Biometrics xx, 1–14 DOI: 10.1111/j.1541-0420.2005.00454.x
xxxxxx xxxx

Supporting Information for "Estimating treatment importance in multidrug-resistant tuberculosis using Targeted Learning: an observational individual patient data network meta-analysis"

Guanbo Wang[1], Mireille E. Schnitzer[2,3,*], Dick Menzies[4,5], Piret Viiklepp[6], Timothy H. Holtz[7], and Andrea Benedetti[1,4,5]

[1]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec H3A1A2, Canada

[2]Faculty of Pharmacy, Université de Montréal, Montréal, Québec H3C3J7, Canada

[3]Department of Social and Preventive Medicine, Université de Montréal, Montréal, Québec H3N1X9, Canada

[4]Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montréal, Québec H4A 3S5, Canada

[5]Department of Medicine, McGill University, Montréal, Québec H3H2R9, Canada

[6]National Institute for Health Development, Hiiu 42, Tallinn 11619, Estonia

[7]Division of Global HIV and TB, Centers for Disease Control and Prevention, Atlanta, GA, USA.

**email:* mireille.schnitzer@umontreal.ca

This paper has been submitted for consideration for publication in *Biometrics*

1 Web Appendix

A Intuition about the counterfactual notation

We use an example to further illustrate the counterfactual outcomes and treatments: Suppose that we have only three treatments, and study 1 investigates only treatments 1 and 2 (at least one patient in study 1 used treatment 1 and the same for treatment 2, but none took treatment 3). Subject 1 took treatment 1, and the observed outcome of this subject is 1. Then the counterfactual outcome of subject 1 under the availability of treatment 1 is equal to the outcome that we observed. Similarly, the counterfactual outcome under the availability of treatment 2 is equal to the observed outcome. We thus write $y_{11}\{d^{(1)} = 1\} = y_{11}\{d^{(2)} = 1\} = y_{11} = 1$ because both treatment 1 and 2 are available in study 1, which is expressed as $d_1^{(1)} = d_1^{(2)} = 1$. However, we do not know the value of the counterfactual outcome under the availability of treatment 3, $Y\{d^{(3)} = 1\}$, because treatment 3 is not available in this study (i.e. $d_1^{(3)} = 0$). Additionally, the counterfactual treatment of subject 1 under the availability of treatment 1 is identical to their observed treatment, denoted $a_{11}^{(1)}\{d^{(1)} = 1\} = a_{11}^{(1)} = 1$; similarly, for treatment 2, $a_{11}^{(2)}\{d^{(2)} = 1\} = a_{11}^{(2)} = 0$. But we do not know $A^{(3)}\{d^{(3)} = 1\}$, the value of the counterfactual treatment of subject 1 had treatment 3 been available in study 1.

B Non-parametric structural equation model and proof of identifiability of $\psi^{(k)}$

Following the main manuscript, we define the observed data structure for an arbitrary individual study as

$\mathbf{O} = [Y(1-C), \mathbf{V}, \mathbf{W}, C, \{A^{(k)}, R^{(k)}, D^{(k)}; k = 1, \ldots, 15\}]$. For convenience, we define $\mathbf{A} = \{A^{(k)}; k = 1, ..., 15\}$, and similarly for R and D. We also define $\mathbf{A}^{(k^*)} = \{A^{(l)}; \forall\ l \in (1, ..., 15)$ s.t. $l \neq k\}$ and similarly for R and D. We use lower case letters with indices ij or j to refer to observed realizations of individual patient variables and study-level variables,

respectively. For ease of notation, we define the adjustment set for the estimation of the variable importance of treatment k as $\mathbf{X}^{(k)} = \{\mathbf{V}, \mathbf{W}, \mathbf{A}^{(k^*)}\}$.

In order to define the model for the observed and counterfactual data, we follow an approach related to Schnitzer et al. (2018), section 1.3. Define the jth's study's observed data as $O_j^c = (O_{i,j}; i \in \mathcal{S}_j)$; we suppose these are identically and independently drawn $O^c \sim P_0^c$ where the probability distribution P_0^c is a member of some model space $\mathcal{M}^c$. We write the non-parametric structural equation model (NPSEM) as a conditional mixture with distribution P_0 belonging to model space $\mathcal{M}$.

In the NPSEM, we explicitly include two independent (possibly multivariate) study-level components $\mathbf{U} = (U_1, U_2)$ (that may be thought of as unmeasured study-level variables leading to random or study-specific effects and clustering) that allow for clustering of the data within studies beyond the measured $\mathbf{V}$. We write the data $\{\mathbf{O}, \mathbf{U}\}$ probability density function (pdf) as

$$\prod_{j=1}^{J} f_{U_1}(u_{1j}) f_{U_2}(u_{2j}) f_{\mathbf{V}}(\mathbf{v}_j \mid \mathbf{u}_j) f_{\mathbf{D}}(\mathbf{d}_j \mid \mathbf{v}_j, u_{2j}) \prod_{i \in \mathcal{S}_j} \Big\{ f_{\mathbf{W},\mathbf{R}}(\mathbf{w}_{ij}, \mathbf{r}_{ij} \mid \mathbf{u}_j, \mathbf{v}_j) \times$$
$$f_{\mathbf{A}}(\mathbf{a}_{ij} \mid \mathbf{u}_j, \mathbf{v}_j, \mathbf{d}_j, \mathbf{w}_{ij}, \mathbf{r}_{ij}) f_C(c_{ij} \mid u_{2j}, \mathbf{v}_j, \mathbf{w}_{ij}, \mathbf{a}_{ij}) f_Y(y_{ij} \mid u_{1j}, \mathbf{v}_j, \mathbf{w}_{ij}, \mathbf{r}_{ij}, \mathbf{a}_{ij}, c_{ij})^{(1-c_{ij})} \Big\}, \tag{1}$$

where J is the total number of studies and $\mathcal{S}_j$ is the set of indices of subjects in study j, and Y has a degenerate value when $C = 1$. In particular, we thought it reasonable that a patient's censoring status and outcome do not conditionally depend on treatment availability beyond the exposure received (though this could be relaxed). We assumed that censoring does not depend on resistance. Next, using standard definitions of conditional probability, we divide the relevant parts of the density into the components specific to the medication k

and other medications, denoted k^*.

$$\begin{aligned}
&\prod_{j=1}^{J} f_{U_1}(u_{1j}) f_{U_2}(u_{2j}) f_V(\mathbf{v}_j \mid \mathbf{u}_j) f_{\mathbf{D}^{(k^*)}}\{\mathbf{d}_j^{(k^*)} \mid u_{2j}, \mathbf{v}_j\} f_{D^{(k)}}\{d_j^{(k)} \mid u_{2j}, \mathbf{v}_j, \mathbf{d}_j^{(k^*)}\} \\
&\prod_{i \in \mathcal{S}_j} \Big[f_W(\mathbf{w}_{ij} \mid \mathbf{u}_j, \mathbf{v}_j) f_{\mathbf{R}}(\mathbf{r}_{ij} \mid \mathbf{u}_j, \mathbf{v}_j, \mathbf{w}_{ij}) f_{\mathbf{A}^{(k^*)}}\{\mathbf{a}_{ij}^{(k^*)} \mid \mathbf{u}_j, \mathbf{v}_j, \mathbf{d}_j^{(k^*)}, \mathbf{w}_{ij}, \mathbf{r}_{ij}^{(k^*)}\} \times \\
&f_{A^{(k)}}\{a_{ij}^{(k)} \mid \mathbf{u}_j, \mathbf{v}_j, d_j^{(k)}, \mathbf{w}_{ij}, r_{ij}^{(k)}, \mathbf{a}_{ij}^{(k^*)}\} f_C(c_{ij} \mid u_{2j}, \mathbf{v}_j, \mathbf{w}_{ij}, \mathbf{a}_{ij}) \times \\
&f_Y(y_{ij} \mid u_{1j}, \mathbf{v}_j, \mathbf{w}_{ij}, \mathbf{r}_{ij}, \mathbf{a}_{ij}, c_{ij})^{1-c_{ij}} \Big]
\end{aligned} \tag{2}$$

Here we have made the additional (again, possibly reasonable) assumption that the exposure $A^{(k)}$ is not affected by resistance nor to the availability of non-k medications beyond the actual exposure to non-k medications.

We now write the counterfactual pdf under the imposition of treatment availability $D^{(k)} = 1$ and observed outcomes $C = 0$. We define counterfactual variables $A^{(k)}\{d^{(k)} = 1\}$ and $Y\{d^{(k)} = 1\}$ and their corresponding conditional pdfs, $f_{A^{(k)}}^{(k)}$ and $f_Y^{(k)}$, respectively. Under this intervention, the counterfactual pdf is

$$\begin{aligned}
&\prod_{j=1}^{J} f_{U_1}(u_{1j}) f_{U_2}(u_{2j}) f_{\mathbf{V}}(\mathbf{v}_j \mid \mathbf{u}_j) f_{\mathbf{D}^{(k^*)}}\{\mathbf{d}_j^{(k^*)} \mid u_{2j}, \mathbf{v}_j\} \prod_{i \in \mathcal{S}_j} \Big(f_{\mathbf{W}}(\mathbf{w}_{ij} \mid \mathbf{u}_j, \mathbf{v}_j) \times \\
&f_{\mathbf{R}}(\mathbf{r}_{ij} \mid \mathbf{u}_j, \mathbf{v}_j, \mathbf{w}_{ij}) f_{\mathbf{A}^{(k^*)}}\{\mathbf{a}_{ij}^{(k^*)} \mid \mathbf{u}_j, \mathbf{v}_j, \mathbf{d}_j^{(k^*)}, \mathbf{w}_{ij}, \mathbf{r}_{ij}^{(k^*)}\} \times \\
&f_{A^{(k)}}^{(k)}\big[a_{ij}^{(k)}\{d^{(k)} = 1\} \mid \mathbf{u}_j, \mathbf{v}_j, \mathbf{w}_{ij}, r_{ij}^{(k)}, \mathbf{a}_{ij}^{(k^*)}\big] \times \\
&f_Y^{(k)}\big[y_{ij}\{d^{(k)} = 1\} \mid u_{1j}, \mathbf{v}_j, \mathbf{w}_{ij}, \mathbf{r}_{ij}, \mathbf{a}_{ij}^{(k^*)}, a_{ij}^{(k)}\{d^{(k)} = 1\}\big] \Big).
\end{aligned} \tag{3}$$

Under the counterfactual pdf for $Y\{d^{(k)} = 1\}$, the counterfactual parameter of interest is

$$\begin{aligned}
\psi^{(k)} = E\Big(& E\left[Y\{d^{(k)} = 1\} \mid \mathbf{X}^{(k)}, R^{(k)} = 0, A^{(k)}\{d^{(k)} = 1\} = 1\right] \\
& - E\left\{Y \mid \mathbf{X}^{(k)}, A^{(k)}, R^{(k)} = 0\right\} \mid R^{(k)} = 0 \Big),
\end{aligned}$$

recalling that $\mathbf{X}^{(k)} = \{\mathbf{V}, \mathbf{W}, \mathbf{A}^{(k^*)}\}$. Under the assumptions defined in section 2.2 of the main manuscript and the above pdfs, we rewrite our parameter of interest as:

$$\psi^{(k)} = E\big(E\big[Y\{d^{(k)} = 1\} \mid \mathbf{X}^{(k)}, R^{(k)} = 0, A^{(k)}\{d^{(k)} = 1\} = 1\big]$$

$$
\begin{aligned}
&- E\big\{Y \mid \mathbf{X}^{(k)}, R^{(k)}=0, A^{(k)}\big\} \mid R^{(k)}=0\big) \\
=&E\big(E\big[Y\{d^{(k)}=1\} \mid \mathbf{X}^{(k)}, R^{(k)}=0, A^{(k)}\{d^{(k)}=1\}=1, D^{(k)}=1\big] \\
&- E\big\{Y \mid \mathbf{X}^{(k)}, R^{(k)}=0, A^{(k)}\big\} \mid R^{(k)}=0\big) && \text{by A3(a)} \\
=&E\big[E\big\{Y \mid \mathbf{X}^{(k)}, R^{(k)}=0, A^{(k)}=1, D^{(k)}=1\big\} \\
&- E\big\{Y \mid \mathbf{X}^{(k)}, R^{(k)}=0, A^{(k)}\big\} \mid R^{(k)}=0\big] && \text{by A1} \\
=&E\big[E\big\{Y \mid \mathbf{X}^{(k)}, R^{(k)}=0, A^{(k)}=1\big\} - E\big\{Y \mid \mathbf{X}^{(k)}, R^{(k)}=0, A^{(k)}\big\} \mid R^{(k)}=0\big] \\
&\text{(since } A^{(k)}=1 \text{ implies } D^{(k)}=1\text{)} \\
=&E\big[E\big\{Y \mid \mathbf{X}^{(k)}, R^{(k)}=0, A^{(k)}=1, C=0\big\} \\
&- E\big\{Y \mid \mathbf{X}^{(k)}, R^{(k)}=0, A^{(k)}, C=0\big\} \mid R^{(k)}=0\big] && \text{by A3(b)}
\end{aligned}
$$

Estimation of $\psi^{(k)}$ by outcome modeling additionally requires that $E\{Y|\mathbf{X}^{(k)}, R^{(k)}=0, A^{(k)}, C=0\}$ can be modeled by merging the data across studies where treatment k is available. While we assumed that the form of the pdf for Y conditional on U_1 is common across studies, it is not given that the pdf conditional on only observed data is common. This sheds light on the potential difficulty of correctly modeling the outcome probabilities $Q^{(\tau k)}$ and $Q^{(\mu k)}$ in addition to the propensity score components that require models to be fit where treatment k is available and extrapolated to where k is not observed.

C Proof of Equation (2)

$$
\begin{aligned}
Q^{(\tau k)} &= Pr[Y\{d^{(k)}=1\}=1 \mid \mathbf{X}^{(k)}, R^{(k)}=0, A^{(k)}\{d^{(k)}=1\}=1] \\
&= Pr[Y\{d^{(k)}=1\}=1 \mid \mathbf{X}^{(k)}, R^{(k)}=0, D^{(k)}=1, C=0, A^{(k)}\{d^{(k)}=1\}=1] && \text{by A3(a)} \\
&= Pr\{Y=1 \mid \mathbf{X}^{(k)}, R^{(k)}=0, D^{(k)}=1, C=0, A^{(k)}=1\} && \text{by A1} \\
&= Pr\{Y=1 \mid \mathbf{X}^{(k)}, R^{(k)}=0, C=0, A^{(k)}=1\} \\
&\text{(since } A^{(k)}=1 \text{ implies } D^{(k)}=1\text{).}
\end{aligned}
$$

D Decomposition of $g_{ij}^{(\tau k)}$

$$
\begin{aligned}
g^{(\tau k)} =& Pr\{A^{(k)}=1, C=0 \mid \mathbf{X}^{(k)}, R^{(k)}=0\} \\
=& Pr\{A^{(k)}=1, D^{(k)}=1, C=0 \mid \mathbf{X}^{(k)}, R^{(k)}=0\} \\
=& Pr\{A^{(k)}=1 \mid D^{(k)}=1, \mathbf{X}^{(k)}, R^{(k)}=0\} Pr\{D^{(k)}=1 \mid \mathbf{X}^{(k)}, R^{(k)}=0\} \times \\
& Pr\{C=0 \mid A^{(k)}=1, D^{(k)}=1, \mathbf{X}^{(k)}, R^{(k)}=0\} \\
=& Pr\{A^{(k)}=1 | D^{(k)}=1, \mathbf{X}^{(k)}, R^{(k)}=0\} Pr\{D^{(k)}=1 | \mathbf{v}_j\} Pr\{C=0 | A^{(k)}=1, \mathbf{X}^{(k)}, R^{(k)}=0\} \\
=& g_1^{(\tau k)} \times g_2^{(\tau k)} \times g_3^{(\tau k)}.
\end{aligned}
$$

E Proof of IPTW consistency for $\tau^{(k)}$

First note that without any independence assumptions between variables A (binary) and Y, we have that

$$E\{Y \mathbb{1}(A=1)\} = E(Y \mid A=1)P(A=1).$$

By the law of large numbers, under consistent estimation of the propensity score, the IPTW estimator converges to the expectation

$$
\begin{aligned}
& E\left[Y \frac{\mathbb{1}\{A^{(k)}=1, C=0\}}{Pr\{A^{(k)}=1, C=0 \,\big|\, \mathbf{X}^{(k)}, R^{(k)}=0\}} \,\middle|\, R^{(k)}=0 \right] \\
=& E\left(E\left[\frac{Y \mathbb{1}\{A^{(k)}=1, C=0\}}{Pr\{A^{(k)}=1, C=0 \,\big|\, \mathbf{X}^{(k)}, R^{(k)}=0\}} \,\middle|\, \mathbf{X}^{(k)}, R^{(k)}=0 \right] \,\middle|\, R^{(k)}=0 \right), \\
& \text{(by the law of iterated expectations)}, \\
=& E\left[E\left\{ Y \,\middle|\, A^{(k)}=1, C=0, \mathbf{X}^{(k)}, R^{(k)}=0 \right\} \frac{P\left\{ A^{(k)}=1, C=0 \,\middle|\, \mathbf{X}^{(k)}, R^{(k)}=0 \right\}}{Pr\{A^{(k)}=1, C=0 \,\big|\, \mathbf{X}^{(k)}, R^{(k)}=0\}} \,\middle|\, R^{(k)}=0 \right], \\
& \text{(by the above identity and because the denominator term is a function of } \mathbf{X}^{(k)} \text{ and } R^{(k)}), \\
=& E\left[E\left\{ Y \,\middle|\, A^{(k)}=1, C=0, \mathbf{X}^{(k)}, R^{(k)}=0 \right\} \,\middle|\, R^{(k)}=0 \right], \\
=& \tau^{(k)}
\end{aligned}
$$

with the last step shown in Section B.

F Details of TMLE Algorithms and Variance Estimation

F.1 TMLE Algorithms

Define $Q^{(\tau k)} = Pr[Y\{d^{(k)} = 1\} = 1 \mid \mathbf{X}^{(k)}, R^{(k)} = 0, A^{(k)}\{d^{(k)} = 1\} = 1]$ and $g^{(\tau k)} = Pr\{A^{(k)} = 1, C = 0 \mid \mathbf{X}^{(k)}, R^{(k)} = 0\}$. After estimating these components as discussed in Section 3 in the main manuscript, we update $Q_{ij,n}^{(\tau k)}$ by allowing it to vary according to a parametric submodel with respect to the inverse propensity score. This submodel, which is parametrized by $\epsilon^{(\tau)}$ is

$$logit\left[Q^{(\tau k)^*}\{\epsilon^{(\tau)}\}\right] = logit\{Q^{(\tau k)}\} + \frac{\epsilon^{(\tau)}}{g^{(\tau k)}}. \tag{4}$$

The "size" of the update step ($\epsilon^{(\tau)}$) is selected such that the update minimizes the empirical expectation of a logistic regression loss function

$$\sum_{j=1}^{31} \sum_{i \in \mathcal{S}_j : r_{ij}^{(k)} = 0} a_{ij}^{(k)}(c_{ij} - 1)\left(y_{ij} log\left[Q_{ij,n}^{(\tau k)^*}\left\{\epsilon^{(\tau)}\right\}\right] + (1 - y_{ij}) log\left[1 - Q_{ij,n}^{(\tau k)^*}\left\{\epsilon^{(\tau)}\right\}\right]\right).$$

This is accomplished by fitting an intercept-free logistic regression of Y on the covariate $A^{(k)}/g^{(\tau k)}$ with offset $logit\{Q^{(\tau k)}\}$ in the subset of subjects whose infections were not known to be resistant to medication k, and with observed outcomes. We denote $\epsilon_n^{(\tau)}$ as the estimate of $\epsilon^{(\tau)}$ where $\epsilon_n^{(\tau)}$ is the estimated coefficient of the covariate in the regression. The estimate of $\tau^{(k)}$ is $\tau_{TMLE,n}^{(k)} = 1/n^{(k)} \sum_{j=1}^{31} \sum_{i \in \mathcal{S}_j : r_{ij}^{(k)} = 0} Q_{ij,n}^{(\tau k)^*}$.

For the estimation of $\mu^{(k)}$, the conditional probability of the outcome that we need to model is $Q^{(\mu k)} = Pr\{Y = 1 \mid \mathbf{X}^{(k)}, A^{(k)}, R^{(k)} = 0, C = 0\}$. Since we only need to correct for censoring, the propensity score is $g^{(\mu k)} = Pr\{C = 0 \mid \mathbf{X}^{(k)}, A^{(k)}, R^{(k)} = 0\}$. Following a similar approach, first we conduct an intercept-free logistic regression of Y on the covariate $1/g^{(\mu k)}$ with offset $logit\{Q^{(\mu k)}\}$ in the subset of subjects with an observed outcome. The coefficient is denoted $\epsilon^{(\mu)}$ and its estimate $\epsilon_n^{(\mu)}$. We then update $Q_{ij,n}^{(\mu k)}$ using $logit\left[Q_{ij,n}^{(\mu k)^*}\{\epsilon_n^{(\mu)}\}\right] = logit\{Q_{ij,n}^{(\mu k)}\} + \epsilon_n^{(\mu)}/g_{ij,n}^{(\mu k)}$. Then the estimate of $\mu^{(k)}$ is $\mu_{TMLE,n}^{(k)} = 1/n^{(k)} \sum_{j=1}^{31} \sum_{i \in \mathcal{S}_j : r_{ij}^{(k)} = 0} Q_{ij,n}^{(\mu k)^*}$.

Finally, we get the estimate of the parameter of interest: $\psi^{(k)}_{TMLE,n} = \tau^{(k)}_{TMLE,n} - \mu^{(k)}_{TMLE,n}$, where $\psi^{(k)}_{TMLE,n}$ is the substitution estimate of $\psi^{(k)}$.

F.2 Variance Estimation for TMLE

Clustering by study is taken into consideration in the estimation of the variance. We assume independence between studies but allow for the existence of clustering within studies, which may include clustering of baseline covariates, treatment assignment mechanisms, and study-specific effects by study. Ignoring clustering would cause bias in the estimation of the variance of the TMLE, especially when the cluster sizes (i.e. sample sizes in each study) are large and correlation between patients of the same study is high.

We estimate the variance of our estimator of $\psi^{(k)}$ using the efficient influence curve method (van der Laan and Rose, 2011). The influence curve $IC^{(k)}$ of the estimator is a function of $IC^{(\tau k)}$ and $IC^{(\mu k)}$, the influence curves of $\tau^{(k)}$ and $\mu^{(k)}$ respectively. The influence curves are,

$$IC^{(\tau k)} = \frac{\mathbb{1}\{R^{(k)} = 0, A^{(k)} = 1, C = 0\}}{g^{(\tau k)}}\{Y - Q^{(\tau k)}\} + Q^{(\tau k)} - \tau^{(k)}, \tag{5}$$

$$IC^{(\mu k)} = \frac{\mathbb{1}\{R^{(k)} = 0, C^{(k)} = 0\}}{g^{(\mu k)}}\{Y - Q^{(\mu k)}\} + Q^{(\mu k)} - \mu^{(k)}. \tag{6}$$

By the Delta method (Oehlert, 1992), we have that $IC^{(k)} = IC^{(\tau k)} - IC^{(\mu k)}$.

Let $IC^{(k)}_{ij}$ denote the value of the influence curve of $\psi^{(k)}$ evaluated at random vector $\mathbf{X}^{(k)}_{ij}$ for subjects $j = 1, ..., 31$ with $i \in \mathcal{S}_j$. The estimator's asymptotically linear form is given by

$$\sqrt{n^{(k)}}\{\psi^{(k)}_{TMLE,n} - \psi^{(k)}\} = \frac{1}{\sqrt{n^{(k)}}} \sum_{j=1}^{J} \sum_{i \in \mathcal{S}_j : r^{(k)}_{ij} = 0} IC^{(k)}_{ij} + o_p(1).$$

where $o_p(1)$ is a term that converges in probability to zero as the number of studies increases.

Since $E\{IC_{ij}^{(k)}\} = 0$, the variance of the estimator is (Schnitzer et al., 2018)

$$\{\sigma_{TMLE}^{(k)}\}^2 \approx Var\left\{\frac{1}{n^{(k)}}\sum_{j=1}^{J}\sum_{i\in\mathcal{S}_j:r_{ij}^{(k)}=0} IC_{ij}^{(k)}\right\}$$

$$= \frac{1}{\{n^{(k)}\}^2}\sum_{j=1}^{J}\left[\sum_{\substack{i,m\in\mathcal{S}_j:r_{ij}^{(k)}=0\\ \text{and } r_{mj}^{(k)}=0}} E\{IC_{ij}^{(k)}\cdot IC_{mj}^{(k)}\}\mathbb{1}(i\neq m) + E\{IC_{ij}^{(k)}\}^2\mathbb{1}(i=m)\right]$$

where J is the total number of studies.

The influence curve estimating equation in this case is:

$$\sum_{j=1}^{J}\sum_{i\in\mathcal{S}_j:r_{ij}^{(k)}=0} IC_{ij,n}^{(k)} = 0 \tag{7}$$

where $IC_{ij,n}^{(k)}$ is the difference $IC_{ij,n}^{(\tau k)} - IC_{ij,n}^{(\mu k)}$ of the empirical influence curves in equations (5) and (6), evaluated at the estimated values of the g and Q components at the realizations $\mathbf{x}_{ij}^{(k)}$.

The two TMLE update steps discussed above solve the efficient influence curve estimating equation (7). This occurs because the logistic regression update step solves the logistic regression score equations. This is done separately for the two components $\sum_{j=1}^{J}\sum_{i\in\mathcal{S}_j:r_{ij}^{(k)}=0} IC_{ij}^{(\tau k)} = 0$, and $\sum_{j=1}^{J}\sum_{i\in\mathcal{S}_j:r_{ij}^{(k)}=0} IC_{ij}^{(\mu k)} = 0$, resulting in the solution to equation (7).

G Simulation Study

This simulation study aims to: 1) demonstrate the consistency and double robustness of the estimator under increasingly complex settings, 2) investigate the appropriateness of the variance estimation and the coverage of the Wald-type confidence intervals based on the empirical influence curves, and 3) illustrate the potential importance of considering treatment availability (transportability) in our setting.

G.1 Methods

G.1.1 Data generation

We simulated data with structure and size similar to our real-life data. The sample size of each generated dataset is 9000, comprising 30 studies (clusters) with 300 individuals in each study. A total of three antimicrobial agents occur in the dataset but each study may only have access to one, two, or all three. All R code for the simulation studies are available on `https://doi.org/10.5281/zenodo.1405199`.

We generated three scenarios where Table 1 gives the full data generating mechanisms for each. For each study j, we generated two continuous study-level covariates V and U the latter considered to be unobserved. Note that all subjects in a study (cluster) are considered to have the same study-level variable value. The treatment availability $D^{(k)}$ for each treatment $k = 1, 2, 3$ was generated by study, completely at random in the first scenario and conditional on the realization of V in the second and third scenarios. Note that patients in the same study share the same access to treatments. For each patient i in study j, we then generated one individual level continuous covariate W conditional on the study-level covariate V and an indicator of resistance to antimicrobial agents $k = 1, 2, 3$, $R^{(k)}$ (generated completely at random). For each individual we independently generated three binary indicators of antimicrobial agent use $A^{(k)}, k = 1, 2, 3$ conditional on the values of V, W, and the corresponding $R^{(k)}$. Finally, we generated the binary outcomes, Y, conditional on both study-level covariates V, the individual-level covariate W, and treatment $A^{(k)}$ in such a way that if resistance is present, the treatment is less effective. In the third scenario, we included random effects, meaning that there were interactions between U and each $A^{(k)}$ in the outcome data generating model so that the treatment effects varied by study. Thus, the simulated data structure taking the individual as the unit of analysis is

$\mathbf{O} = [V, U, W, Y, \{A^{(k)}, R^{(k)}, D^{(k)}\}, k = 1...3]$. We denote $\mathbf{D} = \{D^{(1)}, D^{(2)}, D^{(3)}\}$ to be the vector indicating treatment availability, and do similarly for $\mathbf{A}$ and $\mathbf{R}$.

[Table 1 about here.]

G.1.2 Analysis

Our parameter of interest in the simulation study is:

$$\tau^{(1)} = E[E\{Y|\mathbf{X}^{(1)}, R^{(1)} = 0, A^{(1)} = 1\}|R^{(1)} = 0],$$

where $\mathbf{X}^{(1)} = \{V, W, A^{(2)}, A^{(3)}\}$. This corresponds to the first component of the treatment importance parameter of treatment 1, which is our focus for the purpose of the simulation study. In order to find the true value of $\tau^{(1)}$, we generated data as above, but with sample sizes greater than 10^7, and forced all $A^{(1)}$ equal to 1. Then we took the average of the generated y_{ij} within the subset of individuals whose infections were not resistant to medication 1.

To create the simulated observed data, 1000 random seeds were drawn and stored. Using these seeds, 1000 datasets were generated from these seeds and the analyses conducted on the datasets. For the analysis, we used logistic regressions to estimate $g_1^{(\tau 1)}$ in the first scenario. We took the proportion of $D^{(1)} = 1$ to be the estimate of $g_2^{(\tau 1)}$, since in this scenario, $D^{(1)}$ is independently generated. In the second and third scenario, we modeled $g_2^{(\tau 1)} = Pr\{D^{(1)}|V\}$ using a logistic regression. Since there was no censoring outcome involved in this simulation study, we had $g_3^{(\tau 1)} = 1$. Then the TMLE algorithm was applied to update the predicted values of $Q^{(\tau 1)}$.

In order to verify the double robustness property of our proposed estimator, we varied the model specifications used to estimate $g^{(\tau 1)} = g_1^{(\tau 1)} \cdot g_2^{(\tau 1)}$ and $Q^{(\tau 1)}$. In the first scenario, $Q^{(\tau 1)}$and $g^{(\tau 1)}$ were misspecified as null models (the outcome regressed on the intercept only). In the second and third scenario, the misspecification of $g^{(\tau 1)}$ implied that $g_2^{(\tau 1)}$ was also assigned a null model. In order to explore the performance of the proposed methods

with smaller sample sizes, we also applied the methods to data with 10 studies with 300 subjects each and 30 studies with 150 subjects each, respectively.

The standard error of the TMLE was estimated using the influence curve as discussed in Section F.2. For comparison, we also estimated the standard error using an influence curve sandwich estimator that ignores clustering, given by the sample mean of the square of the efficient influence curve. These two sandwich estimators were used to construct Wald-type 95% confidence intervals. The coverage rates (the percentage of times that the confidence intervals contained the true value) for both approaches were computed. We varied the sample sizes in order to observe changes in coverage rates under correct model specification.

Finally, we investigated the importance of considering transportability in the model fitting and prediction. We fit a standard TMLE ignoring treatment availability by study. We estimated $Pr\{Y = 1 \mid \mathbf{X}^{(1)} = \mathbf{x}_{ij}^{(1)}, R^{(1)} = 0, A^{(1)} = 1\}$ and the propensity score $Pr\{A^{(1)} = 1 \mid \mathbf{X} = \mathbf{x}_{ij}^{(1)}, R^{(1)} = 0\}$ directly (without decomposition, meaning that we did not model $Pr\{D^{(1)}\}$). We made predictions on the set where $r^{(1)} = 0$. When treatment availability is conditional on V (second and third scenarios), we expect that this approach will be biased.

G.2 Results

We found that the true value of our parameter of interest was 0.74 for the first two scenarios and 0.72 for the third. In Table 2 we present the average estimate, Monte-Carlo standard error (SE) and the SE estimated by the clustered and the non-clustered sandwich estimators for each of the three scenarios under various specifications of the Q and g models. The first scenario has only four specification combinations because the propensity score only has one component $g_1^{(\tau 1)}$. The notation Qv is used to indicate a correctly specified outcome model and gv denotes a correctly specified g model. Qx and gx respectively denote the incorrectly specified versions. In the second and third scenarios, where treatment availability $D^{(k)}$ was

dependent on study-level covariate V, we must also model $g_2^{(\tau 1)}$. Therefore, we have 8 different model specifications. We denote the correct specifications of the g components as $g1v$ and $g2v$ respectively and the incorrect versions as $g1x$ and $g2x$. By the double robustness property, the estimator should only be consistent when the model for $Q^{(\tau 1)}$ is correctly specified or when models for both $g_1^{(\tau 1)}$ and $g_2^{(\tau 1)}$ are correctly specified.

[Table 2 about here.]

The results in Table 2 confirm that there was no estimation bias when either model for $Q^{(\tau 1)}$ or $g^{(\tau 1)} = g_1^{(\tau 1)} \cdot g_2^{(\tau 1)}$ was correctly specified, which verified the double robustness property of our proposed estimator. In the last two scenarios, when the model for $Q^{(\tau 1)}$ was misspecified, the bias caused by the misspecification of $g_2^{(\tau 1)}$ was less than that caused by the misspecification of $g_1^{(\tau 1)}$. Unsurprisingly, the mean estimate when all models were misspecified diverged from the true value. At this sample size, the clustered sandwich estimator very slightly underestimated the standard error (compared to the Monte Carlo standard error) while the sandwich estimator that does not incorporate clustering greatly underestimated it.

Table 3 gives the results with 10 studies and 300 subjects in each study, while Table 4 gives the results with 30 studies and 150 subjects in each study. We again observe the double robustness of the estimator. The small sample sizes resulted in higher standard errors.

[Table 3 about here.]

[Table 4 about here.]

Table 5 gives the coverage rates of the 95% Wald-type confidence intervals calculated using the clustered and non-clustered sandwich estimators of the standard error for various study and within-study sample sizes. The coverage of the confidence intervals based on the non-clustered sandwich estimator was very low (14.6%-55.4%) in all scenarios and for all sample sizes. In contrast, the clustered sandwich estimator performed much better for all

sample sizes, with coverage increasing with the number of studies. With 10 studies, the clustered sandwich estimator produced coverage between 85.7%-92.4%. With 30 studies, we still see slight undercoverage for all within-study sample sizes (90.1%-92.7%)). With 60 studies, coverage was optimal in the first two simpler scenarios (94.5%-95.2%), though it still remained slightly below the optimal rate in scenario 3 (92.1%-93.4%). In contrast, the non-clustered sandwich estimator's coverage decreased as the number of subjects within studiess increased. Given that the clustered sandwich estimator is only valid for larger numbers of studies, an investigator may prefer to use the clustered bootstrap (resampling studies) in such a setting.

[Table 5 about here.]

Finally, we compare the estimates resulting from the TMLE with and without considering treatment availability in Table 6. When treatment availability is not considered, the results are biased in the last two scenarios where the treatment availability is dependent on study-level covariates.

[Table 6 about here.]

2 Web Figure

1 Details of the data inclusion and exclusion process

[Figure 1 about here.]

3 Web Tables

1 Descriptive statistics of covariates and outcome for the pooled individual patient data.

[Table 7 about here.]

2 Treatment importance, associated standard error and confidence interval for the 15 treatments without considering treatment availability.

[Table 8 about here.]

3 Comparison of main results (variable importance on the odds ratio scale) with conditional odds ratios from a logistic regression, adjusted for confounders and including all treatments as main terms in the same model.

[Table 9 about here.]

References

Oehlert, G. W. (1992). A note on the delta method. *The American Statistician* **46,** 27–29.

Schnitzer, M. E., van der Laan, M. J., Moodie, E. M., and Platt, R. W. (2018). Chapter 15: LTMLE with clustering. In van der Laan, M. J. and Rose, S., editors, *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies* (in press). Springer, New York.

van der Laan, M. J. and Rose, S. (2011). *Targeted learning: causal inference for observational and experimental data.* Springer Science & Business Media.

Received month year. *Revised month* year. *Accepted month* year.

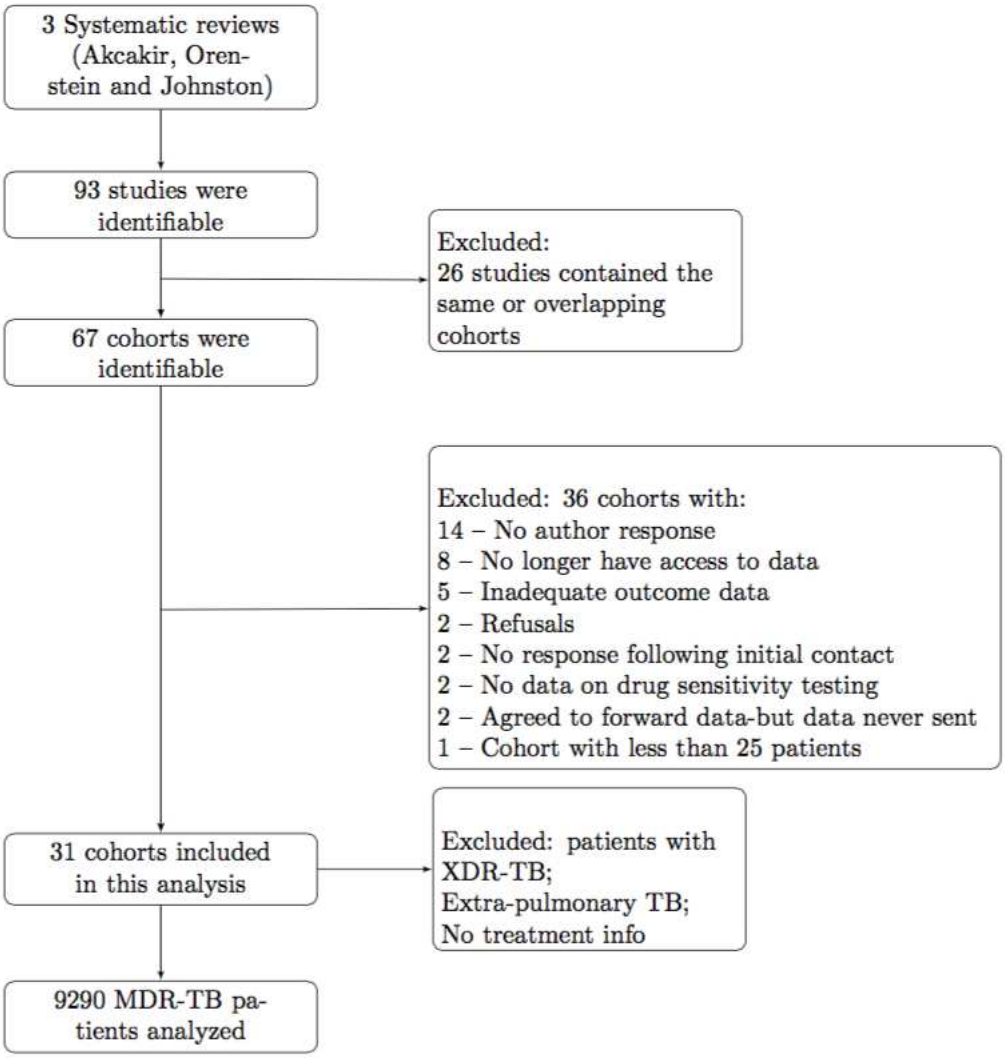

Figure 1: Data Collection.

Table 1: The data generating mechanism of the three scenarios.

Variable	**Generating Mechanism (i.i.d)**
V	$V \sim N(mean = 0.3, sd = 0.3, n = 30)$
U	$U \sim N(mean = 0.2V + 0.1, sd = 0.5, n = 30)$
$D^{(k)}$	**Scenario 1** Within the same study, generate a random number r=1,2, or 3. Randomly select r treatments and set $D^{(k)} = 1$ for these treatments. **Scenario 2 & 3** $D^{(1)} \sim Bin\{logit(p) = 1 + 2V, n = 30\}$ $D^{(2)} \sim Bin\{logit(p) = 0.5 + 1.5V, n = 30\}$ $D^{(3)} \sim Bin\{logit(p) = 1.5 + 0.3V, n = 30\}$ Within the same study impose restriction, $D^{(1)} + D^{(2)} + D^{(3)} > 0$.
For $j = 1, ..., 30$	
W	$W \sim N(mean = 0.1v_j, sd = 0.1, n = 300)$
$R^{(k)}$	$R^{(1)} \sim Ber(p = 0.25, n = 300)$ $R^{(2)} \sim Ber(p = 0.30, n = 300)$ $R^{(3)} \sim Ber(p = 0.25, n = 300)$
$A^{(k)}$	$A^{(1)} \sim Bin\{logit(p) = -0.75 + 2.4v_j + 1.8W - 0.1R^{(1)}, n = 300\}$ $A^{(2)} \sim Bin\{logit(p) = -1 + v_j + 1.7W - 0.15R^{(2)}, n = 300\}$ $A^{(3)} \sim Bin\{logit(p) = -1.5 + 1.7v_j + W - 0.16R^{(3)}, n = 300\}$
Y	**Scenario 1 & 2** $Y \sim Bin[logit(p) = -2 + 3.5v_j + 0.3W - 0.1u_j + 2.2A^{(1)}\{1 - R^{(1)}\} + 0.12A^{(2)}\{1 - R^{(2)}\} + 0.05A^{(3)}\{1 - R^{(3)}\}, n = 300]$ **Scenario 3** $Y \sim Bin[logit(p) = -2 + 3.5v_j + 0.3W - 0.1u_j + 2.2A^{(1)}\{1 - R^{(1)}\} + 0.12A^{(2)}\{1 - R^{(2)}\} + 0.05A^{(3)}\{1 - R^{(3)}\} + 4.5A^{(1)}u_j + 4.1A^{(2)}u_j + 4.15A^{(3)}u_j, n = 300]$

Table 2: Simulated estimation with different model specifications in each scenario (30 studies, 300 subjects in each study). In Model Specification, v indicates the corresponding model is correctly specified and x indicates the model is misspecified.

Model Specification	**Average Estimate**	**Monte Carlo SE**	**Average Clustered SE**	**Average Non-Clustered SE**
Scenario 1	True value: 0.74			
$Qv\ gv$	0.74	0.035	0.033	0.010
$Qv\ gx$	0.74	0.035	0.033	0.008
$Qx\ gv$	0.74	0.040	0.038	0.011
$Qx\ gx$	0.80	0.033	0.031	0.009
Scenario 2	True value: 0.74			
$Qv\ g1v\ g2v$	0.74	0.035	0.035	0.010
$Qv\ g1x\ g2v$	0.74	0.035	0.033	0.007
$Qv\ g1v\ g2x$	0.74	0.034	0.033	0.008
$Qv\ g1x\ g2x$	0.74	0.034	0.032	0.007
$Qx\ g1v\ g2v$	0.74	0.035	0.036	0.010
$Qx\ g1v\ g2x$	0.76	0.035	0.033	0.009
$Qx\ g1x\ g2v$	0.80	0.027	0.028	0.007
$Qx\ g1x\ g2x$	0.81	0.028	0.026	0.007
Scenario 3	True value: 0.72			
$Qv\ g1v\ g2v$	0.72	0.070	0.064	0.009
$Qv\ g1x\ g2v$	0.72	0.070	0.059	0.007
$Qv\ g1v\ g2x$	0.72	0.070	0.063	0.009
$Qv\ g1x\ g2x$	0.72	0.070	0.061	0.007
$Qx\ g1v\ g2v$	0.72	0.070	0.065	0.010
$Qx\ g1v\ g2x$	0.73	0.067	0.063	0.009
$Qx\ g1x\ g2v$	0.76	0.064	0.060	0.008
$Qx\ g1x\ g2x$	0.77	0.064	0.060	0.008

Table 3: Simulated estimation with different model specifications in each scenario (10 studies, 300 subjects in each study). In Model Specification, v indicates the corresponding model is correctly specified and x indicates the model is misspecified.

Model Specification	**Average Estimate**	**Monte Carlo SE**	**Average Clustered SE**	**Average Non-Clustered SE**
Scenario 1	True value: 0.74			
$Qv\ gv$	0.74	0.061	0.053	0.017
$Qv\ gx$	0.74	0.061	0.053	0.015
$Qx\ gv$	0.74	0.070	0.058	0.018
$Qx\ gx$	0.80	0.059	0.050	0.015
Scenario 2	True value: 0.74			
$Qv\ g1v\ g2v$	0.74	0.061	0.055	0.016
$Qv\ g1x\ g2v$	0.74	0.063	0.054	0.013
$Qv\ g1v\ g2x$	0.74	0.059	0.053	0.015
$Qv\ g1x\ g2x$	0.74	0.059	0.052	0.013
$Qx\ g1v\ g2v$	0.74	0.064	0.063	0.017
$Qx\ g1v\ g2x$	0.76	0.061	0.052	0.016
$Qx\ g1x\ g2v$	0.79	0.054	0.050	0.013
$Qx\ g1x\ g2x$	0.81	0.050	0.043	0.013
Scenario 3	True value: 0.72			
$Qv\ g1v\ g2v$	0.71	0.124	0.103	0.016
$Qv\ g1x\ g2v$	0.71	0.123	0.098	0.013
$Qv\ g1v\ g2x$	0.71	0.122	0.102	0.015
$Qv\ g1x\ g2x$	0.71	0.122	0.100	0.013
$Qx\ g1v\ g2v$	0.71	0.123	0.106	0.016
$Qx\ g1v\ g2x$	0.73	0.119	0.102	0.015
$Qx\ g1x\ g2v$	0.75	0.117	0.096	0.013
$Qx\ g1x\ g2x$	0.76	0.115	0.096	0.013

Table 4: Simulated estimation with different model specifications in each scenario (30 studies, 150 subjects in each study). In Model Specification, v indicates the corresponding model is correctly specified and x indicates the model is misspecified.

Model Specification	**Average Estimate**	**Monte Carlo SE**	**Average Clustered SE**	**Average Non-Clustered SE**
Scenario 1	True value: 0.74			
$Qv\ gv$	0.74	0.036	0.034	0.014
$Qv\ gx$	0.74	0.036	0.034	0.012
$Qx\ gv$	0.74	0.041	0.039	0.015
$Qx\ gx$	0.79	0.034	0.032	0.012
Scenario 2	True value: 0.74			
$Qv\ g1v\ g2v$	0.74	0.049	0.045	0.014
$Qv\ g1x\ g2v$	0.74	0.036	0.033	0.010
$Qv\ g1v\ g2x$	0.74	0.036	0.034	0.012
$Qv\ g1x\ g2x$	0.74	0.035	0.033	0.010
$Qx\ g1v\ g2v$	0.74	0.036	0.037	0.014
$Qx\ g1v\ g2x$	0.75	0.036	0.034	0.013
$Qx\ g1x\ g2v$	0.79	0.029	0.029	0.011
$Qx\ g1x\ g2x$	0.81	0.028	0.027	0.010
Scenario 3	True value: 0.72			
$Qv\ g1v\ g2v$	0.72	0.066	0.065	0.014
$Qv\ g1x\ g2v$	0.72	0.066	0.060	0.011
$Qv\ g1v\ g2x$	0.72	0.066	0.063	0.013
$Qv\ g1x\ g2x$	0.72	0.067	0.061	0.011
$Qx\ g1v\ g2v$	0.72	0.066	0.066	0.014
$Qx\ g1v\ g2x$	0.73	0.064	0.064	0.013
$Qx\ g1x\ g2v$	0.76	0.061	0.059	0.011
$Qx\ g1x\ g2x$	0.77	0.060	0.060	0.011

Table 5: Coverage rates with different sample sizes. "Cluster Coverage" denotes the coverage given by the clustered sandwich estimator and "Non-Clustered Coverage" is the coverage given by the sandwich estimator that ignores clustering. All of the models are correctly specified in each scenario.

Cluster Size	**No. of Subjects in Each Cluster**	**Clustered Coverage (%)**	**Non-Clustered Coverage (%)**
Scenario 1			
10	300	87.5	42.4
30	150	92.7	55.4
30	300	92.2	42.8
30	600	92.4	31.6
60	300	94.6	42.9
60	600	95.2	30.8
Scenario 2			
10	300	92.4	41.4
30	150	92.3	53.6
30	300	92.1	39.0
30	600	92.1	27.8
60	300	95.0	39.7
60	600	94.5	28.8
Scenario 3			
10	300	85.7	19.1
30	150	91.8	31.4
30	300	91.4	17.3
30	600	90.1	14.6
60	300	92.1	22.8
60	600	93.4	16.2

Table 6: Simulation study results with & without considering treatment availability in Q and g models (30 clusters, 300 subjects in each cluster). "With TA" indicates the proposed estimator that considers treatment availability and "Without TA" indicates the standard TMLE applied to the subset of available studies. All of the models are correctly specified in each scenario.

Model **Specification**	**Average Estimate**	**Monte Carlo SE**	**Average Clustered SE**	**Average Non-Clustered SE**
Scenario 1	True value: 0.74			
With TA	0.74	0.035	0.033	0.010
Without TA	0.74	0.042	0.041	0.016
Scenario 2	True value: 0.74			
With TA	0.74	0.035	0.035	0.010
Without TA	0.76	0.036	0.035	0.012
Scenario 3	True value: 0.72			
With TA	0.72	0.070	0.064	0.009
Without TA	0.73	0.068	0.078	0.012

Table 7: Descriptive statistics of covariates and outcome for the pooled individual patient data in the MDR-TB application. [a]IQR: inter-quartile range.

Covariates	**Summary**		**Missing N(%)**
Year of Study	Median	2004	NA
	IQR[a]	(2002,2004)	
Age	Median	38	28 (0.3)
	IQR	(29,48)	
Income Group[b] N (%)	Lower middle	404(4.4)	NA
	Upper middle	3106 (33.4)	NA
	High	5780 (62.2)	NA
Sex N(%)	Male	2979 (32.1)	6 (0.06)
	Female	6305 (67.9)	
Positive HIV N (%)		1193 (12.8)	1369 (14.7)
Positive smear N (%)		5836 (62.8)	1439 (15.5)
Positive past TB N (%)		6489 (69.8)	524 (5.6)
Positive cavitation N (%)		6489 (69.8)	2521 (27.1)
Treatment success N (%)		4847 (52.2)	260 (2.8)

Table 8: Treatment importance, associated standard error and confidence intervals for the 15 treatments *without considering treatment availability* in the MDR-TB application.

Treatment	**Estimate**	**Standard Error**	**95% Confidence Interval**
High-generation quinolones	0.073	0.099	(-0.121, 0.267)
Cycloserine	0.057	0.029	(0.000, 0.115)
Streptomycin	0.050	0.021	(0.009, 0.092)
Para-aminosalicylic acid	0.035	0.019	(-0.001, 0.072)
Ciprofloxacin	0.032	0.017	(-0.001, 0.066)
Ethionamide	0.029	0.013	(0.004, 0.055)
Amikacin	0.026	0.016	(-0.005, 0.057)
Ethambutol	0.023	0.021	(-0.018, 0.064)
Kanamycin	0.023	0.026	(-0.028, 0.074)
Ofloxacin	0.021	0.033	(-0.04, 0.085)
Prothionamide	0.004	0.194	(-0.376, 0.384)
Capreomycin	0.004	0.017	(-0.029, 0.036)
Group 5 level drugs	-0.004	0.017	(-0.037, 0.029)
Pyrazinamide	-0.004	0.016	(-0.037, 0.028)
Rifabutin	-0.061	0.015	(-0.090, -0.033)

Table 9: Comparison of main results (variable importance on the odds ratio scale) with conditional odds ratios from a logistic regression, adjusted for confounders and including all treatments as main terms in the same model. While parameters are not directly comparable, we can compare scientific conclusions from both analyses. *: significant hypothesis test.

Treatment	**Main results**	**95% CI**	**Regression**	**95%CI**
Ethambutol	1.086	(0.985, 1.197)	0.361	(0.062, 2.091)
Amikacin	1.367	(0.943, 1.984)	1.552	(0.858, 2.807)
Capreomycin	0.915	(0.673, 1.243)	0.644	(0.342, 1.211)
Ciprofloxacin	1.721*	(1.092, 2.713)	2.087*	(1.192, 3.652)
Cycloserine	1.251	(0.975, 1.605)	1.499	(0.230, 9.761)
Ethionamide	1.325*	(1.018, 1.726)	0.607	(0.250, 1.478)
Ofloxacin	1.111	(0.920, 1.342)	0.087*	(0.014, 0.553)
Para-aminosalicylic acid	1.030	(0.901, 1.177)	2.737	(0.433, 17.297)
Prothionamide	1.040	(0.353, 3.068)	0.406*	(0.166, 0.991)
Rifabutin	1.103	(0.605, 2.001)	0.576	(0.281, 1.182)
Streptomycin	1.303*	(1.043, 1.627)	0.664	(0.276, 1.602)
Pyrazinamide	0.978	(0.889, 1.074)	0.597	(0.032, 11.020)
Kanamycin	1.129	(0.998, 1.279)	17.781*	(1.458, 216.876)
High-generation quinolones	1.351	(0.636, 2.883)	0.645	(0.359, 1.159)
Group 5 level drugs	0.974	(0.724, 1.309)	1.156	(0.562, 2.381)