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
Basics of Causal Inference

- Let $Y$ be an outcome of interest, $Z$ be a binary exposure, and $X$ be a set of covariates

- A dataset is a single realization of the joint process that generates the triple $(Y, Z, X)$

- Associational Inference $\implies$ Relationships among observed quantities

- Causal Inference $\implies$ Manipulation of the mechanism that generates the data regarding answering a causal question
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?

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 $Y(z)$ be the outcome that would be observed if we intervene to set $Z = z$ 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 $Y(z)$ be the outcome that would be observed if we intervene to set $Z = z$ 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?
Let $Y(z)$ be the outcome that would be observed if we intervene to set $Z = z$ 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)

\[
\tau = E(Y(1)) - E(Y(0))
\]

\[
= \int y f_{\varepsilon_1}(y) dy - \int y f_{\varepsilon_0}(y) dy
\]

\[
= \int y \frac{f_{\varepsilon_1}(y)}{f_O(y)} f_O(y) dy - \int y \frac{f_{\varepsilon_0}(y)}{f_O(y)} f_O(y) dy
\]

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

⇒ Infectious disease caused by the *Mycobacterium tuberculosis* bacteria

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

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 $Z_{ji} | \delta_{ji} \sim \text{Bernoulli}(\delta_{ji})$ with
  \[
  \text{logit}(\delta_{ji}) = \log \left( \frac{\delta_{ji}}{1 - \delta_{ji}} \right) = \sum_{k=1}^{q} \gamma_k X_{kji} + v_j, \tag{1}
  \]

- Binary outcome, such that $Y_{ji} | Z_{ji} \sim \text{Bernoulli}(\mu_{ji})$ 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{align*}
    Z_{ji} & = \alpha_0 + \alpha_X X_{ji} + T_j + \varepsilon_{ji}, \quad \varepsilon_{ji} \sim \mathcal{N}(0, \rho^2) \quad (3) \\
    Y_{ji} & = \beta_Z Z_{ji} + \beta_X X_{ji} + W_j + \varepsilon_{ji}, \quad \varepsilon_{ji} \sim \mathcal{N}(0, \kappa^2), \quad (4)
    \end{align*}
  
  - Let $T = (T_1, \ldots, T_m)\top$ and $W = (W_1, \ldots, W_m)\top$ be two cluster-level covariates whose joint distribution is given by
    
    \[
    \begin{pmatrix}
    T \\
    W
    \end{pmatrix}
    \sim \mathcal{N}
    \left[
    \begin{pmatrix}
    \mu_T 1_m \\
    \mu_W 1_m
    \end{pmatrix},
    \begin{pmatrix}
    \sigma_T^2 I_m & \rho_{T,W} \sigma_T \sigma_W I_m \\
    \rho_{T,W} \sigma_T \sigma_W I_m & \sigma_W^2 I_m
    \end{pmatrix}
    \right],
    \]
    
    where $\rho_{T,W}$ is the correlation between $T$ and $W$. The quantities $\rho$ in (3), $\kappa$ in (4), and $\sigma_T$ and $\sigma_W$ 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
    \[ Z_{ji} = \alpha_0 + \alpha_X X_{ji} + \nu_j + \epsilon_{ji}, \]
  - Outcome Model: for $Y_{ji}$ generated according to Equation (4), the following model was fitted
    \[ Y_{ji} = \beta_0 + \beta_Z Z_{ji} + B_{ji} + \eta_j + \epsilon_{ji}, \]
  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 $\hat{BS}$ and $\tilde{BS}$ are the balancing scores estimated from models described in column $E(Z|X)$. The cluster-level random effects $\mathbf{v}$ and $\mathbf{\eta}$ are such that $\mathbf{v} \sim \mathcal{N}(0, \sigma^2_v I_m)$ and $\mathbf{\eta} \sim \mathcal{N}(0, \sigma^2_w I_m)$.

| Model | $E(Z|X)$ | $E(Y|Z, X)$ |
|-------|----------|-------------|
| MD1   | $\alpha_0 1 + \alpha_X X$ | $\beta_0 1_N + \beta_Z Z + \beta_{b_{BS}}$ |
| MD2   | $\alpha_0 1 + \alpha_X X + A \mathbf{v}$ | $\beta_0 1_N + \beta_Z Z + \beta_{b_{BS}}$ |
| MD3   | $\alpha_0 1 + \alpha_X X$ | $\beta_0 1_N + \beta_Z Z + \beta_{b_{BS}} + A \mathbf{\eta}$ |
| MD4   | $\alpha_0 1 + \alpha_X X + A \mathbf{v}$ | $\beta_0 1_N + \beta_Z Z + \beta_{b_{BS}} + A \mathbf{\eta}$ |
Focusing on model MD2, it can be shown that

\[
\text{Bias}(\hat{\beta}_Z) = \left( (H^\top H)^{-1} H^\top \left( \beta X + \rho_{T,W} \sigma_T \sigma_W AA^\top \Sigma_{Z|X}^{-1}(Z - (\alpha_0 + \mu_T)1_N - \alpha_X X) \right) \right)_{(2)},
\]

where \( H = \left[ Z \mid \tilde{\mathcal{B}} \right] \) and \([\cdot]_{(2)}\) indicates the second element of the vector, and \( \Sigma_{Z|X} = \text{Var}(Z|X) \).
An Analytically Tractable Example: The Linear Case

Figure 2: Absolute bias of $\hat{\beta}_Z$ 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 Mechanism

- **Confounders** $X_1$ and $X_2$
  
  Scenario 1: $\nu_{ji} \sim \mathcal{N}(0, 0.1^2)$ and $\zeta_j \sim \mathcal{N}(0, 0.4^2)$;
  
  Scenario 2: $\nu_{ji} \sim \mathcal{N}(0, 0.25^2)$ and $\zeta_j \sim \mathcal{N}(0, 1)$.

- **Exposure** $Z_{ji} | \delta_{ji} \sim \text{Bernoulli}(\delta_{ji})$
  
  $\logit(\delta_{ji}) = \alpha_0 + X_{1ji} \alpha_1 + X_{2ji} \alpha_2 + T_j$.

- **Outcome** $Y_{ji} | \mu_{ji} \sim \text{Bernoulli}(\mu_{ji})$
  
  $\logit(\mu_{ji}) = \beta_0 + X_{1ji} \beta_1 + X_{2ji} \beta_2 + X_{1ji} X_{2ji} \beta_3 + W_j$. 
Model Adjustment

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

<table>
<thead>
<tr>
<th>Model</th>
<th>logit($\delta_{ji}$)</th>
<th>logit($\mu_{ji}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD1</td>
<td>$\gamma_0 + \gamma_1 X_{1ji} + \gamma_2 X_{2ji}$</td>
<td>$\beta + \beta_Z Z_{ji} + \beta_{bPS_{ji}}$</td>
</tr>
<tr>
<td>MD2</td>
<td>$\gamma_0 + \gamma_1 X_{1ji} + \gamma_2 X_{2ji} + \nu_j$</td>
<td>$\beta + \beta_Z Z_{ji} + \beta_{bPS_{ji}}$</td>
</tr>
<tr>
<td>MD3</td>
<td>$\gamma_0 + \gamma_1 X_{1ji} + \gamma_2 X_{2ji}$</td>
<td>$\beta + \beta_Z Z_{ji} + \beta_{bPS_{ji}} + \eta_j$</td>
</tr>
<tr>
<td>MD4</td>
<td>$\gamma_0 + \gamma_1 X_{1ji} + \gamma_2 X_{2ji} + \nu_j$</td>
<td>$\beta + \beta_Z Z_{ji} + \beta_{bPS_{ji}} + \eta_j$</td>
</tr>
</tbody>
</table>
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 PS1: $\text{logit}(\delta_{ji}) = \gamma_0 + \gamma_1 X_{1ji} + \gamma_2 X_{2ji}$ and PS2: $\text{logit}(\delta_{ji}) = \gamma_0 + \gamma_1 X_{1ji} + \gamma_2 X_{2ji} + \nu_j$, over 1000 Monte Carlo data replicates.
A Simulation Study with Binary Exposure and Outcome

Figure 4: Boxplots of the absolute bias (Panel A) and RMSE (Panel B) of the ATE for the models described in Table 2 over 1000 Monte Carlo replicates assuming binary exposure and outcome. The columns correspond to variations in the correlation between $T$ and $W$; rows correspond to the two scenarios of $X$ distributions in the true data generation mechanism. In the data generation model, $X_{kji} = \nu_{ji} + \zeta_j$, for $k = 1, 2$.

The labels 'X Scenario: 1' and 'X Scenario: 2' assume that $\nu_{ji} \sim N(0, 0.1^2)$ and $\zeta_j \sim N(0, 0.4^2)$ and $\nu_{ji} \sim N(0, 0.25^2)$ and $\zeta_j \sim N(0, 1^2)$, respectively.
- $Z_{ij}$ denotes a binary exposure that indicates if individual $i$ in the $j$th city received the DOT

- $Y_{ij}$ denotes the outcome of interest that indicates if individual $i$ in the $j$th city had a diagnosis of cure at the end of the treatment

- Let $\mathbf{X}_{ij} = (X_{1ij}, \ldots, X_{ pij})$ be a $p$-dimensional vector of predictors for individual $i$ in the $j$th city. The vector $\mathbf{X}_{ij}$ comprises both, individual and cluster characteristics
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.
We assume $Z_{ji} | \delta_{ji} \sim \text{Bernoulli}(\delta_{ji})$, 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**: $\nu_j \sim \mathcal{N}(0, \phi^2)$, for all $j$; and
- **PS3**: $\nu \sim \mathcal{N}(0, \phi^2 R(\lambda))$, where $R_{ij} = \text{Corr}(\nu_i, \nu_j) = \exp \left( -\lambda ||s_i - s_j|| \right)$, with $s_j$ denoting the centroid of city $j$ (a two-dimensional vector of coordinates) and $|| \cdot ||$ denoting the Euclidean distance.
Table 3: Exposure model comparison.

<table>
<thead>
<tr>
<th></th>
<th>elpd (WAIC)</th>
<th>pWAIC</th>
<th>WAIC</th>
<th>elpd (LOO)</th>
<th>pLOO</th>
<th>LOO</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS1</td>
<td>-8136.17</td>
<td>14.10</td>
<td>16272.35</td>
<td>-8136.23</td>
<td>14.16</td>
<td>16272.45</td>
</tr>
<tr>
<td>PS2</td>
<td>-6884.15</td>
<td>242.63</td>
<td>13768.30</td>
<td>-6893.51</td>
<td>251.99</td>
<td>13787.02</td>
</tr>
<tr>
<td>PS3</td>
<td>-6877.74</td>
<td>235.82</td>
<td>13755.49</td>
<td>-6885.21</td>
<td>243.28</td>
<td>13770.42</td>
</tr>
</tbody>
</table>
Figure 5: Standardized mean difference (SMD) for the observed baseline covariates between treated and control subjects.
<table>
<thead>
<tr>
<th>Model</th>
<th>Outcome Model</th>
<th>Distribution of the random effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>logit {μ_{ij}} = β + Z_{ij}β_Z</td>
<td>_</td>
</tr>
<tr>
<td>M2</td>
<td>logit {μ_{ij}} = β + Z_{ij}β_Z + η_j</td>
<td>η</td>
</tr>
<tr>
<td>M3</td>
<td>logit {μ_{ij}} = β + Z_{ij}β_Z + η_j</td>
<td>η</td>
</tr>
<tr>
<td>M4</td>
<td>logit {μ_{ij}} = β + Z_{ij}β_Z + \mathbf{x}_{ij}^Tβ_x</td>
<td>_</td>
</tr>
<tr>
<td>M5</td>
<td>logit {μ_{ij}} = β + Z_{ij}β_Z + \mathbf{x}_{ij}^Tβ_x + η_j</td>
<td>η</td>
</tr>
<tr>
<td>M6</td>
<td>logit {μ_{ij}} = β + Z_{ij}β_Z + \mathbf{x}_{ij}^Tβ_x + η_j</td>
<td>η</td>
</tr>
<tr>
<td>M7</td>
<td>logit {μ_{ij}} = β + Z_{ij}β_Z + PS_{1ij}β_{ps}</td>
<td>_</td>
</tr>
<tr>
<td>M8</td>
<td>logit {μ_{ij}} = β + Z_{ij}β_Z + PS_{1ij}β_{ps} + η_j</td>
<td>η</td>
</tr>
<tr>
<td>M9</td>
<td>logit {μ_{ij}} = β + Z_{ij}β_Z + PS_{1ij}β_{ps} + η_j</td>
<td>η</td>
</tr>
<tr>
<td>M10</td>
<td>logit {μ_{ij}} = β + Z_{ij}β_Z + PS_{2ij}β_{ps}</td>
<td>_</td>
</tr>
<tr>
<td>M11</td>
<td>logit {μ_{ij}} = β + Z_{ij}β_Z + PS_{2ij}β_{ps} + η_j</td>
<td>η</td>
</tr>
<tr>
<td>M12</td>
<td>logit {μ_{ij}} = β + Z_{ij}β_Z + PS_{2ij}β_{ps} + η_j</td>
<td>η</td>
</tr>
<tr>
<td>M13</td>
<td>logit {μ_{ij}} = β + Z_{ij}β_Z + PS_{3ij}β_{ps}</td>
<td>_</td>
</tr>
<tr>
<td>M14</td>
<td>logit {μ_{ij}} = β + Z_{ij}β_Z + PS_{3ij}β_{ps} + η_j</td>
<td>η</td>
</tr>
<tr>
<td>M15</td>
<td>logit {μ_{ij}} = β + Z_{ij}β_Z + PS_{3ij}β_{ps} + η_j</td>
<td>η</td>
</tr>
</tbody>
</table>
### Table 4: Outcome model comparison.

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

**Note:** The table compares different models (M1 to M15) using the elpd (WAIC) and elpd (LOO) metrics, along with their associated p-values (pWAIC and pLOO) and the respective WAIC and LOO values. The model with the lowest elpd (WAIC) and LOO values is generally considered the best. The entries with bold and underlined values indicate the most optimized model configurations for WAIC and LOO, respectively.
Figure 6: Posterior distributions of the ATE (Panel A) and Odds Ratio (Panel B) of the models above.
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?

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