The impact of directly observed therapy on the efficacy of Tuberculosis treatment: A Bayesian multilevel approach

Widemberg S. Nobre

DME-IM-UFRJ

Agosto 2023

Jointly with Alexandra Schmidt, Erica Moodie and David Stephens

Paper

- Nobre, WS, Schmidt, AM, Moodie, EEM and Stephens, DA. (2023) The impact of directly observed therapy on the efficacy of Tuberculosis treatment: a Bayesian multilevel approach, Journal of the Royal Statistical Society Series C: Applied Statistics. https://doi.org/10.1093/jrsssc/qlad034

Basics of Causal Inference

- Let Y be an outcome of interest, Z be a binary exposure, and $\mathbf{X}$ be a set of covariates
- A dataset is a single realization of the joint process that generates the triple $(Y, Z, \mathbf{X})$
- Associational Inference $\Rightarrow$ Relationships among observed quantities
- Causal Inference $\Rightarrow$ Manipulation of the mechanism that generates the data regarding answering a causal question

Basics of Causal Inference

- What if we intervene by changing an individual exposure status? how much would the outcome change?
- What if we intervene by changing an individual medication? how much would the outcome change?

$\Rightarrow$ When answering causal questions via observed data we aim to create a "bridge" between the observed setting and some hypothetical (or experimental) setting

Basics of Causal Inference

- Let $Y(\mathrm{z})$ be the outcome that would be observed if we intervene to set $Z = \mathrm{z}$ at a particular unit
- A causal effect is a contrast between $Y(0)$ and $Y(1)$ (or $E(Y(0))$ and $E(Y(1))$)
- A first step when investigating a causal question is to specify conditions necessary to answer it

Basics of Causal Inference

- Let $Y(\mathrm{z})$ be the outcome that would be observed if we intervene to set $Z = \mathrm{z}$ at a particular unit
- A causal effect is a contrast between $Y(0)$ and $Y(1)$ (or $E(Y(0))$ and $E(Y(1))$)
- A first step when investigating a causal question is to specify conditions necessary to answer it
  - Why should we care about that?

Basics of Causal Inference

- Let $Y(\mathrm{z})$ be the outcome that would be observed if we intervene to set $Z = \mathrm{z}$ at a particular unit
- A causal effect is a contrast between $Y(0)$ and $Y(1)$ (or $E(Y(0))$ and $E(Y(1))$)
- A first step when investigating a causal question is to specify conditions necessary to answer it

- Why should we care about that? See Simpson's paradox!

Basics of Causal Inference

- Causal Effect of interest: Average Treatment Effect (ATE)

$$\begin{aligned}\tau &= E(Y(1)) - E(Y(0)) \\ &= \int y f_{\varepsilon_1}(y)dy - \int y f_{\varepsilon_0}(y)dy \\ &= \int y \frac{f_{\varepsilon_1}(y)}{f_O(y)} f_O(y)dy - \int y \frac{f_{\varepsilon_0}(y)}{f_O(y)} f_O(y)dy\end{aligned}$$

- If $\frac{f_{\varepsilon_1}(y)}{f_O(y)}$ and $\frac{f_{\varepsilon_0}(y)}{f_O(y)}$ are equal to one, then τ can be computed based on f_O (randomized experiments)

Basics of Causal Inference

- We focus on the problem of confounding, which is a phenomenon that occurs when, in the data generating process, covariates affect both simultaneously, outcome and exposure of interest

- Under no unmeasured confounding (and other mild assumptions), it can be shown that

$$\tau = E(Y(1)) - E(Y(0)) = E(Y|Z = 1, e(X)) - E(Y|Z = 0, e(X)),$$

where $e(X) = P(Z = 1|X)$ is the (correct) propensity score

- While relaxing the no unmeasured confounding assumption, we investigate the need to include random effects in propensity score and outcome regressions to account for unmeasured confounding

Motivation: Tuberculosis dataset

- Cases of Tuberculosis in the state of São Paulo, Brazil in 2016

$\Rightarrow$ Infectious disease caused by the *Mycobacterium tuberculosis* bacteria

$\Rightarrow$ In Brazil, the treatment against Tuberculosis is given by the public health system and lasts for at least six months

- One of the main challenges with the treatment of Tuberculosis is the drug resistance acquired by patients, which is usually due to mismanagement of medications

Basics of Causal Inference: Simpson's paradox

- In the early 1990s, in order to reduce the odds of treatment failure, the World Health Organization (WHO) introduced the directly observed therapy (DOT)

$\Rightarrow$ A health professional must watch the ingestion of medications during the entire treatment process

- Multilevel structured observations: individual into municipalities

Proposed method

- Binary treatment, Z_{ji}, such that $Z_{ji}|\delta_{ji} \sim Bernoulli(\delta_{ji})$ with

$$\text{logit}(\delta_{ji}) = \log\left(\frac{\delta_{ji}}{1-\delta_{ji}}\right) = \sum_{k=1}^{q} \gamma_k X_{kji} + \nu_j, \tag{1}$$

- Binary outcome, such that $Y_{ji}|Z_{ji} \sim Bernoulli(\mu_{ji})$ with

$$\text{logit}(\mu_{ji}) = \beta_0 + \beta_Z Z_{ji} + B_{ji} + \eta_j. \tag{2}$$

Proposed method

- Literature Review: Proposed methods for modeling (1) and (2)

  1. Joint likelihood $\Rightarrow$ Incorrect inference because of a feedback of the outcome into the propensity score model

  2. Cutting Feedback $\Rightarrow$ Incorrect inference because of a measurement error-like problem on the outcome model

  3. Two-step $\Rightarrow$ Provide correct inference and has a fully Bayesian argument

- We follow the two-step procedure and discuss the need to include random effects in propensity score and outcome models to account for unmeasured confounders

Simulation Studies

- Two settings:

1. We consider both Z and Y to be normally distributed yielding a scenario in which bias calculations are analytically tractable

2. We consider both Z and Y to be binary as this is commonly found in epidemiological studies and is representative of our motivating example

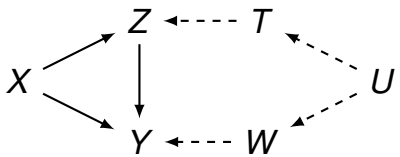


Figure 1: DAG of the data generation mechanism for the simulation studies.

An Analytically Tractable Example: The Linear Case

- Data Generating Mechanism
  - Let X_{ji} be an individual-level covariate generated from a standard normal distribution for $j = 1, \ldots, m$ and $i = 1, \ldots, n_j$, and assume that

$$
\begin{aligned}
Z_{ji} &= \alpha_0 + \alpha_X X_{ji} + T_j + \varepsilon_{ji},\ \varepsilon_{ji} \sim \mathcal{N}(0, \rho^2) && (3)\\
Y_{ji} &= \beta_Z Z_{ji} + \beta_X X_{ji} + W_j + \varepsilon_{ji},\ \varepsilon_{ji} \sim \mathcal{N}(0, \kappa^2), && (4)
\end{aligned}
$$

  - Let $\mathbf{T} = (T_1, \ldots, T_m)^\top$ and $\mathbf{W} = (W_1, \ldots, W_m)^\top$ be two cluster-level covariates whose joint distribution is given by

$$
\begin{pmatrix} \mathbf{T} \\ \mathbf{W} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mu_T \mathbf{1}_m \\ \mu_W \mathbf{1}_m \end{pmatrix}, \begin{pmatrix} \sigma_T^2 \mathbf{I}_m & \rho_{T,W} \sigma_T \sigma_W \mathbf{I}_m \\ \cdot & \sigma_W^2 \mathbf{I}_m \end{pmatrix} \right],
$$

    where $\rho_{T,W}$ is the correlation between $\mathbf{T}$ and $\mathbf{W}$. The quantities ρ in (3), κ in (4), and σ_T and σ_W in (13) are assumed to be known.

An Analytically Tractable Example: The Linear Case

- Model Adjustment
  - Exposure model: For Z_{ji}'s generated according to Equation (3), the following model was fitted

$$Z_{ji} = \alpha_0 + \alpha_X X_{ji} + \nu_j + \varepsilon_{ji},$$

  - Outcome Model: for Y_{ji} generated according to Equation (4), the following model was fitted

$$Y_{ji} = \beta_0 + \beta_Z Z_{ji} + B_{ji} + \eta_j + \varepsilon_{ji},$$

    where B_{ji} indicates how the adjustment for confounding is implemented in the model.

An Analytically Tractable Example: The Linear Case

Table 1: Continuous outcome simulation: fitted models. Data generated as in (3)-(4). The quantities $\widehat{\mathbf{BS}}$ and $\widetilde{\mathbf{BS}}$ are the balancing scores estimated from models described in column $E(\mathbf{Z}|\mathbf{X})$. The cluster-level random effects $\boldsymbol{\nu}$ and $\boldsymbol{\eta}$ are such that $\boldsymbol{\nu} \sim \mathcal{N}(0, \sigma_T^2 \mathbf{I}_m)$ and $\boldsymbol{\eta} \sim \mathcal{N}(0, \sigma_W^2 \mathbf{I}_m)$.

Model	$E(\mathbf{Z}\|\mathbf{X})$	$E(\mathbf{Y}\|\mathbf{Z},\mathbf{X})$
MD1	$\alpha_0\mathbf{1} + \alpha_X\mathbf{X}$	$\beta_0\mathbf{1}_N + \beta_Z\mathbf{Z} + \beta_b\widehat{\mathbf{BS}}$
MD2	$\alpha_0\mathbf{1} + \alpha_X\mathbf{X} + \mathbf{A}\boldsymbol{\nu}$	$\beta_0\mathbf{1}_N + \beta_Z\mathbf{Z} + \beta_b\widetilde{\mathbf{BS}}$
MD3	$\alpha_0\mathbf{1} + \alpha_X\mathbf{X}$	$\beta_0\mathbf{1}_N + \beta_Z\mathbf{Z} + \beta_b\widehat{\mathbf{BS}} + \mathbf{A}\boldsymbol{\eta}$
MD4	$\alpha_0\mathbf{1} + \alpha_X\mathbf{X} + \mathbf{A}\boldsymbol{\nu}$	$\beta_0\mathbf{1}_N + \beta_Z\mathbf{Z} + \beta_b\widetilde{\mathbf{BS}} + \mathbf{A}\boldsymbol{\eta}$

An Analytically Tractable Example: The Linear Case

Focusing on model MD2, it can be shown that

$$\text{Bias}(\widehat{\beta}_Z) = \left[(\mathbf{H}^\top \mathbf{H})^{-1}\mathbf{H}^\top \left(\beta_X \mathbf{X} + \rho_{T,W}\sigma_T\sigma_W \mathbf{A}\mathbf{A}^\top \Sigma^{-1}_{\mathbf{Z}|\mathbf{X}}(\mathbf{Z} - (\alpha_0 + \mu_T)\mathbf{1}_N - \alpha_X \mathbf{X})\right)\right]_{(2)},$$

where $\mathbf{H} = \left[\mathbf{Z} \mid \widetilde{\mathbf{BS}}\right]$ and $[\cdot]_{(2)}$ indicates the second element of the vector, and $\Sigma_{\mathbf{Z}|\mathbf{X}} = \text{Var}(\mathbf{Z}|\mathbf{X})$.

An Analytically Tractable Example: The Linear Case

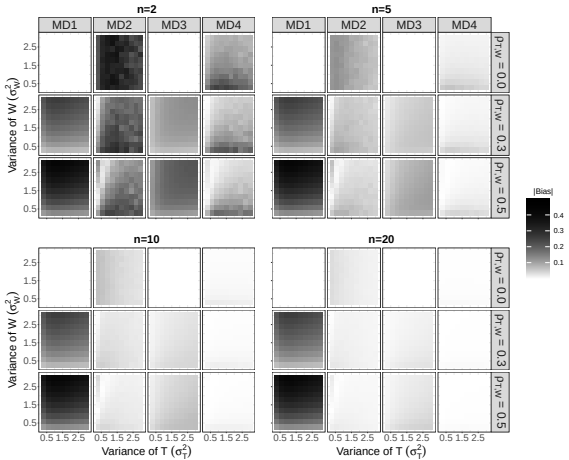


Figure 2: Absolute bias of $\widehat{\beta}_Z$ under the models described in Table 1. These results are averaged over 1000 Monte Carlo replicates.

A Simulation Study with Binary Exposure and Outcome

- Data Generating Mechamism
  - Confounders X_1 and X_2

    Scenario 1: $v_{ji} \sim \mathcal{N}(0, 0.1^2)$ and $\zeta_j \sim \mathcal{N}(0, 0.4^2)$;

    Scenario 2: $v_{ji} \sim \mathcal{N}(0, 0.25^2)$ and $\zeta_j \sim \mathcal{N}(0, 1)$.

  - Exposure $Z_{ji}|\delta_{ji} \sim Bernoulli(\delta_{ji})$

$$\text{logit}(\delta_{ji}) = \alpha_0 + X_{1ji}\alpha_1 + X_{2ji}\alpha_2 + T_j.$$

  - Outcome $Y_{ji}|\mu_{ji} \sim Bernoulli(\mu_{ji})$

$$\text{logit}(\mu_{ji}) = \beta_0 + X_{1ji}\beta_1 + X_{2ji}\beta_2 + X_{1ji}X_{2ji}\beta_3 + W_j.$$

A Simulation Study with Binary Exposure and Outcome

- Model Adjustment

Table 2: Binary outcome simulation: fitted models. Data generated as in (**??**)-(**??**). The quantities $\widehat{PS}_{ji}$ and $\widetilde{PS}_{ji}$ are the propensity scores estimated from models described in column $\text{logit}(\delta_{ji})$. The cluster-level random effects ν_j and η_j are such that $\nu_j \sim \mathcal{N}(0, \varphi^2)$ and $\eta_j \sim \mathcal{N}(0, \phi^2)$, for $j = 1, \ldots, m$.

Model	$\text{logit}(\delta_{ji})$	$\text{logit}(\mu_{ji})$
MD1	$\gamma_0 + \gamma_1 X_{1ji} + \gamma_2 X_{2ji}$	$\beta + \beta_Z Z_{ji} + \beta_b \widehat{PS}_{ji}$
MD2	$\gamma_0 + \gamma_1 X_{1ji} + \gamma_2 X_{2ji} + \nu_j$	$\beta + \beta_Z Z_{ji} + \beta_b \widetilde{PS}_{ji}$
MD3	$\gamma_0 + \gamma_1 X_{1ji} + \gamma_2 X_{2ji}$	$\beta + \beta_Z Z_{ji} + \beta_b \widehat{PS}_{ji} + \eta_j$
MD4	$\gamma_0 + \gamma_1 X_{1ji} + \gamma_2 X_{2ji} + \nu_j$	$\beta + \beta_Z Z_{ji} + \beta_b \widetilde{PS}_{ji} + \eta_j$

A Simulation Study with Binary Exposure and Outcome

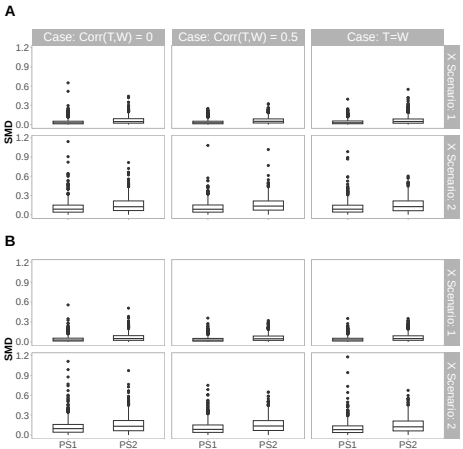
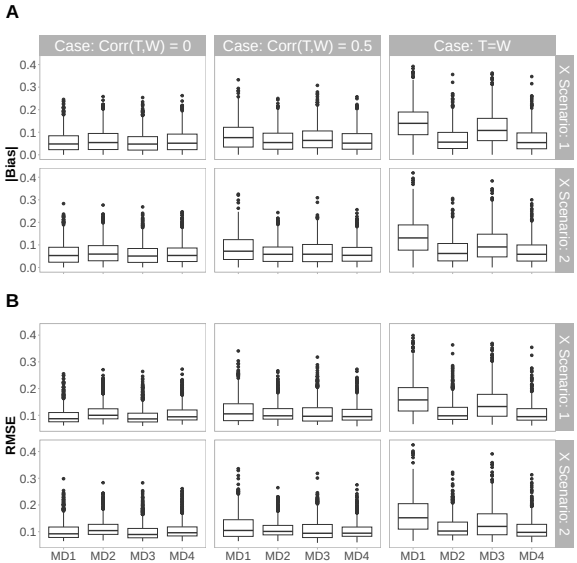


Figure 3: Boxplots of the weighted SMD of X_1 (Panel A) and X_2 (Panel B) under propensity scores estimated from models $\text{PS1} : \text{logit}(\delta_{ji}) = \gamma_0 + \gamma_1 X_{1ji} + \gamma_2 X_{2ji}$ and $\text{PS2} : \text{logit}(\delta_{ji}) = \gamma_0 + \gamma_1 X_{1ji} + \gamma_2 X_{2ji} + v_j$, over 1000 Monte Carlo data replicates.

A Simulation Study with Binary Exposure and Outcome



TB Data Analysis

- Z_{ij} denotes a binary exposure that indicates if individual i in the jth city received the DOT
- Y_{ij} denotes the outcome of interest that indicates if individual i in the jth city had a diagnosis of cure at the end of the treatment
- Let $\mathbf{X}_{ij} = (X_{1ij}, \cdots, X_{pij})$ be a p-dimensional vector of predictors for individual i in the jth city. The vector $\mathbf{X}_{ij}$ comprises both, individual and cluster characteristics

TB Data Analysis

⇒ Individual-level characteristics: indicator variables for diagnosis of Acquired Immunodeficiency Syndrome (AIDS), diagnosis of diabetes, reporting (illicit) drug use, diagnosis of alcoholism, being homeless, gender, whether currently a prisoner, diagnosis of a mental illness, and current smoking status. Additionally, type of TB, and age (in years) are available.

⇒ At the cluster-level, only the Human Development Index (HDI) is available.

TB Data analysis: Exposure model specification

We assume $Z_{ji}|\delta_{ji} \sim Bernoulli(\delta_{ji})$, with

$$\text{logit}(\delta_{ji}) = \log\left(\frac{\delta_{ji}}{1-\delta_{ji}}\right) = \sum_{k=1}^{q} \gamma_k X_{kji} + \nu_j,$$

- PS1: $\nu_j = 0$, for all j;
- PS2: $\nu_j \sim \mathcal{N}(0, \varphi^2)$, for all j; and
- PS3: $\nu \sim \mathcal{N}(\mathbf{0}, \varphi^2 \mathbf{R}(\lambda))$, where $R_{ij} = \text{Corr}(\nu_i, \nu_j) = \exp\left(-\lambda ||\mathbf{s}_i - \mathbf{s}_j||\right)$, with $\mathbf{s}_j$ denoting the centroid of city j (a two-dimensional vector of coordinates) and $||\cdot||$ denoting the Euclidean distance.

TB Data analysis: Exposure

Table 3: Exposure model comparison.

	elpd (WAIC)	pWAIC	WAIC	elpd (LOO)	pLOO	LOO
PS1	-8136.17	14.10	16272.35	-8136.23	14.16	16272.45
PS2	-6884.15	242.63	13768.30	-6893.51	251.99	13787.02
PS3	-6877.74	235.82	**13755.49**	-6885.21	243.28	**13770.42**

TB Data analysis: Exposure

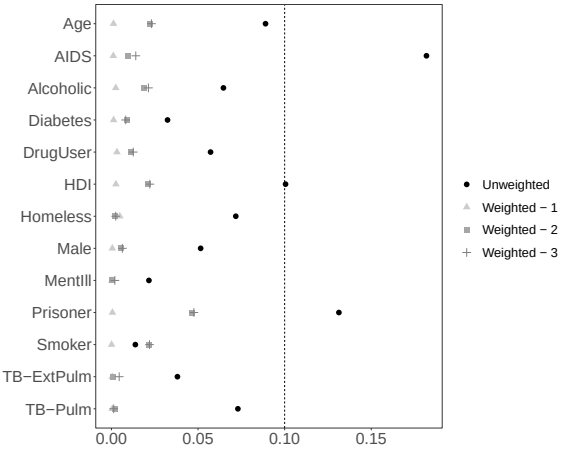


Figure 5: Standardized mean difference (SMD) for the observed baseline covariates between treated and control subjects.

TB Data analysis: Outcome model specification

Model	Outcome Model	Distribution of the random effect
M1	$\text{logit}\{\mu_{ij}\} = \beta + Z_{ij}\beta_Z$	–
M2	$\text{logit}\{\mu_{ij}\} = \beta + Z_{ij}\beta_Z + \eta_j$	$\eta \vert \phi \sim \mathcal{N}(\mathbf{0}, \phi^2 \mathbf{I}_N)$
M3	$\text{logit}\{\mu_{ij}\} = \beta + Z_{ij}\beta_Z + \eta_j$	$\eta \vert \phi, \lambda_y \sim \mathcal{N}(\mathbf{0}, \phi^2 \mathbf{R}(\lambda_y))$
M4	$\text{logit}\{\mu_{ij}\} = \beta + Z_{ij}\beta_Z + \mathbf{X}_{ij}^T\beta_X$	–
M5	$\text{logit}\{\mu_{ij}\} = \beta + Z_{ij}\beta_Z + \mathbf{X}_{ij}^T\beta_X + \eta_j$	$\eta \vert \phi \sim \mathcal{N}(\mathbf{0}, \phi^2 \mathbf{I}_N)$
M6	$\text{logit}\{\mu_{ij}\} = \beta + Z_{ij}\beta_Z + \mathbf{X}_{ij}^T\beta_X + \eta_j$	$\eta \vert \phi, \lambda_y \sim \mathcal{N}(\mathbf{0}, \phi^2 \mathbf{R}(\lambda_y))$
M7	$\text{logit}\{\mu_{ij}\} = \beta + Z_{ij}\beta_Z + PS1_{ij}\beta_{ps}$	–
M8	$\text{logit}\{\mu_{ij}\} = \beta + Z_{ij}\beta_Z + PS1_{ij}\beta_{ps} + \eta_j$	$\eta \vert \phi \sim \mathcal{N}(\mathbf{0}, \phi^2 \mathbf{I}_N)$
M9	$\text{logit}\{\mu_{ij}\} = \beta + Z_{ij}\beta_Z + PS1_{ij}\beta_{ps} + \eta_j$	$\eta \vert \phi, \lambda_y \sim \mathcal{N}(\mathbf{0}, \phi^2 \mathbf{R}(\lambda_y))$
M10	$\text{logit}\{\mu_{ij}\} = \beta + Z_{ij}\beta_Z + PS2_{ij}\beta_{ps}$	–
M11	$\text{logit}\{\mu_{ij}\} = \beta + Z_{ij}\beta_Z + PS2_{ij}\beta_{ps} + \eta_j$	$\eta \vert \phi \sim \mathcal{N}(\mathbf{0}, \phi^2 \mathbf{I}_N)$
M12	$\text{logit}\{\mu_{ij}\} = \beta + Z_{ij}\beta_Z + PS2_{ij}\beta_{ps} + \eta_j$	$\eta \vert \phi, \lambda_y \sim \mathcal{N}(\mathbf{0}, \phi^2 \mathbf{R}(\lambda_y))$
M13	$\text{logit}\{\mu_{ij}\} = \beta + Z_{ij}\beta_Z + PS3_{ij}\beta_{ps}$	–
M14	$\text{logit}\{\mu_{ij}\} = \beta + Z_{ij}\beta_Z + PS3_{ij}\beta_{ps} + \eta_j$	$\eta \vert \phi \sim \mathcal{N}(\mathbf{0}, \phi^2 \mathbf{I}_N)$
M15	$\text{logit}\{\mu_{ij}\} = \beta + Z_{ij}\beta_Z + PS3_{ij}\beta_{ps} + \eta_j$	$\eta \vert \phi, \lambda_y \sim \mathcal{N}(\mathbf{0}, \phi^2 \mathbf{R}(\lambda_y))$

TB Data analysis: Outcome

Table 4: Outcome model comparison.

	elpd (WAIC)	pWAIC	WAIC	elpd (LOO)	pLOO	LOO
M1	-3788.84	1.99	7577.67	-3788.84	2.00	7577.69
M2	-3622.66	120.45	7245.32	-3623.69	121.48	7247.38
M3	-3622.60	116.19	7245.20	-3623.55	117.15	7247.10
M4	-3538.21	15.07	7076.42	-3538.27	15.13	7076.54
M5	-3435.76	115.90	**6871.52**	-3436.61	116.74	**6873.21**
M6	-3436.19	111.72	6872.37	-3436.94	112.47	6873.88
M7	-3649.03	2.88	7298.06	-3649.04	2.89	7298.07
M8	-3535.18	104.18	7070.36	-3535.90	104.90	7071.81
M9	-3535.41	101.62	7070.82	-3536.10	102.31	7072.19
M10	-3734.95	2.90	7469.90	-3734.96	2.91	7469.92
M11	-3612.22	138.22	7224.45	-3613.87	139.87	7227.74
M12	-3613.57	131.75	7227.15	-3615.02	133.20	7230.05
M13	-3735.10	2.99	7470.21	-3735.11	3.00	7470.23
M14	-3613.55	137.26	7227.09	-3615.14	138.86	7230.28
M15	-3614.23	131.34	7228.47	-3615.66	132.77	7231.32

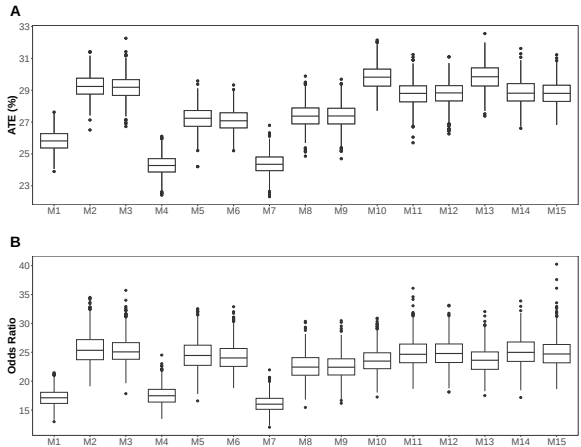


Figure 6: Posterior distributions of the ATE (Panel A) and Odds Ratio (Panel B) of the models above.

Discussion

- We investigate the inclusion of a random effect in the propensity score and outcome models for multilevel models
- (Non-collapsibility$\times$Causal Inference) and Spatial Confounding
- Should we advocate for the inclusion of a random effect in the propensity score model? and what about the outcome model?

$\Rightarrow$ If we have strong indication of potential for unmeasured confounders, and balancing diagnostics for observed confounders are not penalized, the answer might be yes for both

Perguntas?

Muito Obrigado!!!

Áreas de Interesse

- Bayesian theory: semi- and non-parametric methods
- Spatial statistics
- Spatio-temporal analysis
- Causal inference