanced baseline participant characteristics among different arms
- Straightforward estimation of the *causal* intervention effect
- *Cons:*
  - differ from real-world setting
  - lack of generalizability

Registry Study

- Post-market study focusing on post-approval medical products
- Pre-specified protocol and research questions to assess the benefit/risk profile in the real work
- Physicians treat conditions based on their expert knowledge
- Sufficient size and provide information for non-typical populations
- *Cons:*
  - participants receiving different interventions not ensured to be similar
  - produce biased estimations without proper adjustment

Research Goal



Outline

Potential Outcome

- For subject i, let $Y_i(0)$ and $Y_i(1)$ denote outcomes under control ($Z = 0$) and treatment ($Z = 1$)
- $(Y_i(0), Y_i(1))$ termed as *potential outcome*
- Average treatment effect (ATE) can be defined as

$$E(Y_i(1) - Y_i(0))$$

- In RCTs
  - $Z \perp\!\!\!\perp (Y(0), Y(1))$
  - ATE *identified* by observed outcome

$$\begin{aligned} E(Y_i(1) - Y_i(0)) &= E(Y(1)) - E(Y(0)) \\ &= E(Y(1)|Z = 1) - E(Y(0)|Z = 0) \\ &= E(Y|Z = 1) - E(Y|Z = 0) \end{aligned}$$

Propensity Score

$$e(\boldsymbol{v}) = P(Z = 1|\boldsymbol{V} = \boldsymbol{v}) = E(Z|\boldsymbol{V} = \boldsymbol{v})$$

- Conditional probability of assignment to treatment $Z = 1$ given baseline $\boldsymbol{V}$ covariates
- Rosenbaum and Rubin (1983) proved
  - $Z \perp\!\!\!\perp \boldsymbol{V}|e(\boldsymbol{V})$
  - under strongly ignorable treatment assignment assumptions $Y(0), Y(1) \perp\!\!\!\perp Z|\boldsymbol{V}$ and $0 < P(Z = 1|\boldsymbol{V}) < 1$

$$Y(0), Y(1) \perp\!\!\!\perp Z|e(\boldsymbol{V})$$

- ATE *identified* by

$$E(Y_i(1) - Y_i(0)) = E_{e(\boldsymbol{V})}\{E(Y|Z = 1, e(\boldsymbol{V})) - E(Y|Z = 0, e(\boldsymbol{V}))\}$$

Outline

Notation

- S: clinical study
  - $S = 1, \ldots, J-1$: RCTs
  - $S = J$: register study
- For subject i in study $S = j$
  - Y_{ij}: outcome
  - Z_{ij}: treatment assignment
  - $\boldsymbol{V}_{ij}$: baseline covariates
- *Goal:* $p(Y(Z)|S = J)$ for $Z = 0, 1$

PS Estimation

- $e_J(\boldsymbol{v}) = P(Z = 1 | \boldsymbol{V} = \boldsymbol{v}, S = J)$: propensity score on the register study $S = J$
- Estimate $e_J(\boldsymbol{v})$ by logistic regression

$$\text{logit}\, P(Z_{ij} = 1 | \boldsymbol{V}_{ij} = \boldsymbol{v}_{ij}, S = J) = \alpha_{0,J} + \boldsymbol{v}'_{ij}\boldsymbol{\alpha}_{v,J}$$

- Fit the model to the *register study*

RCT and $e_J(\boldsymbol{v})$

- Consider a subject with $\boldsymbol{v}_{ij}$ on a RCT study $S = j$
- $e_J(\boldsymbol{v}_{ij})$
  - *not* the probability of treatment assignment on $S = j$
  - a summary of baseline covariates
- $[Y_{ij}(Z)|e_J(\boldsymbol{v}_{ij})]$?

Propositions

- $Z \perp\!\!\!\perp \boldsymbol{V} | e_J(\boldsymbol{V}), S = j$
- $Y(0), Y(1) \perp\!\!\!\perp Z | e_J(\boldsymbol{V}), S = j$

Proof: (condition $S = j$ omitted)

$$f(z|\boldsymbol{v}, e_J(\boldsymbol{v})) = f(z|\boldsymbol{v}) = f(z) = f(z|e_J(\boldsymbol{v}))$$

$$f(y(0), y(1)|z, e_J(\boldsymbol{v}))$$

$$= \frac{f(z|y(0), y(1), e_J(\boldsymbol{v}))f(y(0), y(1)|e_J(\boldsymbol{v}))f(e_J(\boldsymbol{v}))}{f(z, e_J(\boldsymbol{v}))}$$

$$= \frac{f(z)f(y(0), y(1)|e_J(\boldsymbol{v}))f(e_J(\boldsymbol{v}))}{f(z)f(e_J(\boldsymbol{v}))}$$

$$= f(y(0), y(1)|e_J(\boldsymbol{v}))$$

Proposal

Share information among studies on the the association of Y_{ij}, $e_J(\boldsymbol{V}_{ij})$, and Z_{ij}

A Typical Bayesian Hierarchical Model

$$[Y_{ij}|e_J(\boldsymbol{V}_{ij}), Z_{ij}, S = j] = \beta_{0,j} + \beta_{z,j} Z_{ij} + \beta_{e,j} e_J(\boldsymbol{V}_{ij}) + \epsilon_{ij}$$

- $\boldsymbol{\beta}_j = \{\beta_{0,j}, \beta_{z,j}, \beta_{e,j}\}$: study-specific parameters
- ϵ_{ij}: follow parametric distribution with mean 0
- $\boldsymbol{\beta}_j \sim \pi(\boldsymbol{\beta})$: assumed to be exchangeable among studies
- Limitations
  - normality?
  - linear?

Dirichlet Process

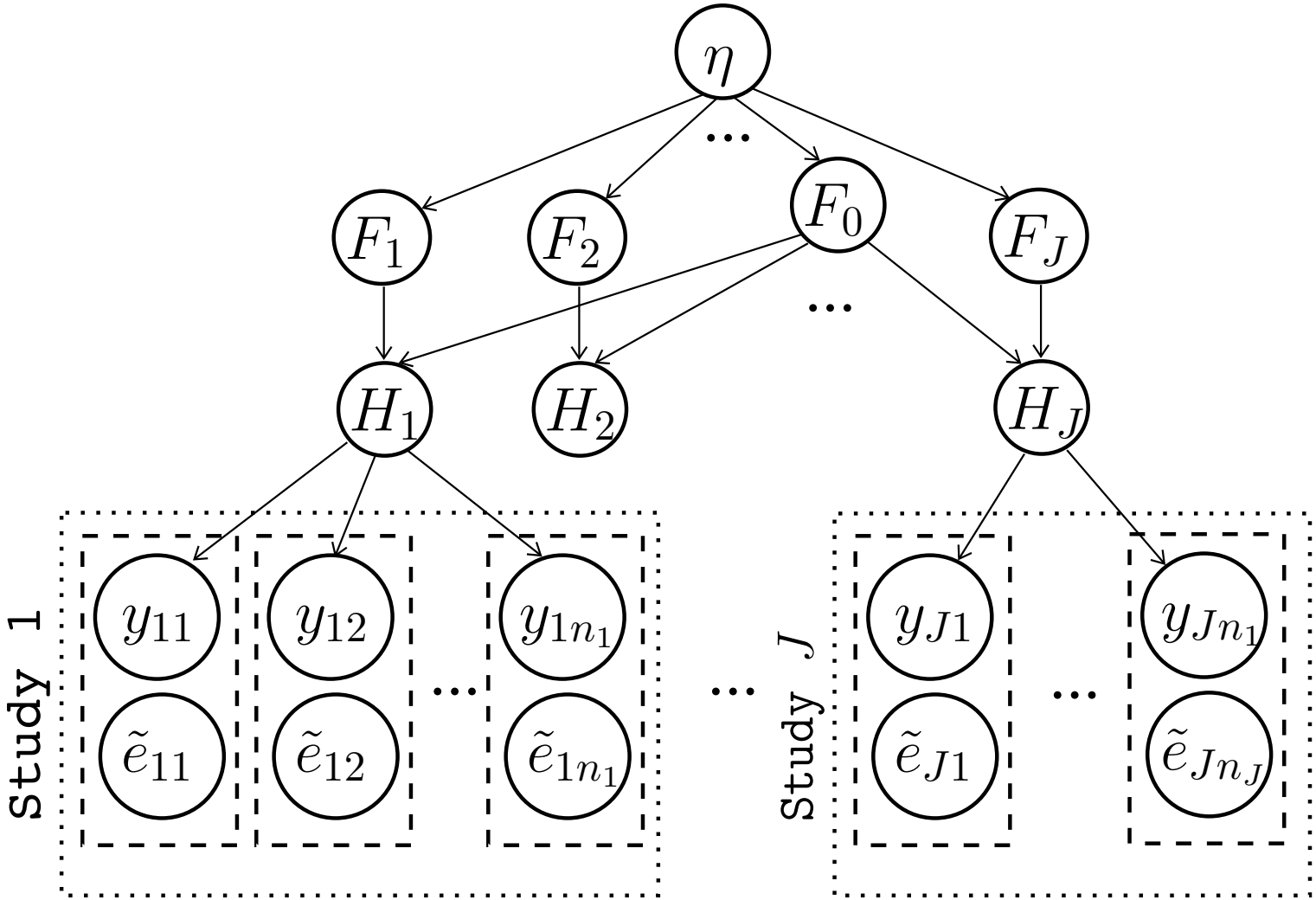
- Distribution on space of distributions
- $G \sim \mathscr{D}(G_0, M)$
- For any partition $(B_1, B_2, \ldots, B_m)$ on the support of G_0, $(G(B_1), \ldots, G(B_m))$ follows a Dirichlet distribution with parameters $(MG_0(B_1), \ldots, MG_0(B_m))$
- As M increases, the variance of G decreases and G is "closer" to G_0

Hierarchical DP

- DP approach often shares information through the full distribution
- More challenging to construct hierarchical model
- Müller et al. (2004)

$$H_j(\cdot \mid \eta) = \varepsilon \underbrace{F_0(\cdot \mid \eta)}_{\substack{\text{common} \\ \text{measure}}} + (1-\varepsilon) \underbrace{F_j(\cdot \mid \eta)}_{\substack{\text{idiosyncratic} \\ \text{measure}}}$$

HDP Model


η
F_1
F_2
F_0
F_J
H_1
H_2
H_J
Study 1
y_{11}
y_{12}
y_{1n_1}
$\tilde{e}_{11}$
$\tilde{e}_{12}$
$\tilde{e}_{1n_1}$
Study J
y_{J1}
y_{Jn_1}
$\tilde{e}_{J1}$
$\tilde{e}_{Jn_J}$

Model Specification

$$\tilde{e}_{ij} = \log \frac{e_J(\boldsymbol{v}_{ij})}{1 - e_J(\boldsymbol{v}_{ij})}$$

$$(y_{ij}, \tilde{e}_{ij} | Z = z, S = j) \sim H_j^{(z)}$$

$$H_j^{(z)} = \epsilon^{(z)} F_0^{(z)} + (1 - \epsilon^{(z)}) F_j^{(z)}$$

$$F_j^{(z)}(\cdot) = \int N(\mu, S^{(z)}) dG_j^{(z)}(\mu)$$

$$G_j^{(z)} \sim \mathscr{D}(G^{(z)}, M_j^{(z)})$$

$$G^{(z)} \sim N(m^{(z)}, B^{(z)})$$

Priors

$$
\begin{aligned}
m^{(z)} &\sim \text{Normal}(m_0, A) \\
B^{(z)} &\sim \text{Inv-Wishart}(c, [cC]^{-1}) \\
S^{(z)} &\sim \text{Inv-Wishart}(q, [qR]^{-1}) \\
M_j^{(z)} &\sim \text{Gamma}(a, b) \\
p(\epsilon^{(z)}) &= \psi_0 \delta_0(\epsilon^{(z)}) + \psi_1 \delta_1(\epsilon^{(z)}) \\
&\quad + (1 - \psi_0 - \psi_1)\, \text{Beta}(a_e, b_e)
\end{aligned}
$$

Predictive Distribution of $Y(z)|S = j$

$$
\begin{aligned}
& p(y(z)|S = j, \boldsymbol{Y}, \boldsymbol{V}) \\
&= \int p(y(z)|S = j, \theta^{(z)}, \boldsymbol{Y}, \boldsymbol{V}) p(\theta^{(z)}|S = j, \boldsymbol{Y}, \boldsymbol{V}) d\theta^{(z)} \\
&= \int p(y(z)|S = j, \theta^{(z)}) p(\theta^{(z)}|S = j, \boldsymbol{Y}, \boldsymbol{V}) d\theta^{(z)} \\
& p(y(z)|S = j, \theta^{(z)}) \\
&= \int p(y(z)|e_J(\boldsymbol{v}), S = J, \theta^{(z)}) p(e_J(\boldsymbol{v})|S = j, \theta^{(z)}) dF(\boldsymbol{v}) \\
&= \int p(y|Z = z, e_J(\boldsymbol{v}), S = J, \theta^{(z)}) p(e_J(\boldsymbol{v})|S = j, \theta^{(z)}) dF(\boldsymbol{v})
\end{aligned}
$$

*$\boldsymbol{\theta}^{(z)} = \{\boldsymbol{\eta}^{(z)}, \boldsymbol{\alpha}_J\}$

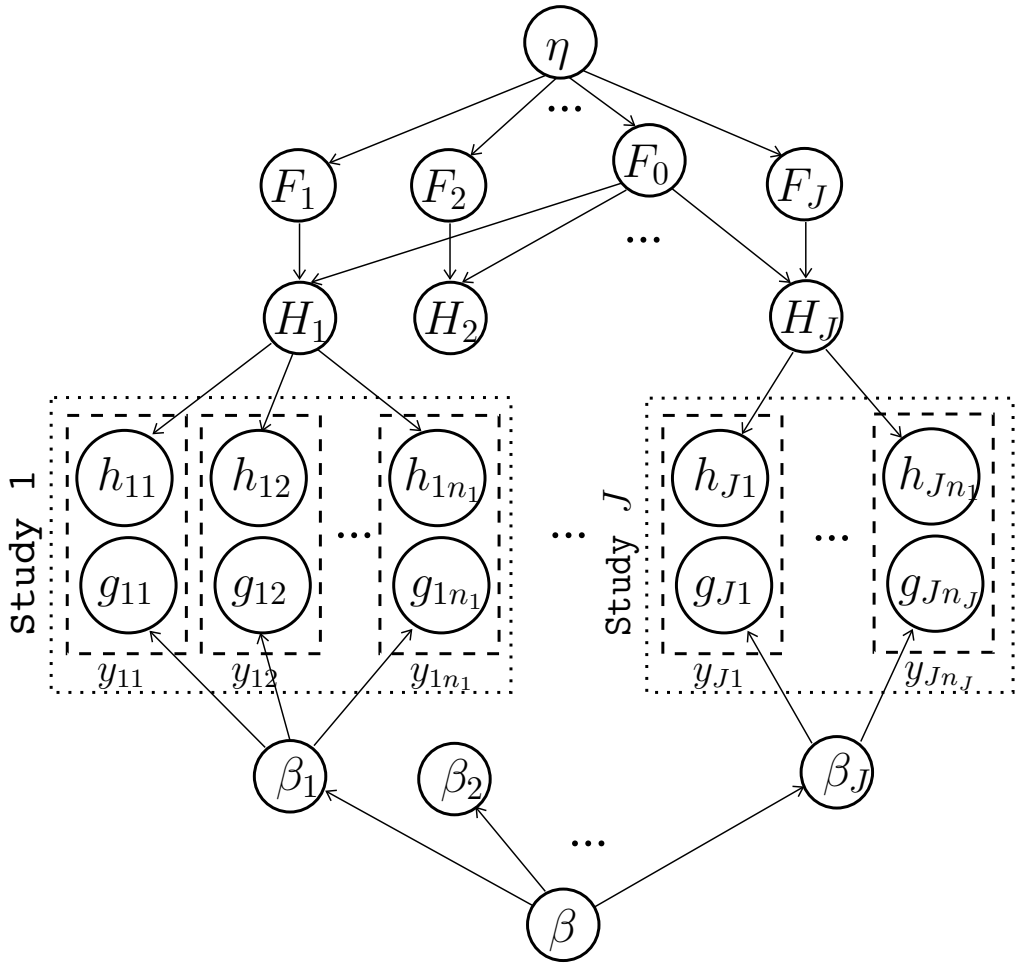
Incorporate Covariates

- Let $\boldsymbol{X} \subset \boldsymbol{V}$ denote the baseline covariates of clinical interest
- For continuous $\boldsymbol{X}$, straightforward to incorporate $\boldsymbol{X}$ in the HDP model

$$(y_{ij}, \tilde{e}_{ij}, \boldsymbol{x}_{ij} | Z = z, S = j) \sim H_j^{(z)}$$

- For discrete $\boldsymbol{X}$, may need alternative approach

Alternative Approach



$$y_{ij} = h_{ij} + g_{ij} + \epsilon_{ij}$$
$$h_{ij} \sim H_j^{(z)}$$
$$g_{ij} \sim \beta_{x,j}\boldsymbol{x}_{ij} + \beta_{e,j}e_J(\boldsymbol{v}_{ij})$$
$$\epsilon_{ij} \sim N(0, \sigma^2)$$
$$\boldsymbol{\beta}_j \sim N(\beta, S)$$

Sensitivity Analysis

- PS only balances covariates that are used to construct the score
- Strong ignorable treatment assignment assumption fails to hold when there are *unmeasured confounders*
- Sensitivity analysis desirable to investigate the robustness of the study findings

Exponential Tilting Model

- Let U denote an unmeasured confounder
- For participants with $Z = 1$, assume

$$P(Z = z|\boldsymbol{V}, U) = \frac{P(Z = z|\boldsymbol{V})e^{\Delta z}}{\sum_{z=0}^{1} P(Z = z|\boldsymbol{V})e^{\Delta z}}$$

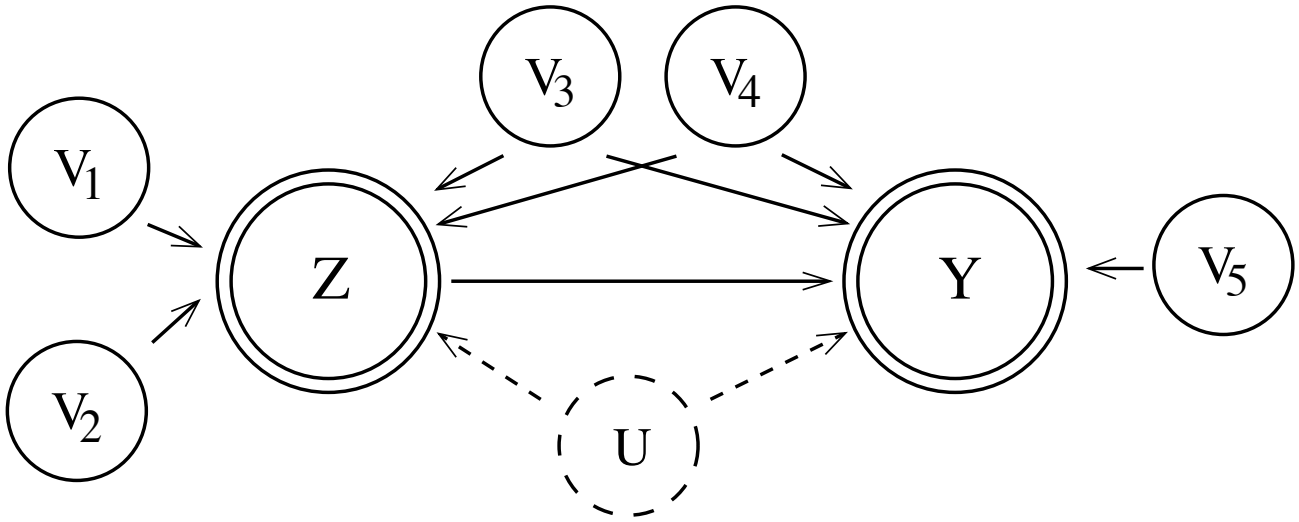
- Equivalent as

$$\text{logit}\, P(Z = 1|\boldsymbol{V}, U) = \alpha_0 + \boldsymbol{V}\boldsymbol{\alpha}_V + \Delta U$$

where $P(U = 1) = 1$ and 0 for $Z = 1$ and 0, respectively

Outline

Configuration



$$Y_{ij} = \beta_{0,j} + \boldsymbol{\beta}_{v,j}\boldsymbol{V}_{ij} + \beta_{u,j}U_{ij} + \beta_{z,j}Z_{ij} + \epsilon_{ij}$$

$$\text{logit}\, P(Z_{iJ} = 1) = \alpha_{0,J} + \boldsymbol{\alpha}_{v,J}\boldsymbol{V}_{iJ} + \alpha_{u,J}U_{iJ}$$

$$\epsilon_{ij} \sim N(0, \sigma^2)$$

Scenario I

- 1 RCT with $N = 200$
- Registry study with $N = 3000$
- No unmeasured confounders
- Identical distributions of $\boldsymbol{V}$
- same fixed β
- 100 replications
- *Stratification approach*

Scenario I: Results

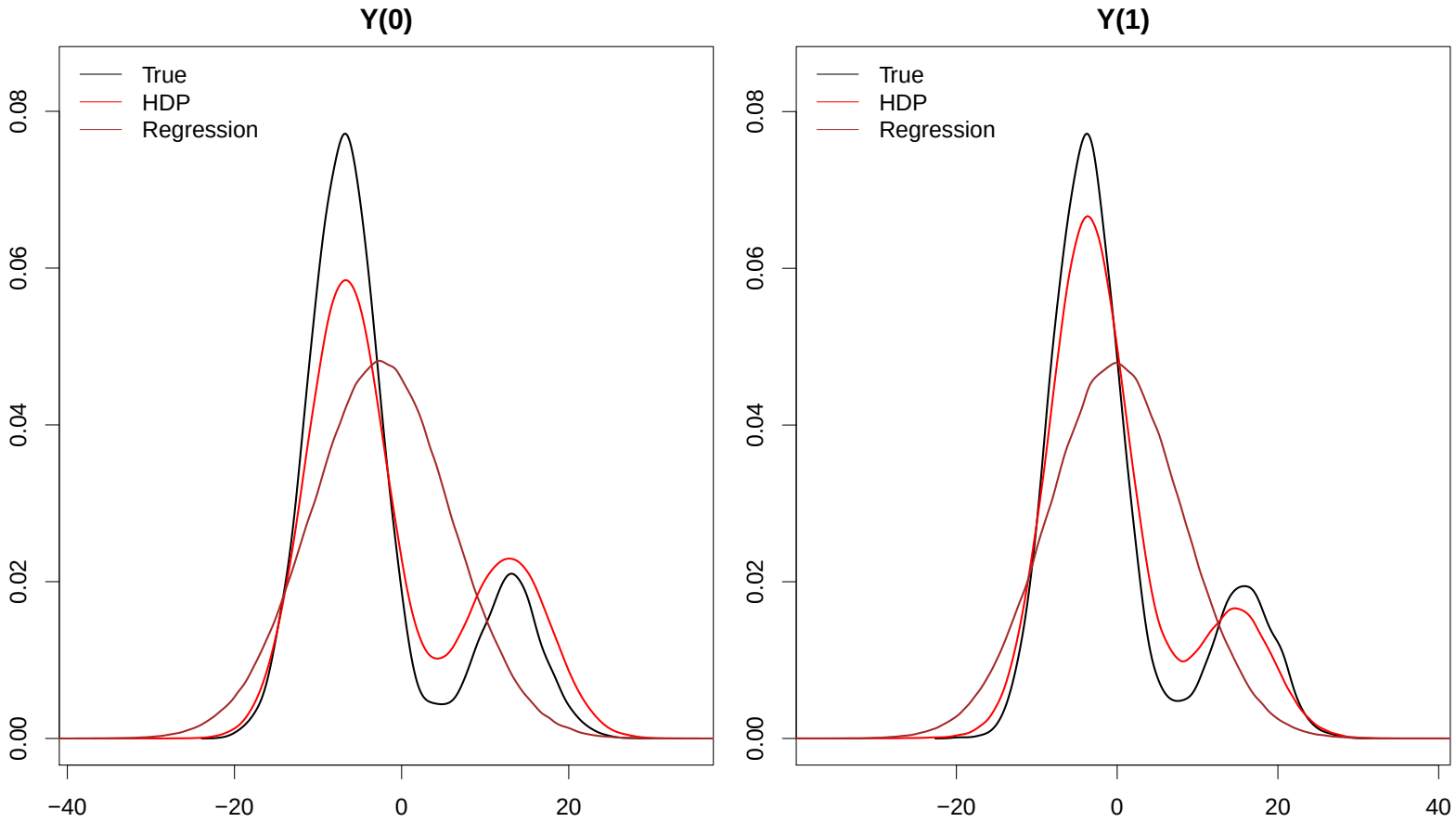
Description	$\widehat{Y}(1) - \widehat{Y}(0)$	$\sqrt{\text{MSE}}$
RCT, no PS	3.02	0.55
RCT, PS	3.02	0.51
Register, PS	2.99	0.15
Register+RCT, PS	2.99	0.12

Scenarios II-V

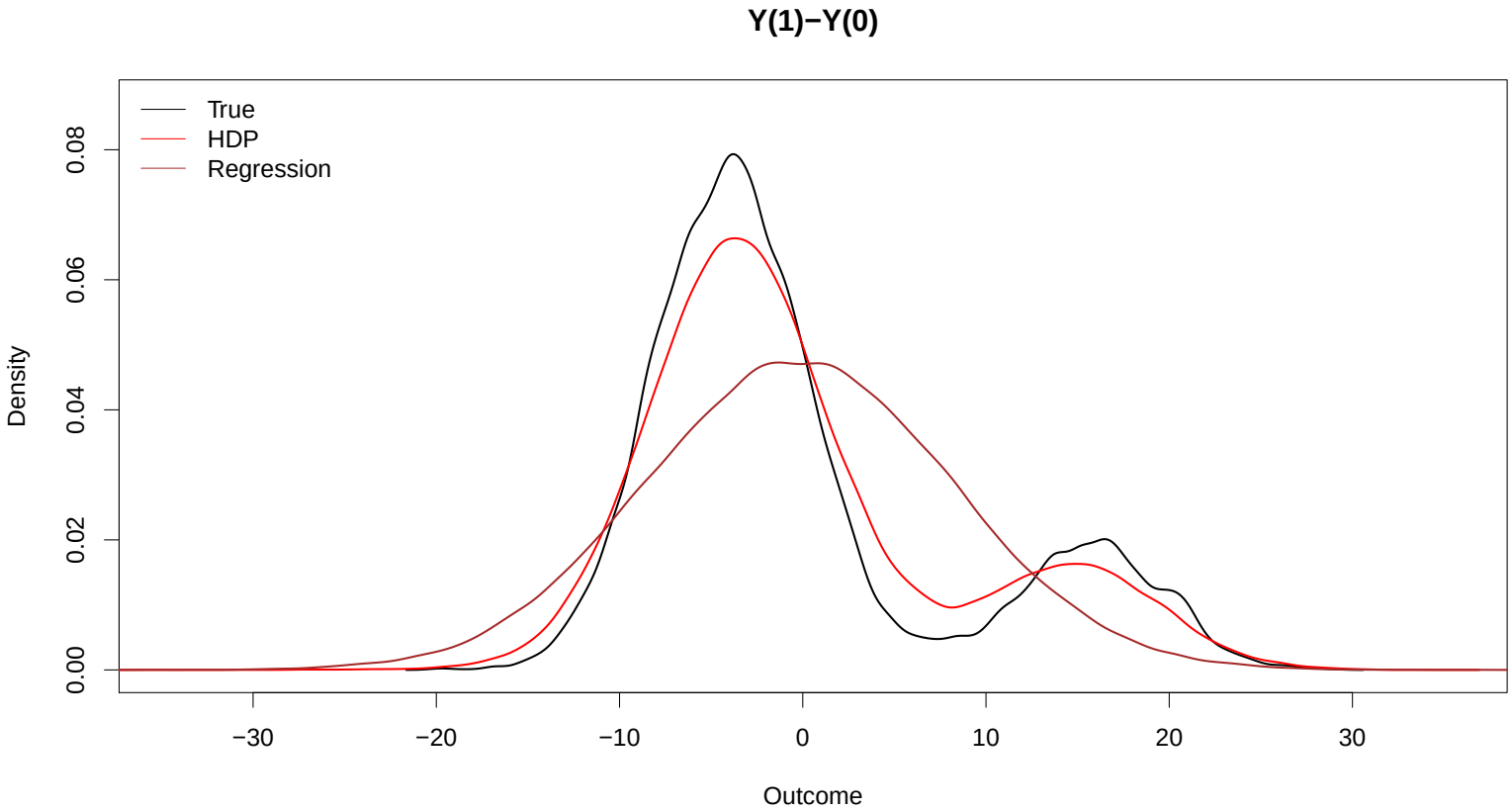
- 3 RCTs each with $N = 200$
- Registry study with $N = 3000$
- No unmeasured confounders
- $\beta_{0,J} = \phi_0 N_1 + \phi_1 N_2$
- Scenarios
  - II: same $p(\boldsymbol{v})$, covariate effects, treatment effect
  - III: different $p(\boldsymbol{v})$
  - IV: different covariate effects
  - V: different treatment effect (0 on $S = J$, positive otherwise)
- Compared with a typical hierarchical model

$$Y_{ij} = \beta_{0,j} + \beta_{z,j} Z_{ij} + \beta_{e,j} e_J(\boldsymbol{V}_{ij}) + \epsilon_{ij}$$

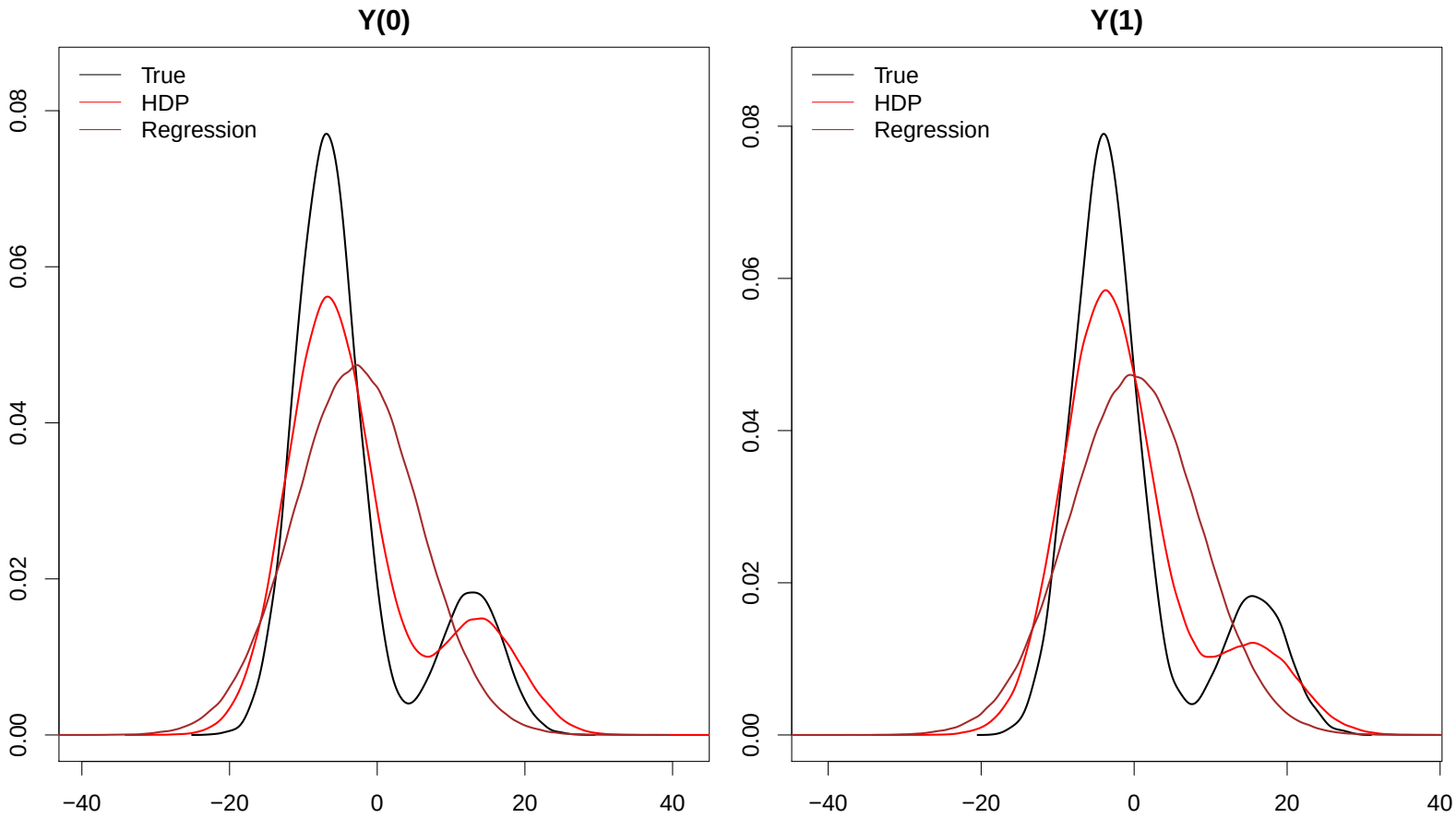
Scenario II: Results



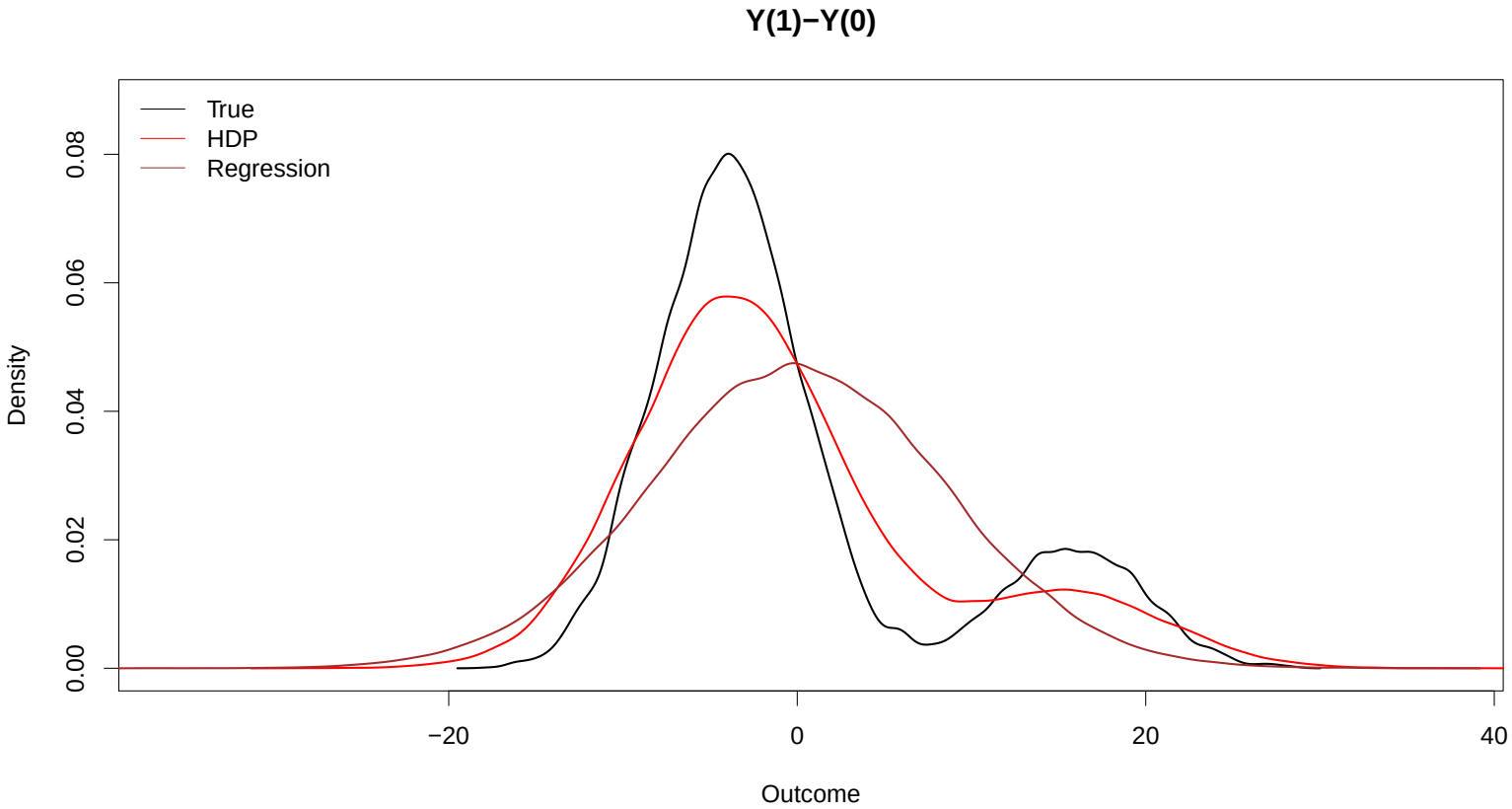
Scenario II: Results



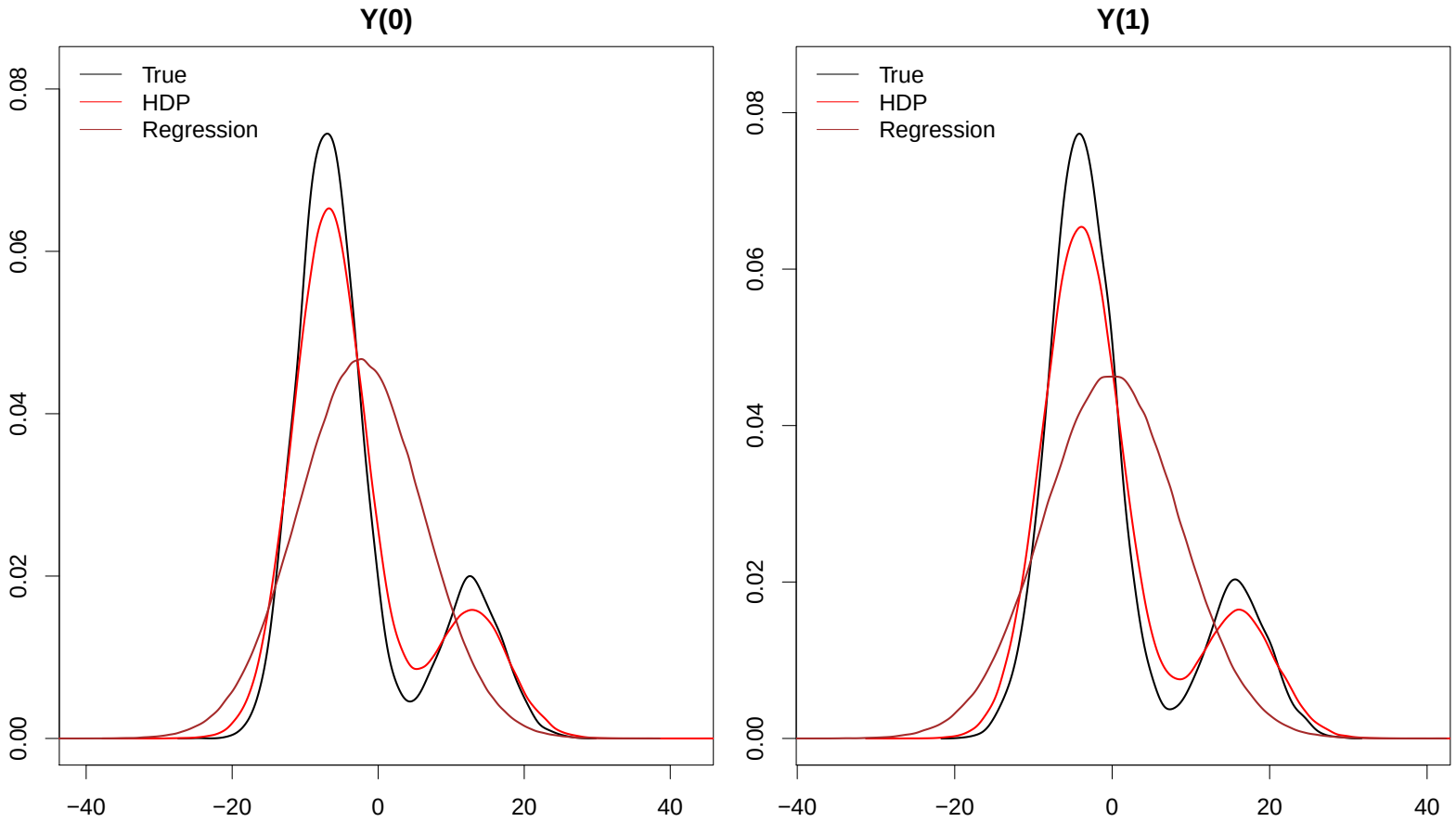
Scenario III: Results



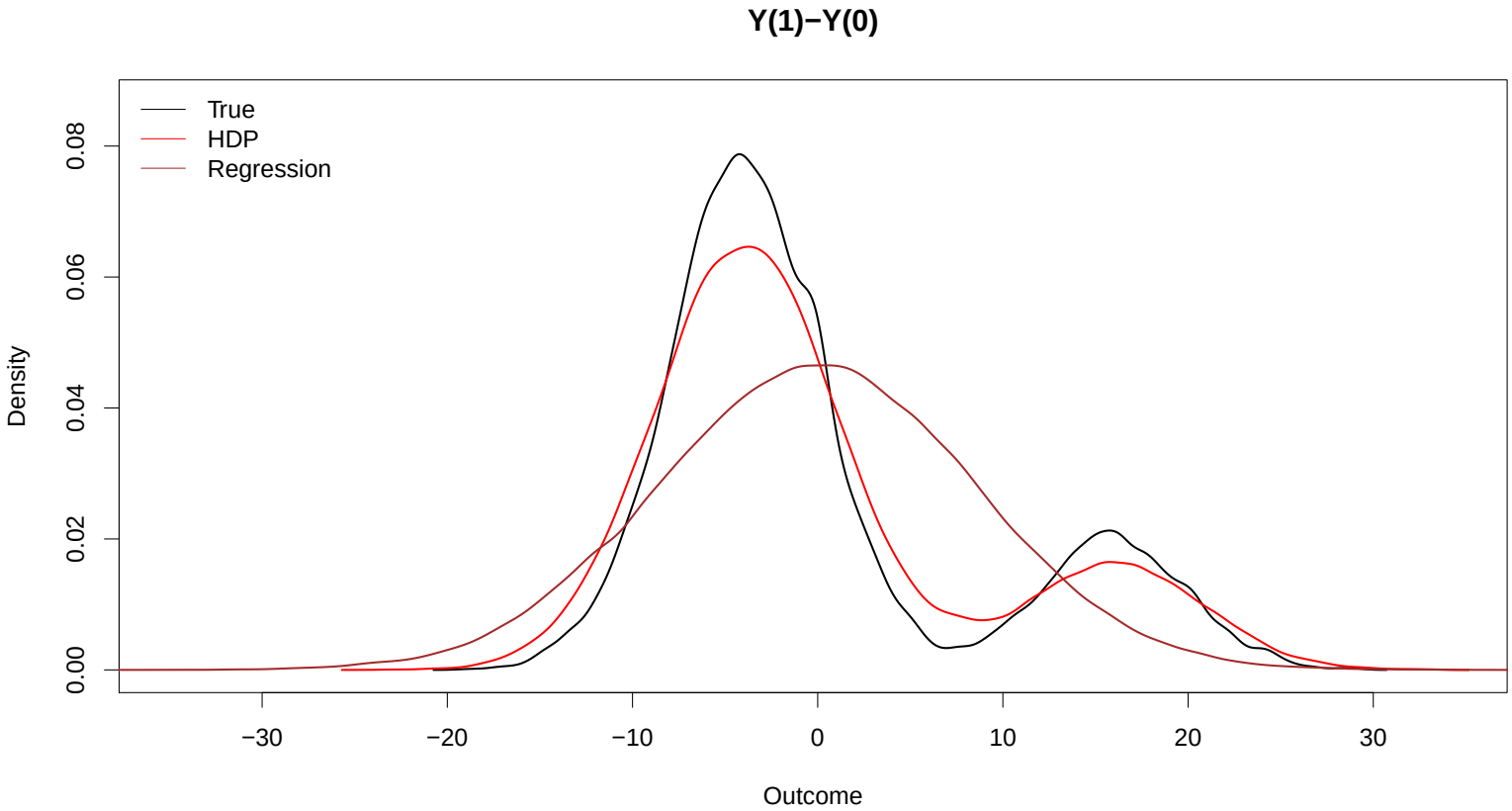
Scenario III: Results



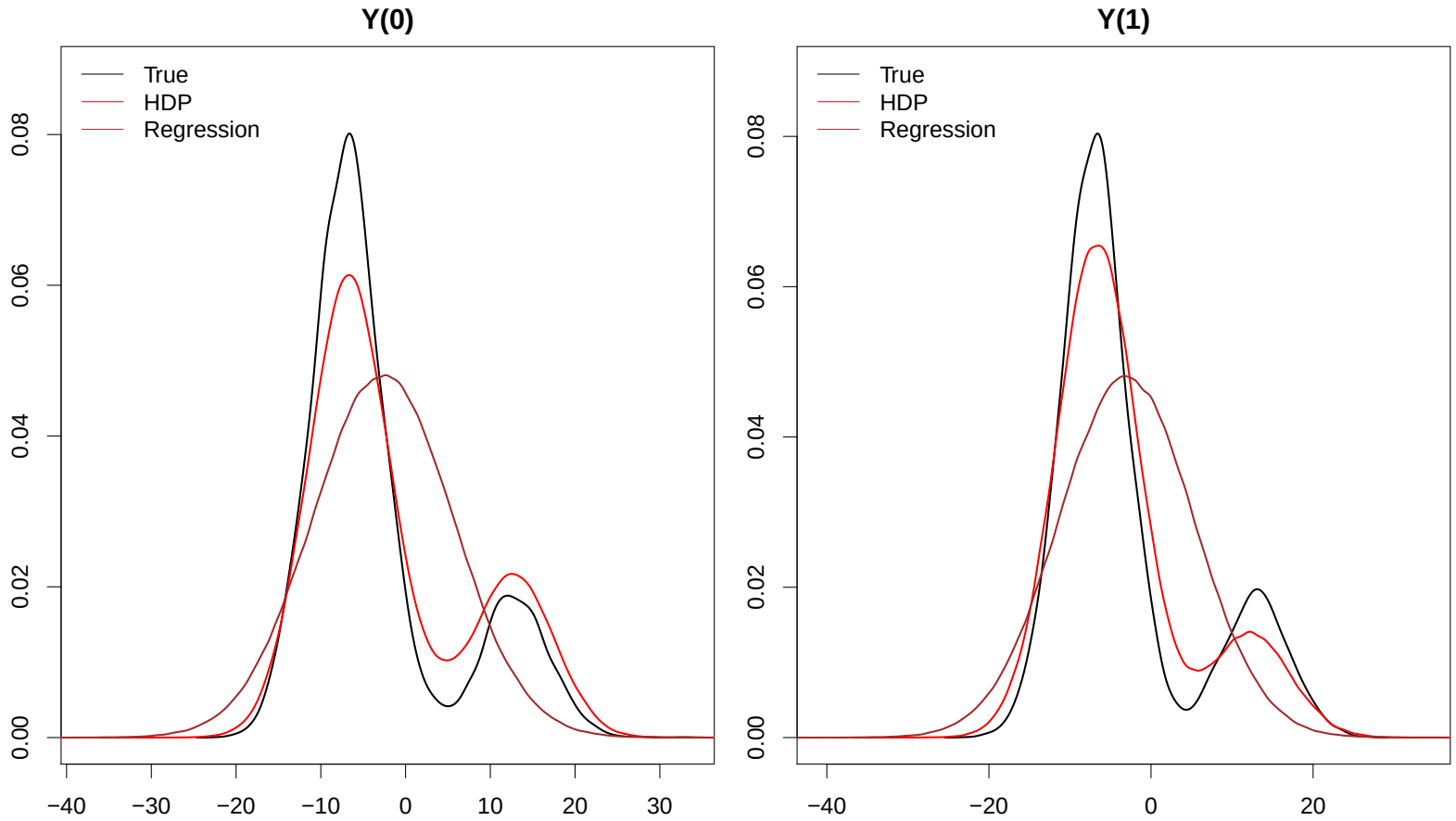
Scenario IV: Results



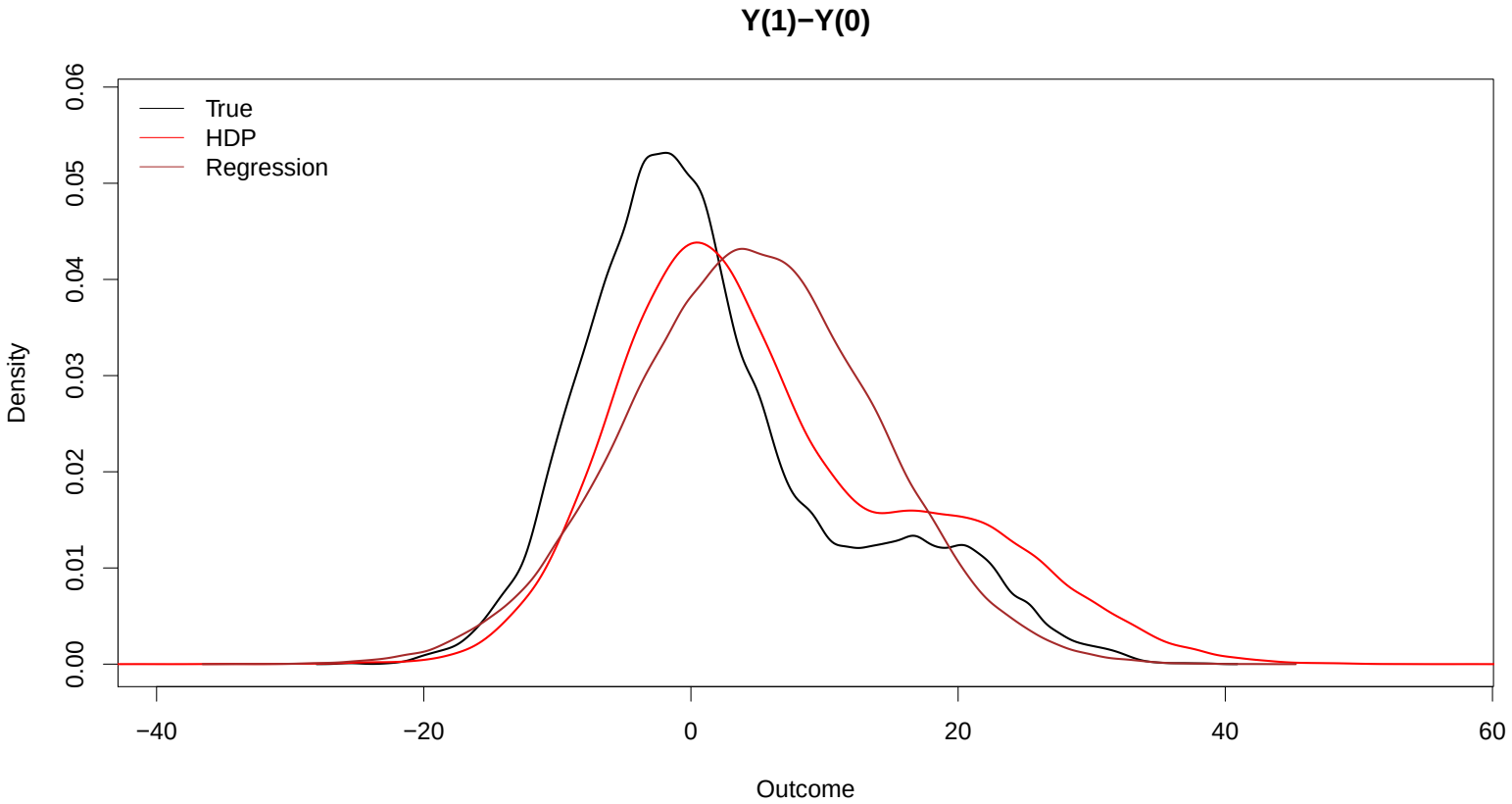
Scenario IV: Results



Scenario V: Results



Scenario V: Results



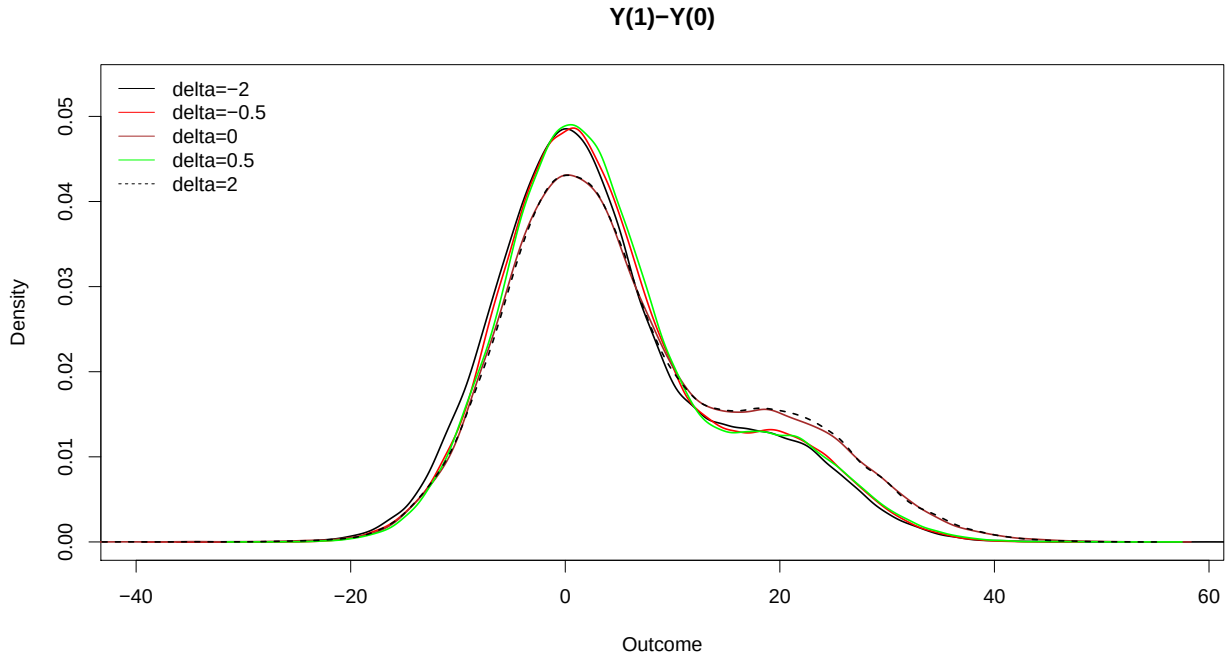
Some Observations

The significance of RCT implications on register study is as follows

1. Different treatment effect (strong)
2. Different $p(\boldsymbol{v})$ (medium)
3. Different covariate effects (weak)

Example of Sensitivity Analysis

- Unmeasured confounder exists
- Correlated with the other covariates
- Effects moderate



Outline

Summary

- Propose Bayesian non-parametric approach to address issues in CER
  - “Propensity score” associate RCTs and registry study
  - Hierarchical DP mixture model
  - Causal inference on the registry study (real world setting)
- Extension
  - Incorporate covariates
  - Better sensitivity analysis strategy

THE END