Integrate Meta-analysis into Specific Study (InMASS) for Estimating Conditional Average Treatment Effect

Keisuke Hanada[1,2], Masahiro Kojima[3,4]

[1]Wakayama Medical University, [2] Osaka University, [3]Chuo University, [4] The Institute of Statistical Mathematics

Motivation: Data Borrowing

Average Treatment Effect (ATE) for the overall population $\mathcal{P}$:

$$\delta_{\mathcal{P}} = \mathrm{E}_{\mathcal{P}}[Y_1 - Y_0] = \mathrm{E}_{\mathcal{P}}[Y \mid Z=1] - \mathrm{E}_{\mathcal{P}}[Y \mid Z=0],$$

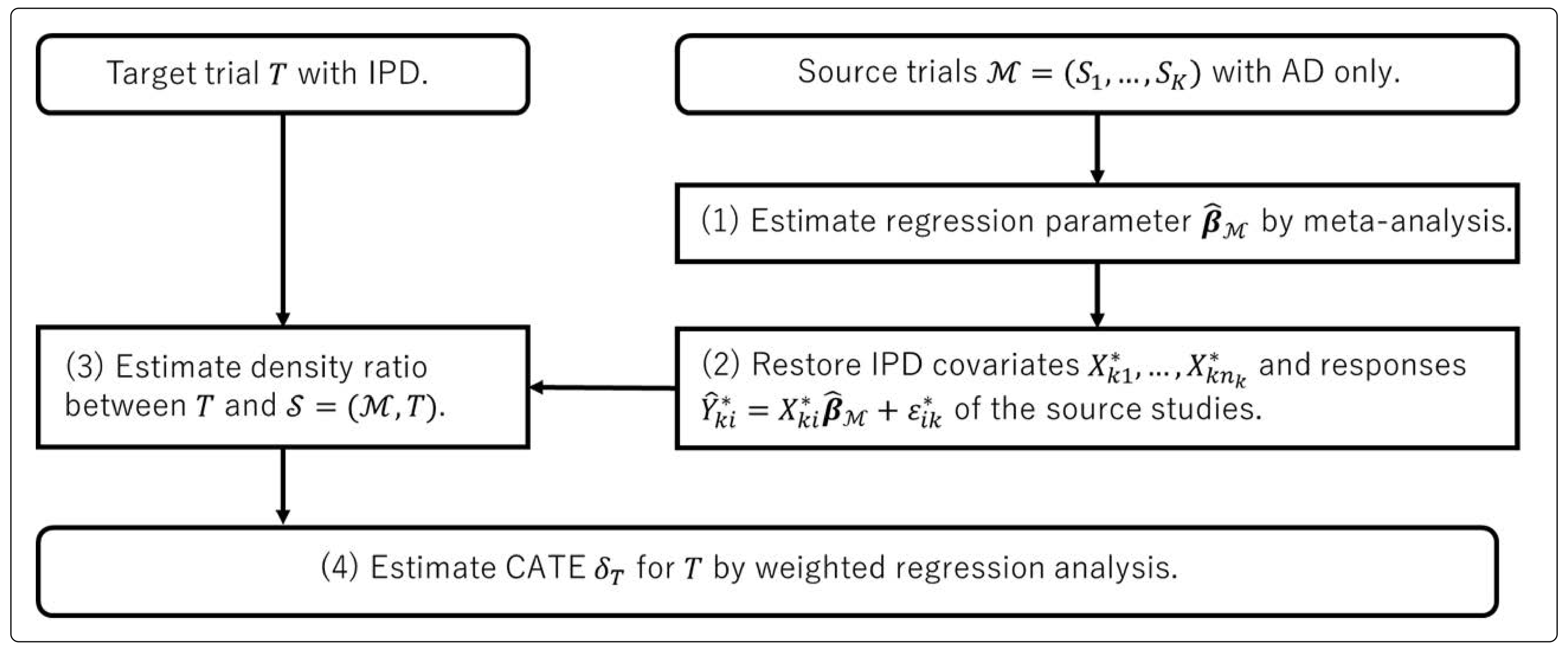
where Y_a $(a=0,1)$ represents the potential outcomes, $\mathrm{E}_{\mathcal{P}}[\cdot] = \int_{\mathcal{P}} \cdot\, dP_{\mathcal{P}}(x,y,z)$ denotes the expectation over the population $\mathcal{P}$, and $Y = ZY_1 + (1-Z)Y_0$.

Aim: To efficiently estimate δ_T by combining individual patient data (IPD) from study T with aggregate data (AD) from external studies $\mathcal{M} = (S_1, \ldots, S_K)$.

Accessible information:

- **IPD for** T: $\{y_{Ti}, \boldsymbol{x}_{Ti}\}_{i=1,\ldots,n_T}$ i.i.d. for each subjects.
- **AD for** $\mathcal{M}$: $\{n_{kj}, \bar{y}_{kj}, \bar{\boldsymbol{x}}_{kj}, \sigma^2_{\bar{y}_{kj}}, \boldsymbol{\sigma}^2_{\bar{\boldsymbol{x}}_{kj}}\}_{j=0,1,\ k=1,\ldots,K}$ i.i.d. for each studies.

Our procedure (InMASS): information borrowing via meta-analysis, domain adaptation by transfer learning, and covariate adjustment using weighted regression



Assumptions & Conditions

InMASS relies on two primary assumptions and three technical conditions:

(A1) Conditional independence: There exists at least one study $S \in \mathcal{S}$ such that $0 < P(Z=1 \mid S) < 1$ and $\{Y_1, Y_0\} \perp\!\!\!\perp Z \mid S$.

(A2) Covariate shift: Suppose that $(\boldsymbol{X}_{T,i}, Y_{T,i}, Z_{T,i}) \sim P_T(\boldsymbol{X}, Y, Z)$ and $(\boldsymbol{X}_{\mathcal{M},i}, Y_{\mathcal{M},i}, Z_{\mathcal{M},i}) \sim P_{\mathcal{M}}(\boldsymbol{X}, Y, Z)$. Then,

$$P_T(Y \mid \boldsymbol{X}, Z) = P_{\mathcal{M}}(Y \mid \boldsymbol{X}, Z).$$

(C1) For any trial $S_k \in \mathcal{M}$, $\bar{\boldsymbol{x}}_k \to \boldsymbol{\mu}_k$ and $\hat{\boldsymbol{\sigma}}^2_{\bar{\boldsymbol{x}}_k} \to \boldsymbol{\sigma}^2_{\boldsymbol{\mu}_k}$ in probability as $n_k \to \infty$.

(C2) Boundedness of Density Ratio: For any subject set $S \in \mathcal{S}$,

$$\sup_{i \in S} \left| \frac{dP_T(\boldsymbol{x}_{Si}, z_{Si})}{dP_{\mathcal{S}}(\boldsymbol{x}_{Si}, z_{Si})} \right| < \infty.$$

(C3) $\frac{1}{N}(\boldsymbol{X}_{\mathcal{S}}^T \boldsymbol{W} \boldsymbol{X}_{\mathcal{S}})^{-1} \to \boldsymbol{Q}$ in probability as $N \to \infty$, where $\boldsymbol{Q}$ is a $(p+2)\times(p+2)$ positive-definite matrix.

Main Results: Our estimator and properties

Univariate case:

The proposed estimator and its variance are given by

$$\hat{\delta}_R = \bar{Y}_1 - \bar{Y}_0, \quad \sigma^2_R = \tilde{N}_1^{-1}\sigma^2_{R1} + \tilde{N}_0^{-1}\sigma^2_{R0}$$

- Weighted mean of group j: $\bar{Y}_j = \tilde{N}_j^{-1} \sum_{S \in \mathcal{S}} \sum_{i=1}^{n_S} w_{Si}\, \mathbb{1}(z_{Si}=j)\, \hat{Y}_{Si}$.
- Outcome for subject i in set S:

$$\hat{Y}_{Si} = \begin{cases} y_{Ti}, & \text{if } S = T, \\ \hat{Y}^*_{ki}, & \text{if } S = S_k, \end{cases}$$

- Weighted sample size: $\tilde{N}_j = \sum_{S \in \mathcal{S}} \tilde{n}_{Sj}$, with $\tilde{n}_{Sj} = \sum_{i=1}^{n_S} w_{Si}\, \mathbb{1}(z_{Si}=j)$.
- the variance for each group j:

$$\sigma^2_{Rj} = \frac{\sum_{S \in \mathcal{S}} \sum_{i=1}^{n_S} w^2_{Si}\, \mathbb{1}(z_{Si}=j) \left(\hat{Y}_{Si} - \mathrm{E}_T[Y \mid Z=j]\right)^2}{\tilde{N}_j}.$$

Theorem 1 (Consistency of $\hat{\delta}_R$ and $\hat{\sigma}^2_R$) *Under (A1),(A2), (C1),(C2), the estimator $\hat{\delta}_R$ is a consistent estimator of δ_T. Moreover, the variance estimator $\hat{\sigma}^2_R$ converges to σ^2_R. Furthermore, if $\delta_T < \infty$ and $\sigma^2_T < \infty$, then $\hat{\delta}_R$ converges in distribution to a standard normal distribution as $N_j \to \infty$ for $j = 0, 1$.*

Theorem 2 *Under assumptions (A1) and (A2) and conditions (C1) and (C2),*

$$\frac{\mathrm{V}[\hat{\delta}_R]}{\mathrm{V}[\hat{\delta}_T]} = O\left(\frac{n_T}{\tilde{N}}\right) \quad \textit{and} \quad \frac{\mathrm{V}[\hat{\delta}_R]}{\mathrm{V}[\hat{\delta}_T]} = O\left(K^{-1}\right). \tag{1}$$

Moreover, if $w_{S_k i} = 0$ for all $k = 1, \ldots, K$ and $w_{Ti} = 1$, then $\mathrm{V}[\hat{\delta}_R]/\mathrm{V}[\hat{\delta}_T] = 1$.

Regression case:

We extend the univariate model to the following linear regression form:

$$\hat{\boldsymbol{Y}}_{\mathcal{S}} = \boldsymbol{X}_{\mathcal{S}}\boldsymbol{\beta}_T + \boldsymbol{\epsilon}_{\mathcal{S}}, \tag{2}$$

where $\boldsymbol{X}_{\mathcal{S}} = (\boldsymbol{1}, \boldsymbol{z}, \boldsymbol{x})$, $\boldsymbol{1} = (1, \ldots, 1)^T$. $\boldsymbol{\beta}_T$ is estimated via weighted least squares as

$$\hat{\boldsymbol{\beta}}_T = (\boldsymbol{X}_{\mathcal{S}}^T \boldsymbol{W} \boldsymbol{X}_{\mathcal{S}})^{-1} \boldsymbol{X}_{\mathcal{S}}^T \boldsymbol{W} \hat{\boldsymbol{Y}}_{\mathcal{S}},$$

where $\boldsymbol{W} = \mathrm{diag}\,(w_{T1}, \ldots, w_{Tn_T}, w_{S_1 1}, \ldots, w_{S_K n_{S_K}})$ is an $N \times N$ diagonal matrix of density ratios. The variance-covariance matrix of $\hat{\boldsymbol{\beta}}_T$ is estimated by

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}_T} = (\boldsymbol{X}_{\mathcal{S}}^T \boldsymbol{W} \boldsymbol{X}_{\mathcal{S}})^{-1} \left(\boldsymbol{X}_{\mathcal{S}}^T \boldsymbol{W} \boldsymbol{\Sigma}_{\varepsilon} \boldsymbol{W} \boldsymbol{X}_{\mathcal{S}}\right) (\boldsymbol{X}_{\mathcal{S}}^T \boldsymbol{W} \boldsymbol{X}_{\mathcal{S}})^{-1}.$$

As $\boldsymbol{\Sigma}_{\varepsilon}$ is unknown, we use a heteroskedasticity-consistent estimator under (A2):

$$\hat{\boldsymbol{\Sigma}}_{\varepsilon} = \mathrm{diag}\left(\boldsymbol{W}^2 (\hat{\boldsymbol{Y}}_{\mathcal{S}} - \boldsymbol{X}_{\mathcal{S}} \hat{\boldsymbol{\beta}}_T)^2\right).$$

Theorem 3 *Under assumptions (A1) and (A2) and conditions (C1) and (C2), if the model (2) is correctly specified, then*

$$\mathrm{E}_{\mathcal{S}}[\hat{\boldsymbol{\beta}}_T] = \boldsymbol{\beta}_T.$$

Furthermore, if condition (C3) holds, then the variance estimator satisfies $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}_T} \to \boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}_T}$ in probability, and $\sqrt{N}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_T)$ converges in distribution to $N(\boldsymbol{0}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}_T})$ as $N \to \infty$.

Experiment: InMASS outperforms IPD-only analysis

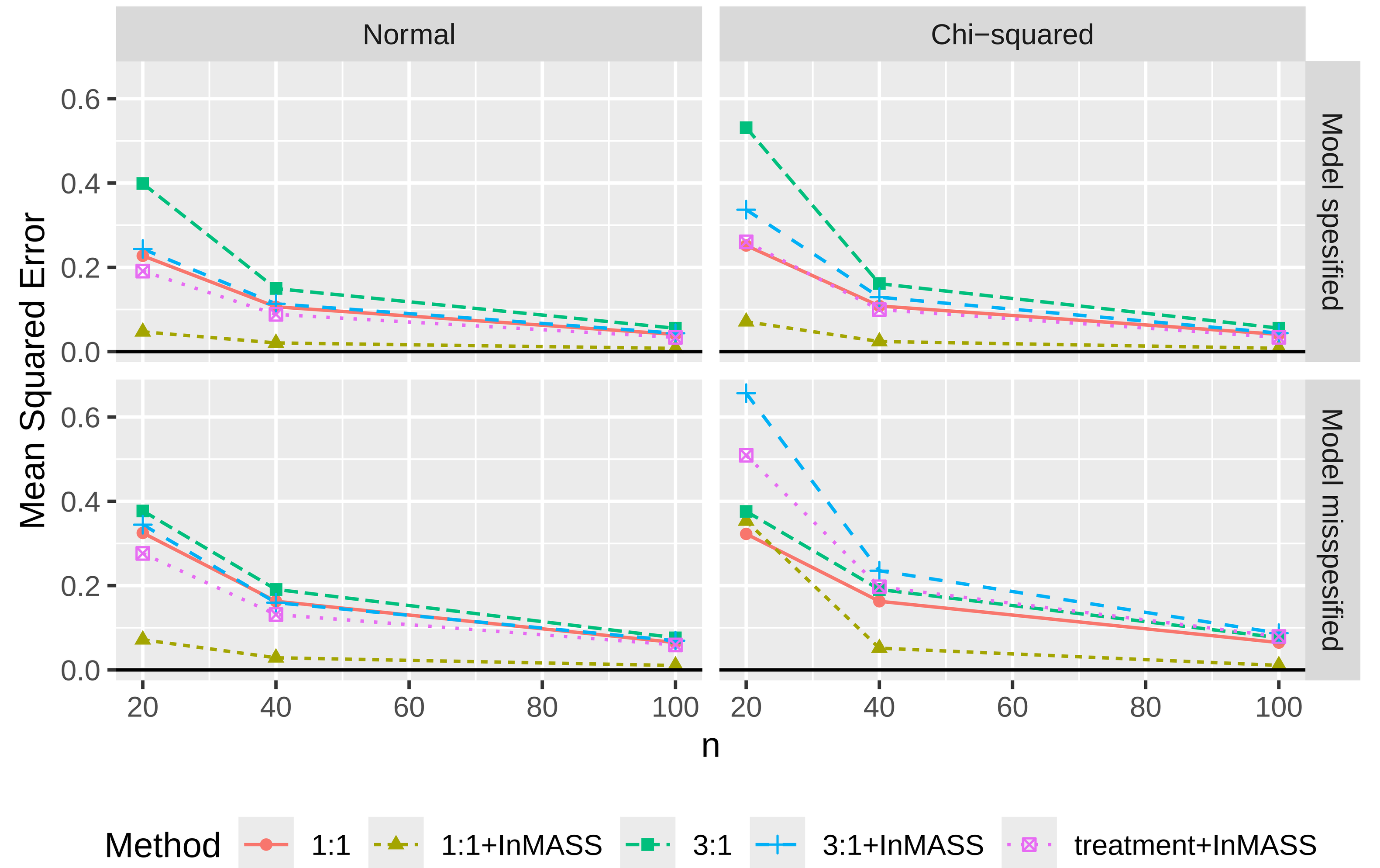
Comparison methods: linear regression and InMASS

Allocation in T:

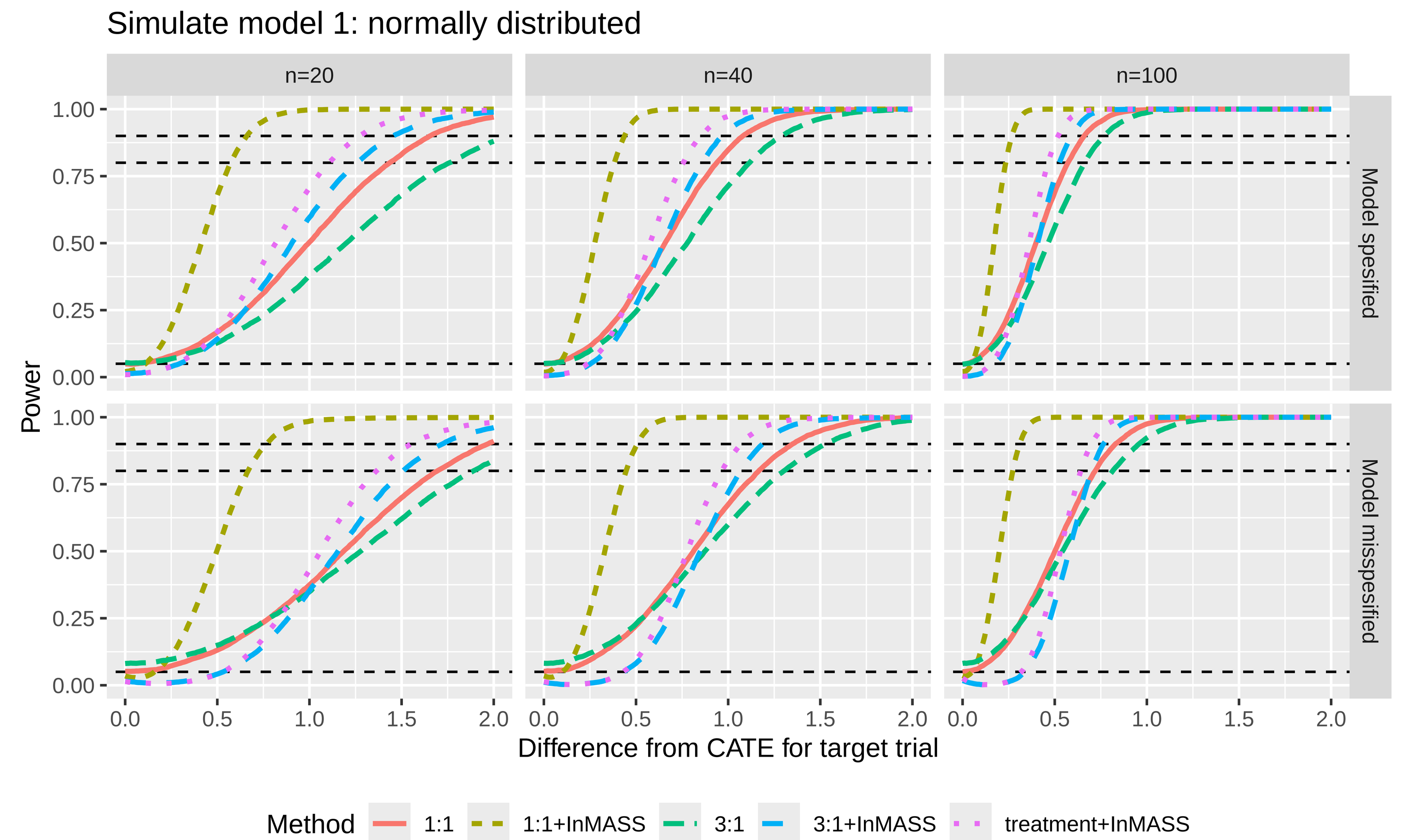
– 1:1: InMASS is applied to both treatment and control groups

– 3:1 and treatment-only: InMASS is applied to the control group only

Simulation of MSE (top) and Power (bottom):



Rows are model specific or misspecified for $P(Y|\boldsymbol{X}, Z)$, columns are normal or chi-square distributed for covariate X.



Rows are model specific or misspecified for $P(Y|\boldsymbol{X}, Z)$, columns correspond to the sample size of target study T.

Results:

- Improves both MSE and power by incorporating AD from meta-analysis.
- Leverages both IPD and AD; provides a consistent estimator.