Supporting Information for "Modeling Treatment Effect Modification in Multidrug-Resistant Tuberculosis in an Individual Patient Data Meta-Analysis"

Yan Liu[1], Mireille E. Schnitzer[2,3,*], Guanbo Wang[1], Edward Kennedy[4], Piret Viiklepp[5], Mario H. Vargas[6], Giovanni Sotgiu[7], Dick Menzies[8,9] and Andrea Benedetti[1,8,10]

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

[2] *Faculty of Pharmacy, Université de Montréal, Canada*

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

[4] *Department of Statistics & Data Science, Carnegie Mellon University, USA*

[5] *National Institute for Health Development, Estonia*

[6] *Instituto Nacional de Enfermedades Respiratorias, Mexico*

[7] *Clinical Epidemiology and Medical Statistics Unit, Department of Medical, Surgical and Experimental Sciences, University of Sassari, Sassari, Italy*

[8] *Respiratory Epidemiology and Clinical Research Unit, Centre for Outcomes Research & Evaluation, Research Institute of the McGill University Health Centre, Canada*

[9] *Montréal Chest Institute & McGill International TB Centre, Research Institute of the McGill University Health Centre, Canada*

[10] *Department of Medicine, McGill University, Canada*

Appendix A: Non-parametric structural equation model

Suppose we observe J i.i.d. copies of $\boldsymbol{O} = (\boldsymbol{S}, \boldsymbol{D}, \boldsymbol{W}, \boldsymbol{R}, \boldsymbol{A}, Y)$ with true distribution P_0 which belongs to some model space ω. For a given treatment $k, for\ k \in \{1, 2, \cdots, 14\}$, the hierarchical non-parametric structural equation model (NPSEM) that we assume can be written as:

$$\begin{aligned}
\boldsymbol{S} &= g_{\boldsymbol{S}}(\varepsilon_{\boldsymbol{S}}) \\
D^{(k)} &= g_D(\boldsymbol{S}, \varepsilon_D) \\
\{\boldsymbol{W}, \boldsymbol{R}\} &= g_{\boldsymbol{W},\boldsymbol{R}}(\boldsymbol{S}, \varepsilon_{\boldsymbol{W},\boldsymbol{R}}) \\
A^{(k)} &= g_A(\boldsymbol{S}, \boldsymbol{W}, R^{(k)}, \varepsilon_A) \cdot D^{(k)} \\
Y &= g_Y(A^{(k)}, \boldsymbol{S}, \boldsymbol{W}, \boldsymbol{R}, \varepsilon_Y)
\end{aligned}$$

Specifically, we start from the study-level random variables $\boldsymbol{S}$ where the values for observations in the same study are equal. The treatment availability D for each study is generated conditioning on $\boldsymbol{S}$. We assume the individual-level variable $\boldsymbol{W}$ and antimicrobial resistance indicator $\boldsymbol{R}$ are jointly distributed based on $\boldsymbol{S}$. The treatment assignment A of each subject is jointly distributed conditioning on $\boldsymbol{S}, \boldsymbol{W}, \boldsymbol{R}$. Note that $fg_{A^{(k)}|A^{(k^*)}}$ is degenerate with all mass at zero when $D^{(k)} = 0$. Finally, we can generate the outcome on the basis of all the variables except the treatment availability. We also assume that all the errors $\varepsilon_{\boldsymbol{S}}, \varepsilon_D, \varepsilon_{\boldsymbol{W},\boldsymbol{R}}, \varepsilon_A$ are independent. In this NPSEM model, we assume there are no unmeasured covariates affecting both the outcome and the treatment assignment and/or availability.

We use lower case letters with indices ij or j to represent the observed realizations of individual patient variables and study-level variables. Let C_j be the set of indices of subjects in study j. Then, we can write the data probability density function (pdf) of observed data as

$$\prod_{j=1}^{J} f_{\boldsymbol{D}}(\boldsymbol{d}_j|\boldsymbol{s}_j) f_{\boldsymbol{S}}(\boldsymbol{s}_j) \prod_{i \in C_j} \left\{ f_{\boldsymbol{W},\boldsymbol{R}}(\boldsymbol{w}_{ij}, \boldsymbol{r}_{ij}|\boldsymbol{s}_j) f_{\boldsymbol{A}}(\boldsymbol{a}_{ij}|\boldsymbol{s}_j, \boldsymbol{w}_{ij}, \boldsymbol{r}_{ij}, \boldsymbol{d}_j) f_Y(y_{ij}|\boldsymbol{a}_{ij}, \boldsymbol{s}_j, \boldsymbol{w}_{ij}, \boldsymbol{r}_{ij}) \right\}$$

Then we separate the relevant components of the density into the parts specific to the treatment k and other treatments k^*. And we assume that the distribution of treatment k is not affected by other treatment availability and resistance. We define $f_{\boldsymbol{A}^{(k^*)}}$ as the counterfactual pdf of treatment k^* such that $k^* \neq k$ under the intervention $\boldsymbol{d}^{(k^*)} = 1$. Then

the pdf can be written as

$$\prod_{j=1}^{J} f_{D^{(k)}}(d_j^{(k)}|\boldsymbol{s}_j) f_{\boldsymbol{D}^{(k^*)}}(\boldsymbol{d}_j^{(k^*)}|\boldsymbol{s}_j) f_{\boldsymbol{S}}(\boldsymbol{s}_j) \prod_{i\in C_j} \Big\{ f_{\boldsymbol{W},\boldsymbol{R}}(\boldsymbol{w}_{ij}, \boldsymbol{r}_{ij}|\boldsymbol{s}_j)$$
$$f_{\boldsymbol{A}^{(k^*)}}(\boldsymbol{a}_{ij}^{(k^*)}|\boldsymbol{s}_j, \boldsymbol{w}_{ij}, \boldsymbol{r}_{ij}^{(k^*)}, \boldsymbol{d}_j^{(k^*)}) f_{A^{(k)}}(a_{ij}^{(k)}|\boldsymbol{s}_j, \boldsymbol{w}_{ij}, r_{ij}^{(k)}, d_j^{(k)}, \boldsymbol{a}_{ij}^{(k^*)}) f_Y(y_{ij}|\boldsymbol{a}_{ij}, \boldsymbol{s}_j, \boldsymbol{w}_{ij}, \boldsymbol{r}_{ij}) \Big\}$$

We define $f_Y^{(k)}$ as the counterfactual pdf of outcome given treatment k. We assume the counterfactual pdf under the intervention settings $d^{(k)} = 1$ and $a^{(k)} = 1$. Note that $a^{(k)} = 1$ implies $d^{(k)} = 1$. Then the pdf can be written as

$$\begin{aligned}
&\prod_{j=1}^{J} f_{\boldsymbol{D}^{(k^*)}}(\boldsymbol{d}_j^{(k^*)}|\boldsymbol{s}_j) f_{\boldsymbol{S}}(\boldsymbol{s}_j) \prod_{i\in C_j} \Big\{ f_{\boldsymbol{W},\boldsymbol{R}}(\boldsymbol{w}_{ij}, \boldsymbol{r}_{ij}|\boldsymbol{s}_j) \\
&f_{\boldsymbol{A}^{(k^*)}}(\boldsymbol{a}_{ij}^{(k^*)}|\boldsymbol{s}_j, \boldsymbol{w}_{ij}, \boldsymbol{r}_{ij}^{(k^*)}, \boldsymbol{d}_j^{(k^*)}) f_Y^{(k)}(y_{ij}(d_j^{(k)} = 1, a_{ij}^{(k)} = 1)|\boldsymbol{a}_{ij}^{(k^*)}, \boldsymbol{s}_j, \boldsymbol{w}_{ij}, \boldsymbol{r}_{ij}) \Big\} \\
=&\prod_{j=1}^{J} f_{\boldsymbol{D}^{(k^*)}}(\boldsymbol{d}_j^{(k^*)}|\boldsymbol{s}_j) f_{\boldsymbol{S}}(\boldsymbol{s}_j) \prod_{i\in C_j} \Big\{ f_{\boldsymbol{W},\boldsymbol{R}}(\boldsymbol{w}_{ij}, \boldsymbol{r}_{ij}|\boldsymbol{s}_j) \\
&f_{\boldsymbol{A}^{(k^*)}}(\boldsymbol{a}_{ij}^{(k^*)}|\boldsymbol{s}_j, \boldsymbol{w}_{ij}, \boldsymbol{r}_{ij}^{(k^*)}, \boldsymbol{d}_j^{(k^*)}) f_Y^{(k)}(y_{ij}(a_{ij}^{(k)} = 1)|\boldsymbol{a}_{ij}^{(k^*)}, \boldsymbol{s}_j, \boldsymbol{w}_{ij}, \boldsymbol{r}_{ij}) \Big\}
\end{aligned}$$

Appendix B: Proof of identifiability of $\psi(\boldsymbol{V}^{(k)})$

We aim to estimate the parameters of interest on a global population based on the observed data. Under the assumptions defined in Section 2.2 in the main manuscript, we can rewrite the CATE as:

$$
\begin{aligned}
&\psi\{\boldsymbol{V}^{(k)}\} \\
=&\mathbb{E}[Y\{a^{(k)}=1\}|\boldsymbol{V}^{(k)}] - E[Y\{a^{(k)}=0\}|\boldsymbol{V}^{(k)}] \\
=&\mathbb{E}(\mathbb{E}[Y\{a^{(k)}=1\}|\boldsymbol{X}^{(k)}]|\boldsymbol{V}^{(k)}) - E(E[Y\{a^{(k)}=0\}|\boldsymbol{X}^{(k)}]|\boldsymbol{V}^{(k)}) \\
=&\mathbb{E}(\mathbb{E}[Y\{a^{(k)}=1,d^{(k)}=1\}|\boldsymbol{X}^{(k)}]|\boldsymbol{V}^{(k)}) - \mathbb{E}(\mathbb{E}[Y\{a^{(k)}=0\}|\boldsymbol{X}^{(k)}]|\boldsymbol{V}^{(k)}) \\
=&\mathbb{E}(\mathbb{E}[Y\{a^{(k)}=1,d^{(k)}=1\}|\boldsymbol{X}^{(k)},D^{(k)}=1]|\boldsymbol{V}^{(k)}) \\
&-\mathbb{E}(\mathbb{E}[Y\{a^{(k)}=0\}|\boldsymbol{X}^{(k)}]|\boldsymbol{V}^{(k)}) && \textit{by assumption}(c) \\
=&\mathbb{E}(\mathbb{E}[Y\{a^{(k)}=1,d^{(k)}=1\}|A^{(k)}=1,\boldsymbol{X}^{(k)},D^{(k)}=1]|\boldsymbol{V}^{(k)}) \\
&-\mathbb{E}(\mathbb{E}[Y\{a^{(k)}=0\}|\boldsymbol{X}^{(k)}]|\boldsymbol{V}^{(k)}) && \textit{by assumption}(d) \\
=&\mathbb{E}[\mathbb{E}\{Y|A^{(k)}=1,\boldsymbol{X}^{(k)},D^{(k)}=1\}|\boldsymbol{V}^{(k)}] - \mathbb{E}[\mathbb{E}\{Y|A^{(k)}=0,\boldsymbol{X}^{(k)}\}|\boldsymbol{V}^{(k)}] && \textit{by assumption}(a) \\
=&\mathbb{E}[\mathbb{E}\{Y|A^{(k)}=1,\boldsymbol{X}^{(k)}\}|\boldsymbol{V}^{(k)}] - \mathbb{E}[\mathbb{E}\{Y|A^{(k)}=0,\boldsymbol{X}^{(k)}\}|\boldsymbol{V}^{(k)}]
\end{aligned}
$$

Both the two terms can be estimated from the observed data, so the CATE is identifiable.

Appendix C: Efficient Influence Function for $\boldsymbol{\beta_V}$

We derive the EIF for the coefficients $\boldsymbol{\beta_V}$ in a working i.i.d. setting. Depending on the geometry of influence function, the EIF is given by $\prod(\mathcal{D}|T_Q)$ where $\mathcal{D}$ is an arbitrary influence function, T_Q is the tangent space which is tangential to model space $\mathcal{M}$ and has the same dimension with $\mathcal{M}$. $\prod(\mathcal{D}|T_Q)$ refers to the projection of $\mathcal{D}$ onto space T_Q [5]. Here, we use the score function for IPTW as $\mathcal{D}_{IPTW}$ which is acturally an EIF up to a normalizing matrix and decompose the tangent space T_Q to the direct sum of T_Y and $T_{\boldsymbol{X}}$. As discussed by Tsiatis [5], we have

$$
\begin{aligned}
\prod(\mathcal{D}_{IPTW}|T_Q) =& \prod(\mathcal{D}_{IPTW}|T_Y \oplus T_{\boldsymbol{X}}) \\
=& \underbrace{\mathbb{E}(\mathcal{D}_{IPTW}|\boldsymbol{X},A,Y)}_{(i)} - \underbrace{\mathbb{E}(\mathcal{D}_{IPTW}|\boldsymbol{X},A)}_{(ii)} + \underbrace{\mathbb{E}(\mathcal{D}_{IPTW}|\boldsymbol{X})}_{(iii)}
\end{aligned} \tag{1}
$$

where

$$
\mathcal{D}_{IPTW} = \left[\left\{\frac{I_{A=1}}{g(1|\boldsymbol{X})} - \frac{I_{A=0}}{g(0|\boldsymbol{X})}\right\}Y - \boldsymbol{V}^{\intercal}\boldsymbol{\beta_V}\right]\boldsymbol{V}
$$

Note that $\boldsymbol{V}$ is a $(p+1)$-dimensional random covariate vector and the symbol $^{\intercal}$ indicates a transpose.

Obviously, for the term (i), $\mathbb{E}(\mathcal{D}_{IPTW}|\boldsymbol{X}, A, Y) = \mathcal{D}_{IPTW}$. And the term (ii),

$$\mathbb{E}(\mathcal{D}_{IPTW}|\boldsymbol{X}, A) = \left[\left\{\frac{I_{A=1}}{g(1|\boldsymbol{X})} - \frac{I_{A=0}}{g(0|\boldsymbol{X})}\right\}\mathbb{E}(Y|\boldsymbol{X}, A) - \boldsymbol{V}^\intercal\boldsymbol{\beta}_{\boldsymbol{V}}\right]\boldsymbol{V}$$

Then for the term (iii),

$$\begin{aligned}
\mathbb{E}(\mathcal{D}_{IPTW}|\boldsymbol{X}) =& \mathbb{E}\left(\left[\left\{\frac{I_{A=1}}{g(1|\boldsymbol{X})} - \frac{I_{A=0}}{g(0|\boldsymbol{X})}\right\}Y - \boldsymbol{V}^\intercal\boldsymbol{\beta}_{\boldsymbol{V}}\right]\boldsymbol{V}\Big|\boldsymbol{X}\right) \\
=& \mathbb{E}\left(\left[\left\{\frac{I_{A=1}}{g(1|\boldsymbol{X})} - \frac{I_{A=0}}{g(0|\boldsymbol{X})}\right\}Y\Big|\boldsymbol{X}\right] - \boldsymbol{V}^\intercal\boldsymbol{\beta}_{\boldsymbol{V}}\right)\boldsymbol{V} \\
=& \left[\left\{\frac{\mathbb{E}(Y, I_{A=1}|\boldsymbol{X})}{g(1|\boldsymbol{X})} - \frac{\mathbb{E}(Y, I_{A=0}|\boldsymbol{X})}{g(0|\boldsymbol{X})}\right\} - \boldsymbol{V}^\intercal\boldsymbol{\beta}_{\boldsymbol{V}}\right]\boldsymbol{V} \\
=& \left\{\mathbb{E}(Y|A=1, \boldsymbol{X}) - \mathbb{E}(Y|A=0, \boldsymbol{X}) - \boldsymbol{V}^\intercal\boldsymbol{\beta}_{\boldsymbol{V}}\right\}\boldsymbol{V}
\end{aligned}$$

Finally, according to equation (2), we can combine all three terms as

$$\begin{aligned}
\mathcal{D}(Y, A, \boldsymbol{X}, \boldsymbol{V}) =& \prod(\mathcal{D}_{IPTW}|T_Q) = \mathbb{E}(\mathcal{D}_{IPTW}|\boldsymbol{X}, A, Y) - \mathbb{E}(\mathcal{D}_{IPTW}|\boldsymbol{X}, A) + \mathbb{E}(\mathcal{D}_{IPTW}|\boldsymbol{X}) \\
=& \Bigg[\mathbb{E}(Y|A=1, \boldsymbol{X}) - \mathbb{E}(Y|A=0, \boldsymbol{X}) + \\
& \left\{\frac{I_{A=1}}{g(1|\boldsymbol{X})} - \frac{I_{A=0}}{g(0|\boldsymbol{X})}\right\}\{Y - \mathbb{E}(Y|\boldsymbol{X}, A)\} - \boldsymbol{V}^\intercal\boldsymbol{\beta}_{\boldsymbol{V}}\Bigg]\boldsymbol{V} \\
=& \left[\left\{\frac{I_{A=1}}{g(1|\boldsymbol{X})} - \frac{I_{A=0}}{g(0|\boldsymbol{X})}\right\}\{Y - \overline{Q}(A, \boldsymbol{X}\} + \overline{Q}(1, \boldsymbol{X}) - \overline{Q}(0, \boldsymbol{X}) - \boldsymbol{V}^\intercal\boldsymbol{\beta}_{\boldsymbol{V}}\right]\boldsymbol{V}
\end{aligned}$$

$\mathcal{D}(Y, A, \boldsymbol{X}, \boldsymbol{V})$ is a $(p+1)$-dimensional random vector. We define a normalizing matrix M

$$M = -\mathbb{E}\left\{\frac{\partial}{\partial\boldsymbol{\beta}_{\boldsymbol{V}}}\mathcal{D}(Y, A, \boldsymbol{X}, \boldsymbol{V})\right\} = \mathbb{E}(\boldsymbol{V}^\intercal\boldsymbol{V})$$

The dimension of M is $(p+1)\times(p+1)$. The efficient influence function for coefficients $\boldsymbol{\beta}_{\boldsymbol{V}}$ is $\mathcal{D}_{\boldsymbol{\beta}_{\boldsymbol{V}}} = M^{-1}\mathcal{D}(Y, A, \boldsymbol{X}, \boldsymbol{V})$, i.e.

$$\mathcal{D}_{\boldsymbol{\beta}_{\boldsymbol{V}}} = M^{-1}\left[\left\{\frac{I_{A=1}}{g(1|\boldsymbol{X})} - \frac{I_{A=0}}{g(0|\boldsymbol{X})}\right\}\{Y - \overline{Q}(A, \boldsymbol{X})\} + \overline{Q}(1, \boldsymbol{X}) - \overline{Q}(0, \boldsymbol{X}) - \boldsymbol{V}^\intercal\boldsymbol{\beta}_{\boldsymbol{V}}\right]\boldsymbol{V}$$

An equivalent result was also given by Rosenblum and van der Laan [1].

Appendix D: Augmented inverse probability of treatment weighted estimator (A-IPTW)

In simple single-treatment single-dataset settings, A-IPTW involves a transformation of the outcomes using two components: 1) an outcome regression conditional on treatment and covariates, and 2) weights comprising of the inverse of the propensity score, where the propensity is the probability of treatment conditional on covariates [2]. A-IPTW allows for doubly-robust and locally efficient estimation of causal parameters [3].

The A-IPTW estimator can be defined as a linear regression of the transformed outcome [4, 2] on the set of covariates $\boldsymbol{V}$. We can construct the doubly robust transformed outcome as:

$$Y^* = \frac{2A-1}{g_n(A|\boldsymbol{X})}\{Y - \overline{Q}_n(A, \boldsymbol{X})\} + \overline{Q}_n(1, \boldsymbol{X}) - \overline{Q}_n(0, \boldsymbol{X})$$

Here, a subscript n is used to denote a fitted value of the propensity score or outcome regression quantities. The coefficient estimates in the linear regression of Y^* are denoted as $\hat{\boldsymbol{\beta}}_{\boldsymbol{V}}^{AIPTW}$, and are the A-IPTW estimates of $\boldsymbol{\beta}_{\boldsymbol{V}}$.

Note that the A-IPTW estimator satisfies the equation $\sum_{j=1}^{J}\sum_{i\in C_j}\mathcal{D}_{ij,n}(\hat{\boldsymbol{\beta}}_{\boldsymbol{V}}^{AIPTW}) = 0$ where $\mathcal{D}_{ij,n}(\hat{\boldsymbol{\beta}}_{\boldsymbol{V}}^{AIPTW})$ is the empirical influence curve evaluated at $\overline{Q}_n(a_{ij}, \boldsymbol{x}_{ij})$ and $g_n(a_{ij}|\boldsymbol{x}_{ij})$, with parameter estimate $\hat{\boldsymbol{\beta}}_V$. The estimator is doubly robust, meaning that if either the propensity scores or the outcome regression estimates converge to their true values, $\hat{\boldsymbol{\beta}}_{\boldsymbol{V}}^{AIPTW}$ converges to $\boldsymbol{\beta}_{\boldsymbol{V}}$ [2] as the sample sizes of both the in-study populations and the number of studies increase. If both of the estimators $\overline{Q}_n$ and g_n are consistent and converge overall at sub-parametric $n^{-1/4}$ rates under regularity conditions, the A-IPTW is an asymptotically linear estimator. To estimate the variance of the parameter of interest, it suffices to consider the influence function of the estimator [5].

Appendix E: Simulation results

Table S1: The simulation data generating mechanism for $J \in \{10, 30, 50\}$ studies and $n_j = 300$ $(for\ j = 1, 2, \cdots, J)$ subjects in each study. C_j is the set of indices of subjects in study j.

Variable	**Generating Mechanism**
$S_{1,ij}$	$S_{1,j} \sim N(mean = 0.5, sd = 0.8, n = J)$ Set $S_{1,ij} = S_{1,j}$ for all i in C_j
$S_{2,ij}$	$S_{2,j} \sim N(mean = 0.5S_{1,ij} + 0.1, sd = 0.5, n = J)$ Set $S_{2,ij} = S_{2,j}$ for all i in C_j
$W_{1,ij}$	$W_{1,ij} \sim N(mean = 0.3S_{1,j}, sd = 0.7, n = n_j)$
$W_{2,ij}$	$W_{2,ij} \sim Bin(logit(p) = 0.5 + 0.1S_{1,j}, n = n_j)$
$W_{3,ij}$	$W_{3,ij} \sim Bin(logit(p) = 0.4 + 0.2S_{1,j}, n = n_j)$
$D_{ij}^{(k)}$	$D_j^{(1)} \sim Bin(logit(p) = 0.8 + 0.5S_{1,ij}, n = J)$ $D_j^{(2)} \sim Bin(logit(p) = 0.9 + 0.4S_{1,ij}, n = J)$ $D_j^{(3)} \mid (D_j^{(1)}, D_j^{(2)})$ $\begin{cases} = 1 & if\ d_j^{(1)} = d_j^{(2)} = 0 \\ \sim Bin(logit(p) = 0.2 + 0.3S_{1,ij}, n = J) & otherwise \end{cases}$ Set $D_{ij}^{(k)} = D_j^{(k)}$ for all i in C_j
$A_{ij}^{(k)}$	$A_{ij}^{(1)} \sim Bin(logit(p) = -2 + 1.6S_{1,ij} + 2.5W_{1,ij} + 1.2W_{2,ij}, n = n_j)$ $A_{ij}^{(2)} \sim Bin(logit(p) = 0.6 + 0.3S_{1,ij} + 0.4W_{3,ij}, n = n_j)$ $A_{ij}^{(3)} \sim Bin(logit(p) = 0.5 + 0.3S_{1,ij} + 0.2W_{1,ij}, n = n_j)$
Y_{ij}	Without random effects: $Y_{ij} \sim \{2 + 0.3W_{1,ij} + 0.6W_{2,ij} + 0.9W_{3,ij} + 0.8S_{1,ij} - 0.4S_{2,ij}$ $+0.5A_{ij}^{(1)} + A_{ij}^{(2)} + 0.5A_{ij}^{(3)} + A_{ij}^{(1)}(0.65W_{1,ij} + 0.35W_{3,ij})$ $+A_{ij}^{(2)}(0.8W_{1,ij}) + A_{ij}^{(3)}(W_{2,ij}) + \epsilon, n = n_j\}$ $where\ \epsilon \sim N(mean = 0, sd = 1)$ With random effects: $Y_{ij} \sim \{2 + 0.3W_{1,ij} + 0.6W_{2,ij} + 0.9W_{3,ij} + 0.8S_{1,ij} - 0.4S_{2,ij}$ $+0.5A_{ij}^{(1)} + A_{ij}^{(2)} + 0.5A_{ij}^{(3)} + A_{ij}^{(1)}(0.65W_{1,ij} + 0.35W_{3,ij})$ $+\underline{A_{ij}^{(1)}(0.7S_{2,ij})} + A_{ij}^{(2)}(0.8W_{1,ij}) + A_{ij}^{(3)}(W_{2,ij}) + \epsilon, n = n_j\}$ $where\ \epsilon \sim N(mean = 0, sd = 1)$

Table S2: Estimated effect modification coefficients and their corresponding non clustered (noCl), clustered (Cl) and Monte Carlo (MC) standard errors based on TMLE under four scenarios for three sample sizes. The simulated data are generated without random effects. The four scenarios are as follows: Scenario 1 - both Q and g models are correct; Scenario 2 - Q model is correct, g is null; Scenario 3 - Q model is null, g model is correct; Scenario 4 - both Q and g models are null.

			N=3000, no. of cluster=10				N=9000, no. of cluster=30				N=15000, no. of cluster=50			
Scenarios	**Var**	β_V	$\hat{\beta}_V$	SE_{noCl}	SE_{Cl}	SE_{MC}	$\hat{\beta}_V$	SE_{noCl}	SE_{Cl}	SE_{MC}	$\hat{\beta}_V$	SE_{noCl}	SE_{Cl}	SE_{MC}
Scenario 1	W_1	**0.65**	**0.65**	0.15	0.14	0.16	**0.65**	0.10	0.10	0.10	**0.65**	0.09	0.09	0.08
	W_2	**0.00**	**0.00**	0.18	0.16	0.17	**0.00**	0.12	0.11	0.11	**0.00**	0.10	0.10	0.09
	W_3	**0.35**	**0.35**	0.16	0.15	0.14	**0.35**	0.11	0.10	0.09	**0.35**	0.09	0.09	0.08
	$A^{(2)}$	**0.00**	**0.01**	0.16	0.15	0.17	**0.00**	0.11	0.11	0.11	**0.00**	0.09	0.09	0.09
	$A^{(3)}$	**0.00**	**0.00**	0.16	0.16	0.20	**0.00**	0.10	0.11	0.11	**0.00**	0.09	0.09	0.09
Scenario 2	W_1	**0.65**	**0.65**	0.06	0.08	0.08	**0.65**	0.03	0.06	0.05	**0.65**	0.02	0.05	0.04
	W_2	**0.00**	**0.00**	0.08	0.08	0.09	**0.00**	0.04	0.05	0.05	**0.00**	0.04	0.04	0.04
	W_3	**0.35**	**0.35**	0.08	0.08	0.08	**0.35**	0.05	0.05	0.05	**0.35**	0.04	0.04	0.04
	$A^{(2)}$	**0.00**	**0.00**	0.08	0.10	0.12	**0.00**	0.04	0.07	0.06	**0.00**	0.04	0.06	0.05
	$A^{(3)}$	**0.00**	**0.00**	0.08	0.11	0.14	**0.00**	0.05	0.07	0.08	**0.00**	0.04	0.06	0.06
Scenario 3	W_1	**0.65**	**0.65**	0.17	0.19	0.22	**0.65**	0.13	0.15	0.17	**0.66**	0.11	0.13	0.14
	W_2	**0.00**	**-0.02**	0.20	0.19	0.20	**-0.01**	0.14	0.14	0.14	**-0.02**	0.12	0.12	0.11
	W_3	**0.35**	**0.34**	0.18	0.16	0.17	**0.35**	0.13	0.13	0.12	**0.35**	0.11	0.10	0.09
	$A^{(2)}$	**0.00**	**0.08**	0.18	0.23	0.28	**0.04**	0.13	0.18	0.19	**0.03**	0.10	0.15	0.15
	$A^{(3)}$	**0.00**	**0.01**	0.18	0.23	0.30	**0.02**	0.12	0.18	0.20	**0.02**	0.10	0.16	0.17
Scenario 4	W_1	**0.65**	**0.62**	0.06	0.19	0.17	**0.61**	0.04	0.13	0.10	**0.61**	0.03	0.11	0.08
	W_2	**0.00**	**-0.05**	0.09	0.12	0.12	**-0.05**	0.05	0.08	0.07	**-0.06**	0.04	0.06	0.05
	W_3	**0.35**	**0.31**	0.09	0.08	0.09	**0.31**	0.05	0.05	0.06	**0.31**	0.04	0.04	0.05
	$A^{(2)}$	**0.00**	**0.51**	0.09	0.21	0.27	**0.52**	0.05	0.15	0.16	**0.52**	0.04	0.12	0.13
	$A^{(3)}$	**0.00**	**0.03**	0.09	0.22	0.30	**0.05**	0.05	0.16	0.17	**0.05**	0.04	0.13	0.14

Table S3: Estimated effect modification coefficients and their corresponding non clustered (noCl), clustered (Cl) and Monte Carlo (MC) standard errors based on TMLE under four scenarios for three sample sizes. The data are generated with random effects. The four scenarios are as follows: Scenario 1 - both Q and g models are correct; Scenario 2 - Q model is correct, g is null; Scenario 3 - Q model is null, g model is correct; Scenario 4 - both Q and g models are null.

			N=3000, no. of cluster=10				N=9000, no. of cluster=30				N=15000, no. of cluster=50			
Scenarios	**Var**	β_V	$\hat{\beta}_V$	SE_{noCl}	SE_{Cl}	SE_{MC}	$\hat{\beta}_V$	SE_{noCl}	SE_{Cl}	SE_{MC}	$\hat{\beta}_V$	SE_{noCl}	SE_{Cl}	SE_{MC}
Scenario 1	W_1	**0.77**	**0.75**	0.15	0.14	0.17	**0.76**	0.11	0.11	0.11	**0.76**	0.09	0.09	0.08
	W_2	**0.02**	**0.01**	0.18	0.16	0.16	**0.02**	0.12	0.12	0.11	**0.02**	0.10	0.10	0.08
	W_3	**0.38**	**0.38**	0.16	0.15	0.14	**0.38**	0.11	0.11	0.09	**0.38**	0.09	0.09	0.07
	$A^{(2)}$	**0.08**	**0.08**	0.16	0.19	0.24	**0.08**	0.11	0.14	0.14	**0.08**	0.09	0.11	0.11
	$A^{(3)}$	**0.02**	**0.02**	0.16	0.20	0.26	**0.02**	0.10	0.14	0.14	**0.03**	0.09	0.12	0.11
Scenario 2	W_1	**0.77**	**0.75**	0.06	0.07	0.11	**0.76**	0.03	0.04	0.06	**0.77**	0.02	0.04	0.04
	W_2	**0.02**	**0.02**	0.08	0.08	0.09	**0.02**	0.05	0.04	0.05	**0.02**	0.04	0.04	0.04
	W_3	**0.38**	**0.38**	0.08	0.08	0.08	**0.38**	0.05	0.05	0.05	**0.38**	0.04	0.04	0.04
	$A^{(2)}$	**0.08**	**0.07**	0.08	0.15	0.19	**0.08**	0.04	0.10	0.10	**0.08**	0.04	0.08	0.08
	$A^{(3)}$	**0.02**	**0.02**	0.08	0.17	0.22	**0.02**	0.05	0.11	0.12	**0.02**	0.04	0.09	0.09
Scenario 3	W_1	**0.77**	**0.73**	0.18	0.20	0.25	**0.75**	0.14	0.17	0.18	**0.76**	0.11	0.14	0.15
	W_2	**0.02**	**-0.01**	0.22	0.21	0.22	**0.00**	0.16	0.15	0.15	**0.00**	0.13	0.13	0.12
	W_3	**0.38**	**0.36**	0.20	0.18	0.18	**0.38**	0.15	0.14	0.13	**0.38**	0.12	0.12	0.10
	$A^{(2)}$	**0.08**	**0.15**	0.20	0.29	0.37	**0.12**	0.14	0.22	0.25	**0.10**	0.12	0.19	0.19
	$A^{(3)}$	**0.02**	**0.04**	0.20	0.30	0.40	**0.05**	0.14	0.23	0.25	**0.05**	0.11	0.20	0.21
Scenario 4	W_1	**0.77**	**0.64**	0.06	0.23	0.17	**0.63**	0.04	0.16	0.11	**0.63**	0.03	0.13	0.09
	W_2	**0.02**	**-0.06**	0.09	0.14	0.13	**-0.07**	0.05	0.08	0.07	**-0.08**	0.04	0.07	0.06
	W_3	**0.38**	**0.33**	0.09	0.09	0.10	**0.34**	0.05	0.06	0.06	**0.34**	0.04	0.05	0.05
	$A^{(2)}$	**0.08**	**0.57**	0.09	0.27	0.36	**0.57**	0.05	0.20	0.22	**0.58**	0.04	0.16	0.17
	$A^{(3)}$	**0.02**	**0.03**	0.09	0.30	0.41	**0.07**	0.05	0.22	0.25	**0.06**	0.04	0.19	0.20

Table S4: Coverage rates based on clustered standard errors for TMLE under the four scenarios for three sample sizes. $Cov_{10}, Cov_{30}, Cov_{50}$ represent the coverage rates corresponding to three datasets of 10, 30 and 50 studies with total sample sizes of 3000, 9000, 15000, respectively. The results are summarized over 1000 iterations. The four scenarios are as follows: Scenario 1 - both Q and g models are correct; Scenario 2 - Q model is correct, g is null; Scenario 3 - Q model is null, g model is correct; Scenario 4 - both Q and g models are null.

		Without random effects			With random effects		
Scenarios	**Var**	Cov_{10}	Cov_{30}	Cov_{50}	Cov_{10}	Cov_{30}	Cov_{50}
Scenario 1	W_1	0.815	0.896	0.920	0.798	0.890	0.930
	W_2	0.866	0.943	0.951	0.881	0.946	0.957
	W_3	0.893	0.961	0.950	0.901	0.948	0.967
	$A^{(2)}$	0.851	0.940	0.940	0.856	0.922	0.946
	$A^{(3)}$	0.860	0.935	0.953	0.826	0.881	0.896
Scenario 2	W_1	0.908	0.976	0.976	0.750	0.853	0.876
	W_2	0.874	0.945	0.947	0.855	0.917	0.928
	W_3	0.902	0.927	0.922	0.899	0.927	0.938
	$A^{(2)}$	0.908	0.973	0.961	0.883	0.949	0.962
	$A^{(3)}$	0.869	0.926	0.949	0.838	0.897	0.872
Scenario 3	W_1	0.827	0.898	0.915	0.797	0.892	0.911
	W_2	0.864	0.917	0.959	0.865	0.914	0.951
	W_3	0.885	0.943	0.951	0.883	0.932	0.953
	$A^{(2)}$	0.842	0.922	0.930	0.847	0.913	0.935
	$A^{(3)}$	0.847	0.932	0.926	0.827	0.927	0.930
Scenario 4	W_1	0.939	0.972	0.961	0.900	0.891	0.881
	W_2	0.892	0.920	0.873	0.892	0.859	0.744
	W_3	0.864	0.858	0.819	0.849	0.829	0.806
	$A^{(2)}$	0.294	0.119	0.034	0.401	0.305	0.182
	$A^{(3)}$	0.824	0.903	0.908	0.822	0.909	0.917

Table S5: Estimated effect modification coefficients and their corresponding non clustered (noCl), clustered (Cl) and Monte Carlo (MC) standard errors based on A-IPTW under the four scenarios for three sample sizes. The data is generated without random effects. The four scenarios are as follows: Scenario 1 - both Q and g models are correct; Scenario 2 - Q model is correct, g is null; Scenario 3 - Q model is null, g model is correct; Scenario 4 - both Q and g models are null.

			N=3000, no. of cluster=10				N=9000, no. of cluster=30				N=15000, no. of cluster=50			
Scenarios	**Var**	β_V	$\hat{\beta}_V$	SE_{noCl}	SE_{Cl}	SE_{MC}	$\hat{\beta}_V$	SE_{noCl}	SE_{Cl}	SE_{MC}	$\hat{\beta}_V$	SE_{noCl}	SE_{Cl}	SE_{MC}
Scenario 1	W_1	**0.65**	**0.64**	0.25	0.24	0.27	**0.65**	0.14	0.13	0.14	**0.65**	0.10	0.10	0.11
	W_2	**0.00**	**0.00**	0.27	0.26	0.27	**0.00**	0.14	0.14	0.14	**0.00**	0.11	0.11	0.11
	W_3	**0.35**	**0.35**	0.23	0.22	0.22	**0.35**	0.13	0.12	0.12	**0.35**	0.10	0.10	0.10
	$A^{(2)}$	**0.00**	**0.01**	0.23	0.22	0.26	**0.00**	0.13	0.13	0.14	**0.0**0	0.10	0.11	0.11
	$A^{(3)}$	**0.00**	**0.00**	0.23	0.23	0.26	**0.00**	0.12	0.13	0.14	**0.01**	0.10	0.11	0.11
Scenario 2	W_1	**0.65**	**0.65**	0.06	0.09	0.09	**0.65**	0.03	0.06	0.05	**0.65**	0.02	0.05	0.04
	W_2	**0.00**	**0.00**	0.08	0.08	0.09	**0.00**	0.04	0.05	0.05	**0.00**	0.04	0.04	0.04
	W_3	**0.35**	**0.35**	0.08	0.08	0.08	**0.35**	0.05	0.05	0.05	**0.35**	0.04	0.04	0.04
	$A^{(2)}$	**0.00**	**0.00**	0.08	0.12	0.14	**0.00**	0.04	0.07	0.06	**0.00**	0.04	0.06	0.05
	$A^{(3)}$	**0.00**	**0.00**	0.08	0.11	0.15	**0.00**	0.05	0.07	0.08	**0.00**	0.04	0.06	0.06
Scenario 3	W_1	**0.65**	**0.62**	0.87	0.98	0.92	**0.62**	0.52	0.67	0.59	**0.60**	0.39	0.55	0.44
	W_2	**0.00**	**0.03**	0.89	0.90	0.87	**-0.04**	0.49	0.53	0.49	**-0.05**	0.38	0.42	0.38
	W_3	**0.35**	**0.34**	0.82	0.86	0.83	**0.36**	0.47	0.54	0.49	**0.34**	0.36	0.41	0.36
	$A^{(2)}$	**0.00**	**0.06**	0.72	0.78	0.90	**0.04**	0.43	0.56	0.58	**0.05**	0.33	0.46	0.45
	$A^{(3)}$	**0.00**	**0.04**	0.73	0.81	0.88	**0.05**	0.42	0.57	0.60	**0.07**	0.32	0.47	0.48
Scenario 4	W_1	**0.65**	**0.24**	0.10	0.55	0.20	**0.23**	0.06	0.34	0.13	**0.23**	0.04	0.27	0.10
	W_2	**0.00**	**-0.31**	0.12	0.36	0.15	**-0.31**	0.07	0.21	0.08	**-0.32**	0.05	0.16	0.06
	W_3	**0.35**	**0.29**	0.12	0.48	0.12	**0.29**	0.07	0.28	0.07	**0.29**	0.05	0.22	0.06
	$A^{(2)}$	**0.00**	**0.50**	0.12	0.34	0.38	**0.52**	0.07	0.23	0.22	**0.51**	0.05	0.19	0.18
	$A^{(3)}$	**0.00**	**0.04**	0.12	0.37	0.43	**0.06**	0.07	0.26	0.26	**0.05**	0.05	0.21	0.21

Table S6: Estimated effect modification coefficients and their corresponding non clustered (noCl), clustered (Cl) and Monte Carlo (MC) standard errors based on A-IPTW under four scenarios for three sample sizes. The outcomes are generated with random effects. The four scenarios are as follows: Scenario 1 - both Q and g models are correct; Scenario 2 - Q model is correct, g is null; Scenario 3 - Q model is null, g model is correct; Scenario 4 - both Q and g models are null.

			N=3000, no. of cluster=10				N=9000, no. of cluster=30				N=15000, no. of cluster=50			
Scenarios	**Var**	β_V	$\hat{\beta}_V$	SE_{noCl}	SE_{Cl}	SE_{MC}	$\hat{\beta}_V$	SE_{noCl}	SE_{Cl}	SE_{MC}	$\hat{\beta}_V$	SE_{noCl}	SE_{Cl}	SE_{MC}
Scenario 1	W_1	**0.77**	**0.74**	0.23	0.22	0.26	**0.76**	0.14	0.14	0.15	**0.76**	0.10	0.11	0.11
	W_2	**0.02**	**0.01**	0.25	0.24	0.26	**0.02**	0.14	0.14	0.14	**0.02**	0.11	0.11	0.11
	W_3	**0.38**	**0.38**	0.22	0.20	0.22	**0.38**	0.13	0.13	0.13	**0.38**	0.10	0.10	0.10
	$A^{(2)}$	**0.08**	**0.07**	0.22	0.25	0.32	**0.08**	0.13	0.15	0.16	**0.08**	0.10	0.12	0.13
	$A^{(3)}$	**0.02**	**0.03**	0.21	0.26	0.31	**0.02**	0.12	0.15	0.16	**0.03**	0.10	0.12	0.13
Scenario 2	W_1	**0.77**	**0.75**	0.06	0.07	0.11	**0.76**	0.03	0.04	0.06	**0.77**	0.02	0.04	0.04
	W_2	**0.02**	**0.02**	0.08	0.08	0.09	**0.02**	0.05	0.04	0.05	**0.02**	0.04	0.04	0.04
	W_3	**0.38**	**0.38**	0.08	0.08	0.08	**0.38**	0.05	0.05	0.05	**0.38**	0.04	0.04	0.04
	$A^{(2)}$	**0.08**	**0.07**	0.08	0.17	0.24	**0.08**	0.04	0.10	0.10	**0.08**	0.04	0.08	0.08
	$A^{(3)}$	**0.02**	**0.02**	0.08	0.18	0.22	**0.02**	0.05	0.11	0.12	**0.02**	0.04	0.09	0.09
Scenario 3	W_1	**0.77**	**0.70**	0.96	1.06	1.02	**0.72**	0.58	0.73	0.64	**0.70**	0.44	0.59	0.48
	W_2	**0.02**	**0.04**	1.00	1.00	0.97	**-0.03**	0.55	0.59	0.56	**-0.03**	0.43	0.47	0.43
	W_3	**0.38**	**0.37**	0.93	0.95	0.93	**0.40**	0.54	0.59	0.55	**0.38**	0.40	0.45	0.41
	$A^{(2)}$	**0.08**	**0.13**	0.82	0.91	1.04	**0.11**	0.49	0.63	0.67	**0.13**	0.37	0.52	0.52
	$A^{(3)}$	**0.02**	**0.07**	0.83	0.93	1.01	**0.09**	0.47	0.64	0.67	**0.11**	0.36	0.53	0.53
Scenario 4	W_1	**0.77**	**0.25**	0.10	0.60	0.21	**0.24**	0.06	0.38	0.13	**0.24**	0.04	0.30	0.11
	W_2	**0.02**	**-0.33**	0.12	0.37	0.15	**-0.33**	0.07	0.22	0.09	**-0.34**	0.06	0.17	0.07
	W_3	**0.38**	**0.31**	0.13	0.48	0.13	**0.32**	0.07	0.28	0.08	**0.32**	0.06	0.22	0.06
	$A^{(2)}$	**0.08**	**0.56**	0.12	0.40	0.45	**0.57**	0.07	0.28	0.26	**0.58**	0.05	0.22	0.21
	$A^{(3)}$	**0.02**	**0.06**	0.13	0.44	0.52	**0.08**	0.07	0.32	0.33	**0.07**	0.06	0.26	0.26

Table S7: Coverage rates based on clustered standard errors for A-IPTW under four scenarios for three sample sizes. $Cov_{10}, Cov_{30}, Cov_{50}$ represent the coverage rates under 10, 30 and 50 studies with sample sizes 3000, 9000, 15000, respectively. The results summarized are over 1000 iterations. The four scenarios are as follows: Scenario 1 - both Q and g models are correct; Scenario 2 - Q model is correct, g is null; Scenario 3 - Q model is null, g model is correct; Scenario 4 - both Q and g models are null.

		Without random effects			With random effects		
Scenarios	**Var**	Cov_{10}	Cov_{30}	Cov_{50}	Cov_{10}	Cov_{30}	Cov_{50}
Scenario 1	W_1	0.868	0.918	0.937	0.853	0.911	0.939
	W_2	0.893	0.944	0.950	0.903	0.941	0.951
	W_3	0.911	0.951	0.940	0.913	0.940	0.948
	$A^{(2)}$	0.897	0.943	0.943	0.882	0.924	0.943
	$A^{(3)}$	0.893	0.939	0.947	0.884	0.933	0.950
Scenario 2	W_1	0.913	0.977	0.977	0.747	0.856	0.881
	W_2	0.867	0.946	0.950	0.850	0.920	0.925
	W_3	0.897	0.927	0.924	0.902	0.927	0.936
	$A^{(2)}$	0.924	0.972	0.964	0.891	0.952	0.963
	$A^{(3)}$	0.868	0.928	0.947	0.871	0.939	0.952
Scenario 3	W_1	0.963	0.993	0.995	0.953	0.983	0.987
	W_2	0.955	0.973	0.962	0.951	0.968	0.956
	W_3	0.954	0.989	0.985	0.947	0.983	0.984
	$A^{(2)}$	0.910	0.944	0.954	0.904	0.950	0.960
	$A^{(3)}$	0.902	0.943	0.949	0.897	0.946	0.946
Scenario 4	W_1	0.995	0.965	0.860	0.983	0.910	0.690
	W_2	0.948	0.846	0.536	0.946	0.765	0.381
	W_3	1.000	1.000	1.000	1.000	1.000	1.000
	$A^{(2)}$	0.533	0.377	0.213	0.611	0.509	0.356
	$A^{(3)}$	0.905	0.942	0.938	0.903	0.939	0.943

Table S8: Bias of TMLE and A-IPTW estimators with and without random effects in the simulation of the outcome. 10, 30 and 50 are the number of studies in each simulation. S1-4 represent the four scenarios.

		Without random effects						With random effects					
		TMLE			A-IPTW			TMLE			A-IPTW		
Scen	**Var**	10	30	50	10	30	50	10	30	50	10	30	50
S1	W_1	0.00	0.00	0.00	-0.01	0.00	0.00	-0.02	-0.01	-0.01	-0.03	-0.01	-0.01
	W_2	-0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.01	0.00	0.00	-0.01	0.00	0.00
	W_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	$A^{(2)}$	0.01	-0.00	-0.00	0.01	-0.00	-0.00	-0.01	0.00	0.00	-0.01	0.00	0.00
	$A^{(3)}$	-0.00	-0.00	0.00	0.00	-0.00	0.01	0.01	0.00	0.01	0.01	0.00	0.01
S2	W_1	0.00	0.00	0.00	0.00	0.00	0.00	-0.02	-0.01	0.00	-0.02	-0.01	0.00
	W_2	-0.00	0.00	0.00	-0.00	0.00	0.00	-0.01	0.00	0.00	-0.01	0.00	0.00
	W_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	$A^{(2)}$	0.00	0.00	-0.00	0.00	0.00	-0.00	-0.01	0.00	0.00	-0.01	0.00	0.00
	$A^{(3)}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
S3	W_1	0.00	0.00	0.01	-0.03	-0.03	-0.05	-0.04	-0.02	-0.01	-0.07	-0.05	-0.07
	W_2	-0.02	-0.01	-0.02	0.03	-0.04	-0.05	-0.03	-0.02	-0.02	0.02	-0.05	-0.05
	W_3	-0.01	0.00	0.00	-0.01	0.01	-0.01	-0.02	0.00	0.00	-0.01	0.02	0.00
	$A^{(2)}$	0.08	0.04	0.03	0.06	0.04	0.05	0.07	0.04	0.02	0.05	0.03	0.05
	$A^{(3)}$	0.01	0.02	0.02	0.04	0.05	0.07	0.02	0.03	0.03	0.05	0.07	0.09
S4	W_1	-0.03	-0.04	-0.04	-0.41	-0.42	-0.42	-0.13	-0.14	-0.14	-0.52	-0.53	-0.53
	W_2	-0.05	-0.05	-0.06	-0.31	-0.31	-0.32	-0.08	-0.09	-0.09	-0.35	-0.35	-0.36
	W_3	-0.04	-0.04	-0.04	-0.06	-0.06	-0.06	-0.05	-0.04	-0.04	-0.07	-0.06	-0.06
	$A^{(2)}$	0.51	0.52	0.52	0.50	0.52	0.51	0.49	0.50	0.50	0.48	0.50	0.50
	$A^{(3)}$	0.03	0.05	0.05	0.04	0.06	0.05	0.01	0.05	0.04	0.04	0.06	0.05

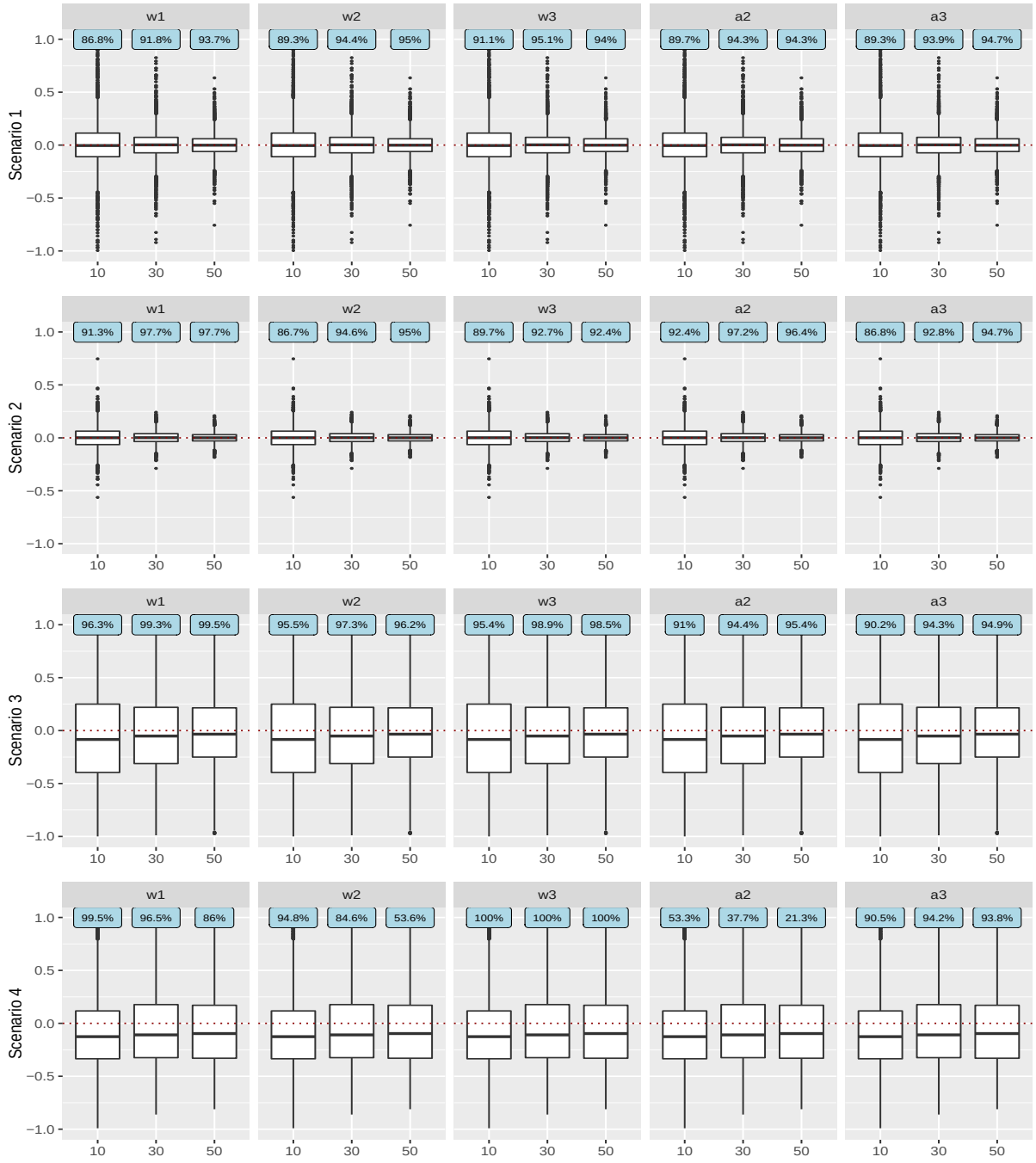

Figure S1: Estimated bias of A-IPTW estimates under four scenarios and three different sample sizes without random effects. x-axis represents the number of studies for three sample sizes. Coverage rates based on the clustered sandwich estimators of the standard error are presented in blue boxes. The four scenarios are as follows: Scenario 1 - both Q and g models are correct; Scenario 2 - Q model is correct, g is null; Scenario 3 - Q model is null, g model is correct; Scenario 4 - both Q and g models are null.

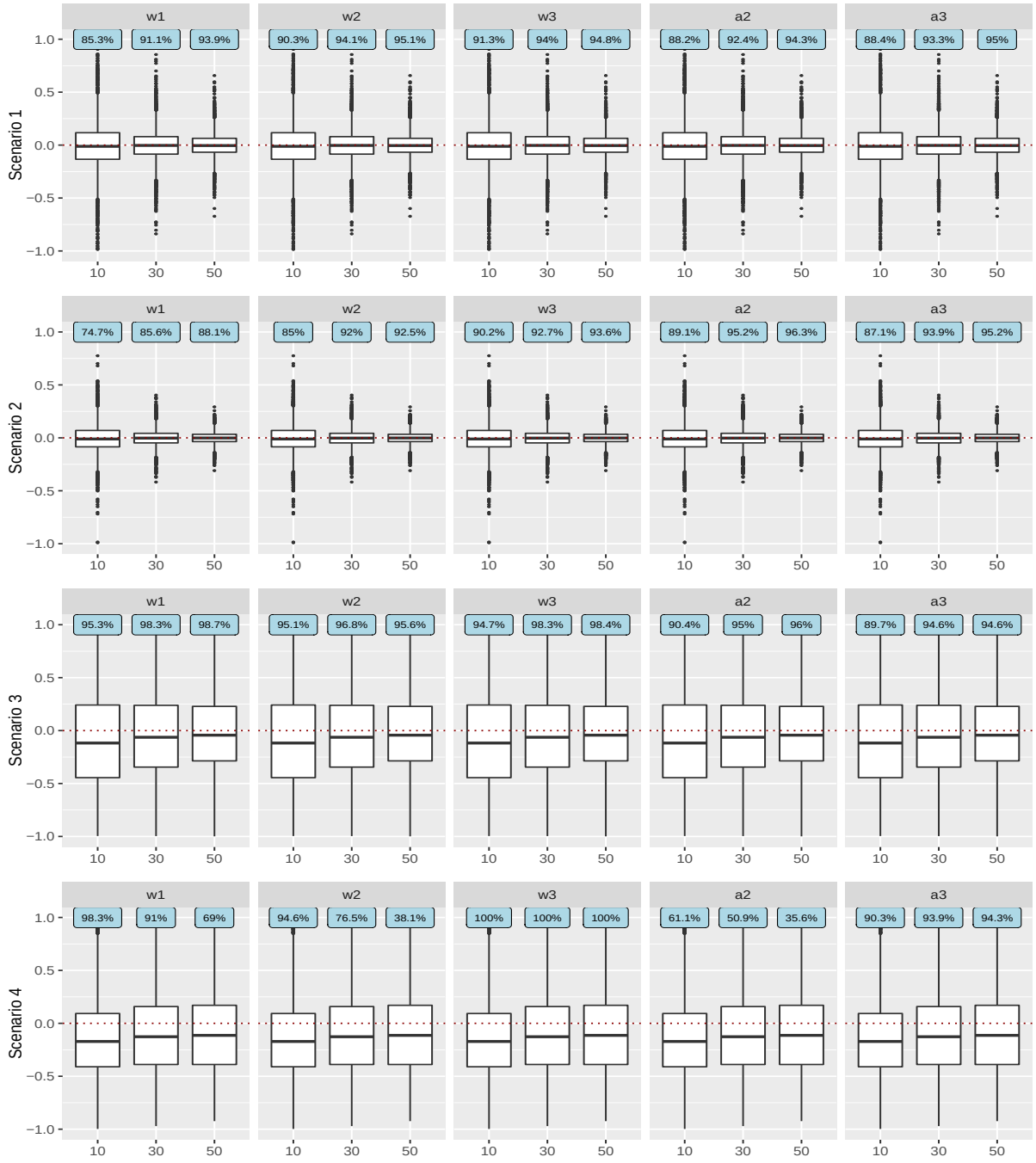

Figure S2: Estimated bias of A-IPTW estimates under four scenarios and three different sample sizes with random effects. x-axis represents the number of studies for three sample sizes. Coverage rates based on the clustered sandwich estimators of the standard error are presented in blue boxes. The four scenarios are as follows: Scenario 1 - both Q and g models are correct; Scenario 2 - Q model is correct, g is null; Scenario 3 - Q model is null, g model is correct; Scenario 4 - both Q and g models are null.

Appendix F: MDR-TB results

Table S9: Sample sizes and the numbers of patients taking each of the 14 medications by study [6].

Study	Size	EMB	CAP	CIP	CS	ETO	OFX	PAS	PTO	RIF	SM	PZA	KM/AM	LgFQ	Gp5
Ahuja	722	589	247	346	319	350	270	143	0	619	149	622	193	33	117
Avendano	66	31	0	1	24	13	7	27	0	0	2	24	61	59	66
Burgos	48	48	28	9	29	17	24	3	0	47	28	47	10	8	5
Chan	202	88	84	54	163	124	111	133	0	24	29	77	78	10	61
Chiang	125	76	0	0	31	0	122	102	117	27	67	99	42	0	0
Cox	77	28	63	0	77	72	77	77	0	2	2	66	22	0	48
Garcia	41	41	0	0	0	0	0	0	0	41	2	41	0	0	0
Granich	101	50	65	16	50	45	29	33	0	33	9	70	19	0	0
Koh	149	47	1	1	142	0	6	101	138	0	40	80	55	139	56
Leung	98	0	0	0	98	0	58	0	0	0	0	57	73	0	0
Migliori	44	12	10	0	7	19	10	15	10	0	4	18	24	30	22
Mitnick	710	191	349	488	615	478	123	597	232	9	146	279	347	235	557
Narita	77	61	11	1	24	26	48	4	0	23	30	55	5	0	24
Palmero	114	38	4	2	82	8	79	52	81	0	15	31	46	0	22
Pasvol	43	27	0	9	8	0	16	1	12	5	2	29	13	5	2
Pena	25	7	0	0	0	0	0	13	0	0	0	0	0	0	7
Perez	34	28	0	17	0	0	7	0	9	0	16	17	13	0	24
Quy	142	123	0	0	0	0	0	0	0	142	142	142	0	0	0
Riekstina	1,026	407	474	0	888	0	962	552	777	0	26	314	601	6	691
Robert	45	28	0	0	13	18	34	4	0	11	6	36	21	0	5
Schaaf	33	28	0	0	1	28	29	0	0	10	11	29	9	0	9
Seung	142	3	0	0	142	0	142	72	142	0	104	71	40	0	0
Shim	1,364	188	0	0	1,199	0	1,113	887	1,186	149	319	610	477	37	45
Shin	608	150	353	0	541	450	545	560	27	0	0	507	254	9	45
Shiraishi	61	16	1	1	44	40	0	34	0	3	16	44	36	56	12
Tabarsi	43	14	0	0	36	0	43	0	7	0	0	35	41	0	26
Tupasi	170	45	28	19	160	40	122	140	151	9	71	123	119	72	79
Van der Werf	43	39	0	4	5	1	36	15	0	19	0	36	33	0	34
Van der Walt	2,182	1,642	0	0	476	2,182	2,182	0	0	0	0	2,182	2,182	0	0
Viiklepp	284	62	150	0	263	0	237	211	240	44	41	253	115	0	84
Yim	211	81	6	0	192	0	32	161	175	44	89	108	86	167	97
Total	9,030	4,188	1,874	968	5,629	3,911	6,464	3,937	3,304	1,261	1,366	6,102	5,015	866	2,138

Table S10: Summary of proportions of patients with each combination of 6 individual level covariates and medications after removing the missing values. There are 9030 patients in total with 4 missing values for sex (0.3%), 26 for age (14.4%), 1304 for HIV (14.4%), 2345 missing data for cavity (26%), 519 for past TB (5.7%) and 1405 for AFB (15.6%). For each medication, the last two rows represent the number of studies (J) that observed the usage of the given medication and the total sample size corresponding to those J studies.

		n	EMB	CAP	CIP	CS	ETO	OFX	PAS	PTO	RIF	SM	PZA	KM/AM	LgFQ	Gp5
Sex	**male**	6,147	0.660	0.699	0.640	0.698	0.653	0.696	0.705	0.716	0.697	0.685	0.676	0.666	0.579	0.679
	female	2,879	0.340	0.301	0.360	0.302	0.347	0.304	0.295	0.284	0.303	0.315	0.324	0.334	0.421	0.321
Age	**0-17**	154	0.018	0.019	0.039	0.017	0.020	0.013	0.019	0.014	0.023	0.029	0.017	0.013	0.025	0.027
	18-25	1,283	0.129	0.157	0.186	0.153	0.170	0.136	0.164	0.137	0.059	0.119	0.137	0.161	0.232	0.175
	26-45	4,892	0.597	0.563	0.597	0.519	0.608	0.560	0.509	0.470	0.620	0.539	0.576	0.565	0.483	0.531
	46-60	1,974	0.198	0.207	0.137	0.231	0.174	0.220	0.226	0.266	0.201	0.217	0.207	0.206	0.192	0.209
	61-91	701	0.057	0.055	0.041	0.081	0.028	0.071	0.081	0.113	0.096	0.096	0.064	0.056	0.068	0.057
HIV	**pos.**	1,159	0.261	0.124	0.285	0.075	0.266	0.147	0.029	0.011	0.397	0.089	0.208	0.179	0.024	0.055
	neg.	6,567	0.739	0.876	0.715	0.925	0.734	0.853	0.971	0.989	0.603	0.911	0.792	0.821	0.976	0.945
Cavity	**yes**	4,568	0.739	0.738	0.874	0.659	0.797	0.666	0.649	0.607	0.537	0.672	0.707	0.732	0.855	0.791
	no	2,117	0.261	0.262	0.126	0.341	0.203	0.334	0.351	0.393	0.463	0.328	0.293	0.268	0.145	0.209
past TB	**yes**	6,380	0.713	0.732	0.612	0.756	0.847	0.796	0.789	0.740	0.309	0.680	0.750	0.819	0.747	0.755
	no	2,131	0.287	0.268	0.388	0.244	0.153	0.204	0.211	0.260	0.691	0.320	0.250	0.181	0.253	0.245
AFB	**pos.**	5,683	0.737	0.769	0.741	0.751	0.736	0.730	0.758	0.720	0.797	0.805	0.744	0.727	0.761	0.755
	neg.	1,942	0.263	0.231	0.259	0.249	0.264	0.270	0.242	0.280	0.203	0.195	0.256	0.273	0.239	0.245
Study	**J**		30	16	14	27	17	27	24	15	19	25	30	28	14	24
Total	**n**		4,188	1,874	968	5,629	3,911	6,464	3,937	3,304	1,261	1,366	6,102	5,015	866	2,138

Table S11: Estimated effect modification coefficients and standard errors for the intercept and six covariates. Each column gives the results for a given medication. In each cell, the upper number represents the estimates, and the lower (in parentheses) is the standard error.

	EMB	**CAP**	**CIP**	**CS**	**ETO**	**OFX**	**PAS**	**PTO**	**RIF**	**SM**	**PZA**	**KM/AM**	**LgFQ**	**Gp5**
β_0	0.013	0.488	0.072	0.232	0.081	0.257	-0.039	-0.045	-0.432	0.124	0.000	0.443	0.529	0.212
	(0.159)	(0.208)	(0.119)	(0.143)	(0.109)	(0.177)	(0.131)	(0.185)	(0.287)	(0.182)	(0.099)	(0.176)	(1.190)	(0.123)
Age	-0.015	-0.020	0.001	-0.002	-0.025	0.022	-0.013	0.008	-0.011	0.005	0.009	0.014	-0.105	-0.013
	(0.022)	(0.017)	(0.025)	(0.014)	(0.021)	(0.016)	(0.012)	(0.022)	(0.032)	(0.031)	(0.011)	(0.033)	(0.213)	(0.032)
Sex	0.095	-0.012	0.052	0.026	0.020	-0.014	0.017	-0.037	0.137	-0.133	-0.001	-0.031	-0.114	0.010
	(0.032)	(0.047)	(0.036)	(0.017)	(0.042)	(0.032)	(0.027)	(0.021)	(0.185)	(0.059)	(0.034)	(0.055)	(0.386)	(0.039)
Afb	-0.064	-0.026	-0.024	-0.002	-0.052	-0.041	0.001	0.073	0.119	0.089	-0.049	0.045	-0.224	0.044
	(0.037)	(0.063)	(0.041)	(0.034)	(0.051)	(0.048)	(0.045)	(0.031)	(0.138)	(0.074)	(0.032)	(0.073)	(0.476)	(0.064)
Cavity	-0.085	-0.145	0.023	0.013	-0.066	0.077	-0.004	0.067	0.013	0.059	0.020	-0.108	0.081	-0.070
	(0.049)	(0.069)	(0.071)	(0.052)	(0.056)	(0.056)	(0.038)	(0.063)	(0.086)	(0.060)	(0.031)	(0.086)	(0.616)	(0.058)
HIV	-0.007	-0.243	-0.270	0.025	0.130	0.050	0.016	0.104	-0.125	-0.151	0.130	0.091	-0.208	-0.135
	(0.046)	(0.095)	(0.116)	(0.046)	(0.183)	(0.198)	(0.095)	(0.105)	(0.078)	(0.120)	(0.102)	(0.132)	(0.915)	(0.127)
pastTB	0.029	0.014	0.034	-0.090	0.068	0.044	0.042	-0.107	0.090	0.081	0.050	0.003	0.172	-0.014
	(0.062)	(0.086)	(0.053)	(0.04)	(0.042)	(0.067)	(0.043)	(0.049)	(0.112)	(0.076)	(0.029)	(0.064)	(0.796)	(0.048)

Table S12: Estimated effect modification coefficients and standard errors. Each column represents the effect modification of the named medication and each row represents the potential effect modifier. In each cell, the upper value represents the estimate, and the lower (in parentheses) is the standard error.

	EMB	**CAP**	**CIP**	**CS**	**ETO**	**OFX**	**PAS**	**PTO**	**RIF**	**SM**	**PZA**	**KM/AM**	**LgFQ**	**Gp5**
EMB	-	-0.025	0.094	-0.031	-0.051	-0.031	0.083	0.103	0.075	0.165	-0.058	-0.115	-0.099	0.079
	-	(0.052)	(0.062)	(0.028)	(0.066)	(0.057)	(0.053)	(0.035)	(0.096)	(0.066)	(0.047)	(0.062)	(0.300)	(0.063)
CAP	-0.021	-	-0.164	-0.052	-0.242	-0.031	-0.049	-0.006	0.071	-0.093	-0.053	-0.059	-0.261	-0.107
	(0.065)	-	(0.054)	(0.048)	(0.052)	(0.105)	(0.055)	(0.101)	(0.168)	(0.109)	(0.039)	(0.111)	(0.578)	(0.056)
CIP	-0.012	-0.017	-	0.019	0.112	0.109	0.104	0.058	-0.235	-0.127	0.022	-0.137	0.277	-0.071
	(0.052)	(0.159)	-	(0.054)	(0.071)	(0.258)	(0.114)	(0.069)	(0.197)	(0.138)	(0.049)	(0.190)	(1.684)	(0.108)
CS	-0.063	-0.064	0.011	-	0.148	-0.051	-0.131	-0.127	-0.054	-0.183	0.092	-0.018	-0.303	-0.051
	(0.053)	(0.067)	(0.054)	-	(0.045)	(0.069)	(0.063)	(0.054)	(0.077)	(0.086)	(0.057)	(0.147)	(0.291)	(0.067)
ETO	0.041	-0.101	0.006	0.089	-	-0.182	0.018	0.103	-0.113	0.080	-0.048	-0.188	-0.052	-0.003
	(0.049)	(0.086)	(0.078)	(0.042)	-	(0.097)	(0.061)	(0.069)	(0.160)	(0.065)	(0.061)	(0.129)	(0.499)	(0.068)
OFX	-0.009	-0.021	0.078	-0.017	0.001	-	0.101	-0.007	0.131	-0.080	-0.052	-0.210	-0.153	0.098
	(0.059)	(0.136)	(0.074)	(0.053)	(0.055)	-	(0.074)	(0.047)	(0.332)	(0.200)	(0.038)	(0.202)	(1.234)	(0.095)
PAS	0.005	-0.026	0.087	-0.050	0.030	0.051	-	0.002	0.185	0.001	-0.012	0.030	0.250	-0.087
	(0.047)	(0.055)	(0.040)	(0.037)	(0.038)	(0.062)	-	(0.045)	(0.073)	(0.060)	(0.033)	(0.139)	(0.519)	(0.073)
PTO	0.096	0.027	0.159	-0.004	0.128	-0.101	0.101	-	-0.036	-0.050	-0.027	-0.157	0.014	0.028
	(0.083)	(0.090)	(0.070)	(0.059)	(0.062)	(0.102)	(0.059)	-	(0.183)	(0.074)	(0.037)	(0.093)	(0.528)	(0.068)
RIF	0.025	0.069	-0.064	-0.126	-0.027	-0.098	0.005	-0.013	-	0.009	0.072	-0.054	0.192	-0.073
	(0.076)	(0.090)	(0.084)	(0.065)	(0.087)	(0.125)	(0.070)	(0.095)	-	(0.148)	(0.057)	(0.126)	(1.654)	(0.068)
SM	0.108	-0.088	-0.085	-0.236	-0.101	-0.045	-0.062	-0.034	-0.242	-	0.058	-0.032	-0.154	-0.030
	(0.078)	(0.089)	(0.048)	(0.037)	(0.068)	(0.051)	(0.043)	(0.062)	(0.166)	-	(0.045)	(0.170)	(1.134)	(0.069)
PZA	-0.036	-0.066	0.101	-0.044	0.029	-0.069	-0.047	-0.080	0.174	0.116	-	0.043	0.000	-0.013
	(0.065)	(0.092)	(0.047)	(0.036)	(0.051)	(0.055)	(0.054)	(0.059)	(0.115)	(0.052)	-	(0.131)	(0.291)	(0.041)
KM/AM	-0.001	-0.107	-0.158	-0.058	-0.166	-0.079	-0.029	0.064	-0.218	-0.14	-0.028	-	-0.017	-0.092
	(0.064)	(0.083)	(0.061)	(0.031)	(0.044)	(0.079)	(0.046)	(0.042)	(0.302)	(0.100)	(0.042)	-	(0.783)	(0.078)
LgFQ	-0.026	-0.273	0.016	-0.103	-0.013	-0.011	0.036	-0.001	0.036	0.019	0.051	-0.232	-	-0.127
	(0.059)	(0.272)	(0.075)	(0.064)	(0.059)	(0.327)	(0.056)	(0.076)	(0.674)	(0.16)	(0.033)	(0.173)	-	(0.088)
Gp5	0.037	-0.121	-0.129	0.012	-0.107	0.037	-0.079	0.049	0.048	-0.088	-0.021	0.090	-0.331	-
	(0.054)	(0.071)	(0.058)	(0.062)	(0.045)	(0.062)	(0.057)	(0.069)	(0.125)	(0.099)	(0.044)	(0.113)	(0.679)	-

Table S13: Summary of the estimated $Pr(A = 1|\boldsymbol{X})$ for each of the 14 medications

	Minimum	5% Quantile	Median	95% Quantile	Maximum
EMB	0.00576	0.02901	0.43321	0.92581	0.96728
CAP	0.00101	0.01462	0.09539	0.44997	0.51206
CIP	0.00022	0.00551	0.06477	0.38125	0.45060
CS	0.00525	0.08334	0.67773	0.86341	0.87079
ETO	0.00166	0.03193	0.32798	0.54188	0.54752
OFX	0.00058	0.06633	0.77979	0.85201	0.86978
PAS	0.00287	0.05702	0.42128	0.77329	0.86634
PTO	0.00003	0.00182	0.23199	0.47145	0.48373
RIF	0.00020	0.00333	0.05279	0.55159	0.74471
SM	0.00024	0.00680	0.06327	0.53681	0.80091
PZA	0.02318	0.11876	0.74804	0.95482	0.96586
KM/AM	0.00116	0.03656	0.54369	0.88310	0.89926
LgFQ	0.00001	0.00069	0.03262	0.35000	0.60993
Gp5	0.00182	0.02494	0.25323	0.65496	0.77199

Appendix G: Multiple testing

In a multiple testing situation, there are a variety of criteria that may be considered. The classic approach to the multiple comparison problem is to control the familywise error rate (FWER) which is the probability of making at least one false discovery. The Bonferroni correction is one of the most common ways to control the FWER. We can find the critical value α for an individual test through dividing the FWER (usually 0.05) by the number of tests. The Bonferroni correction tends to be a bit too conservative. It is mainly useful for a fairly small number of multiple comparisons.

An alternative approach is to control the expected proportion of falsely rejected hypotheses which is also called false discovery rate (FDR). It was developed and demonstrated to be a simple and powerful procedure by Benjamini and Hochberg (1995) [7]. Compared to FWER,

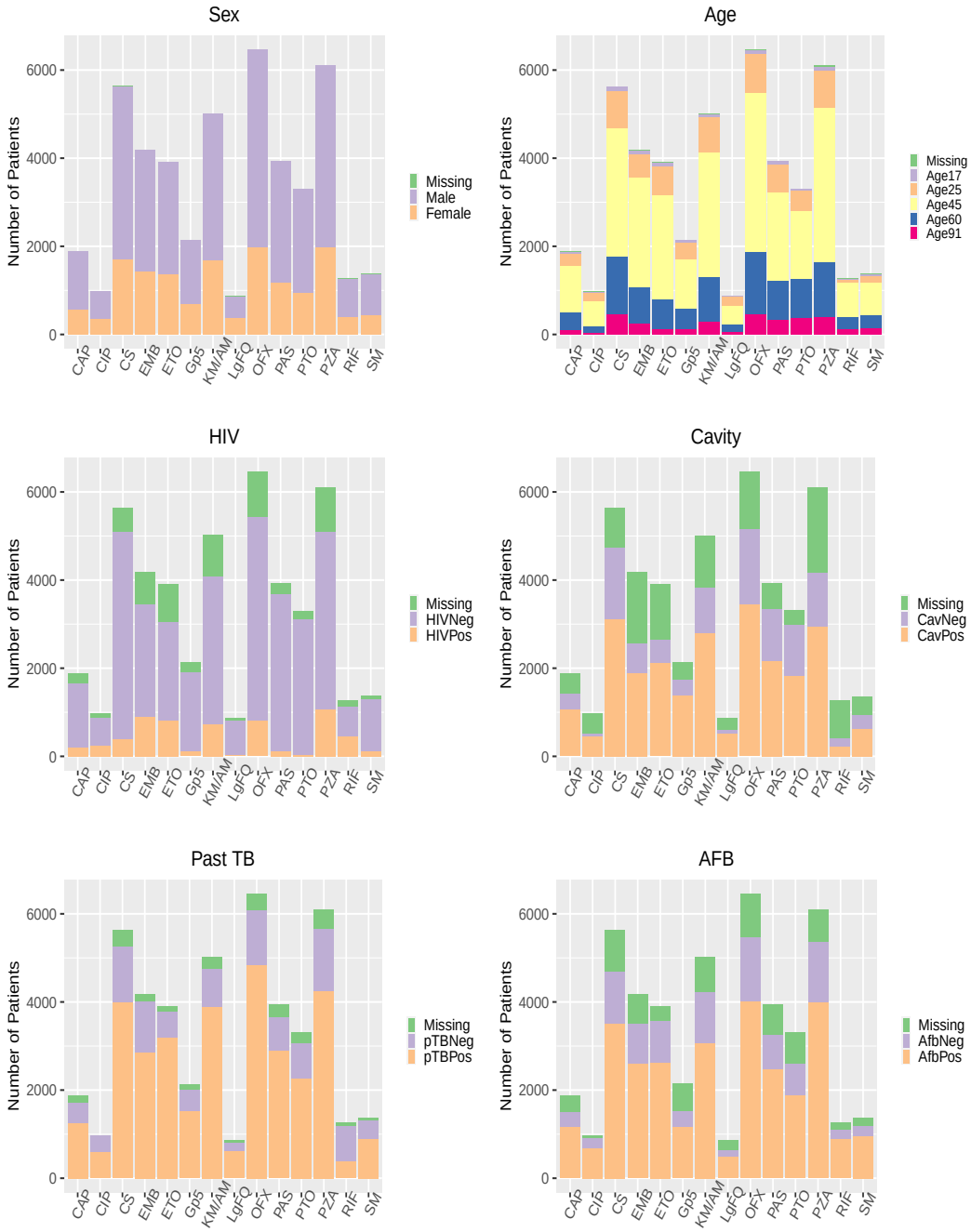

Figure S3: Summary of numbers of patients with combinations among 6 covariates and 14 medications. Age17 represents 0-17 year old group; Age25 represents the 18-25 year old group; Age45 represents the 26-45 year old group; Age60 represents the 46-60 year old group; Age91 represents the 61-91 year old group.

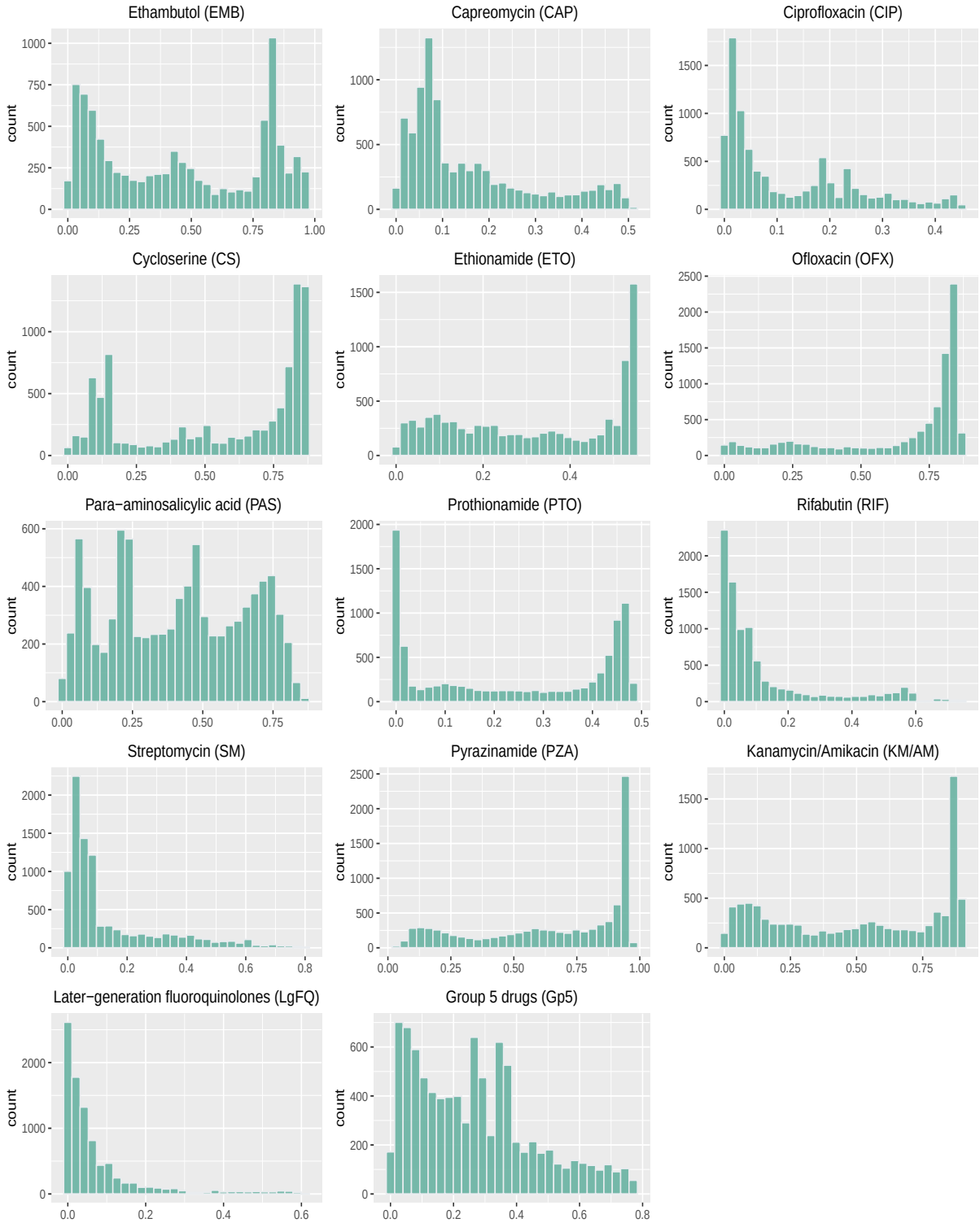

Figure S4: Distribution of propensity scores for 14 medications

the error rate is equivalent when all hypotheses are true but is smaller otherwise. To perform a FDR test, we need to order the individual p values from the smallest to the largest. If there are n total number of tests, the smallest p has a rank of $i = 1$ and the largest has $i = n$. Then the next step is to compare each individual p value to its Benjamini-Hochberg critical value, $(i/n)q$, where q is the false discovery rate we choose. The largest p value that has $p < (i/n)q$ is significant, and all of the p values smaller than it are also significant, even those that are not less than their Benjamini-Hochberg critical value. There is another way to look at this method. We can calculate q which is an estimate of FDR by $q = p(n/i)$ where p is the i-th smallest p value out of n total p values. The q value equation can be treated as the expected false positives divided by the total number of positives actually accepted at that same p value. Then we can use the q value much like a p value. For example, if we choose $q = 0.05$, that means we expect that 5% or less of our accepted results will be false. However, there is a problem when q is not a monotonic function. To address this concern, Yekutieli and Benjamini introduced the FDR-adjustment method in which q was replaced with the smallest value among all lower-rank q values that we calculated then monotonicity is enforced [8].

Bibliography

[1] Rosenblum M and van der Laan MJ. Targeted maximum likelihood estimation of the parameter of a marginal structural model. *Int J Biostat* 2010; **6**(2): Article 19.

[2] van der Laan MJ and Luedtke AR. Targeted learning of an optimal dynamic treatment, and statistical inference for its mean outcome. *UC Berkeley Division of Biostatistics Working Paper Series*. Working Paper 2014; 317.

[3] Kang JD, Schafer JL. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Stat Sci.* 2007; **27**: 523-539.

[4] Bahamyirou A, Schnitzer ME, Kennedy EH, et al. Doubly robust adaptive LASSO for effect modifier discovery. Submitted to journal.

[5] Tsiatis, A. *Semiparametric theory and missing data.* Springer Science & Business Media, 2007.

[6] Ahuja SD, Ashkin D, Avendano M, et al. Multidrug resistant pulmonary tuberculosis treatment regimens and patient outcomes: an individual patient data meta-analysis of 9,153 patients. *PLoS Med* 2012; **9**(8): e1001300.

[7] Benjamini Y and Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 1995; **57**: 289-300.

[8] Yekutieli D and Benjamini Y. Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *J Stat Plan Inference* 1999; **82**: 171-96.