Contents

1 Two-group definitions

1.1 Preference score definition

Let $\mathbf{X}_i$ be a covariate vector and A_i be a binary treatment indicator. Then, $e_i = E[A_i|\mathbf{X}_i]$ is the propensity score. Its expectation is the treatment prevalence by iterative expectation $p = E[e_i] = E[E[A_i|\mathbf{X}_i]] = E[A_i]$. Walker *et al* [1] defined the *preference score* as π_i that satisfied the following relationship.

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \log\left(\frac{e_i}{1-e_i}\right) - \log\left(\frac{p}{1-p}\right)$$

If we solve for π_i, we can obtain the following.

$$\begin{aligned}
\log\left(\frac{\pi_i}{1-\pi_i}\right) &= \log\left(\frac{e_i}{1-e_i}\right) - \log\left(\frac{p}{1-p}\right) \\
\log\left(\frac{\pi_i}{1-\pi_i}\right) &= \log\left(\frac{e_i}{1-e_i}\Big/\frac{p}{1-p}\right) \\
\frac{\pi_i}{1-\pi_i} &= \frac{e_i}{1-e_i}\Big/\frac{p}{1-p} \\
&= \frac{\frac{e_i}{p}}{\frac{1-e_i}{1-p}} \\
\pi_i &= \frac{\frac{\frac{e_i}{p}}{\frac{1-e_i}{1-p}}}{1+\frac{\frac{e_i}{p}}{\frac{1-e_i}{1-p}}} \\
&= \frac{\frac{e_i}{p}}{\frac{1-e_i}{1-p}+\frac{e_i}{p}}
\end{aligned}$$

This form gives insight into its re-centering property. When the treatment is rare, e_i is generally small. The numerator $\frac{e_i}{p}$ corrects this by dividing the generally small e_i with a small p. In particular, those individuals who happen to have the mean PS, *i.e.*, $e_i = p$, receive $pi_i = 0.5$. This transformation brings the "average individuals" to the center of the scale.

Also if we solve for e_i, we can obtain the following.

$$\begin{aligned}
\frac{\pi_i}{1-\pi_i} &= \frac{e_i}{1-e_i}\Big/\frac{p}{1-p} \\
\frac{\pi_i p}{(1-\pi_i)(1-p)} &= \frac{e_i}{1-e_i} \\
e_i &= \frac{\frac{\pi_i p}{(1-\pi_i)(1-p)}}{1+\frac{\pi_i p}{(1-\pi_i)(1-p)}} \\
&= \frac{\pi_i p}{(1-\pi_i)(1-p)+\pi_i p}
\end{aligned}$$

1.2 Intercept-adjustment interpretation of the preference score

Note

$$\begin{aligned}
\log\left(\frac{\pi_i}{1-\pi_i}\right) &= \log\left(\frac{e_i}{1-e_i}\right) - \log\left(\frac{p}{1-p}\right) \\
&= \log\left(\frac{P[A_i=1|\mathbf{X}_i]}{1-P[A_i=1|\mathbf{X}_i]}\right) - \log\left(\frac{P[A_i=1]}{1-P[A_i=1]}\right)
\end{aligned}$$

Assuming logistic models

$$\log\left(\frac{P[A_i=1|\mathbf{X}_i]}{1-P[A_i=1|\mathbf{X}_i]}\right) = \alpha_0 + \mathbf{X}_i^T\boldsymbol{\alpha}_\mathbf{x}$$

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \left[\alpha_0 - \log\left(\frac{p}{1-p}\right)\right] + \mathbf{X}_i^T\boldsymbol{\alpha}_\mathbf{x}$$

The last expression has the same form as the intercept-adjusted logistic regression used for risk prediction from a logistic regression fit on a case-control dataset [2]. It is known that a case-control logistic regression and the corresponding cohort logistic regression give the same coefficients except for the intercepts [3, 4]. The intercept terms have the following relationship.

$$\alpha_0^{cohort} = \alpha_0^{case-control} - \log\left(\frac{\tau_1}{\tau_0}\right)$$

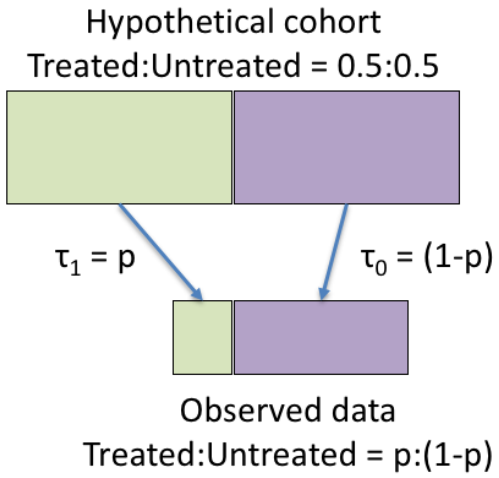
where

τ_1 = case sampling fraction

τ_0 = control sampling fraction

Intuitively, the case-control intercept is an overestimate because of the artificially high case prevalence in the case-control data. $\log\left(\frac{\tau_1}{\tau_0}\right) > 0$ if we oversample cases ($\tau_1 > \tau_0$).

We can consider the current study with a marginal treatment prevalence of p is a biased sample from a hypothetical population in which the covariate effects on the logit of treatment $\boldsymbol{\alpha}_\mathbf{x}$ are preserved but the marginal treatment prevalence is 0.5. The sampling fraction for the treated would be $\tau_1 = p$ and the sampling fraction for the untreated would be $\tau_1 = 1 - p$. We would obtain the desired ratio because $\frac{0.5p}{0.5(1-p)} = \frac{p}{1-p}$.



Under this framework, the initial PS model is the treatment assignment model for the biased sample with a treatment prevalence of p. The preference score model is the treatment assignment model for the super-population with a treatment prevalence of 0.5.

When the covariates have no role in determining treatment assignment (random treatment assignment), the right-hand side is always zero (preference score of 0.5) [1] because $P[A_i = 1|\mathbf{X}_i] = P[A_i = 1]$.

2 Multi-group definitions

2.1 Generalized preference score

Each generalized preference score is the following.

$$\pi_{ji} = \frac{\frac{e_{ji}}{p_j}}{\sum_{k=0}^{J} \frac{e_{ki}}{p_k}}$$

This expression came from the following proposed generalization of the defining equations (J simultaneous equations) using the baseline logit multinomial logistic regression in place of the binary logistic regression in the two-group definition.

$$\begin{gathered}
\text{For } j \in \{1, ..., J\} \\
\log\left(\frac{\pi_{ji}}{\pi_{0i}}\right) = \log\left(\frac{e_{ji}}{e_{0i}}\right) - \log\left(\frac{p_j}{p_0}\right) \\
\text{where} \\
\sum_{k=0}^{J} \pi_{ki} = 1
\end{gathered}$$

The sum constraint is necessary to maintain the interpretation as the prevalence-adjusted PS. For each $j \in \{1, ..., J\}$, we have the following.

$$\begin{aligned}
\log\left(\frac{\pi_{ji}}{\pi_{0i}}\right) &= \log\left(\frac{e_{ji}}{e_{0i}}\right) - \log\left(\frac{p_j}{p_0}\right) \\
&= \log\left(\frac{e_{ji}}{e_{0i}} \Big/ \frac{p_j}{p_0}\right) \\
\frac{\pi_{ji}}{\pi_{0i}} &= \frac{e_{ji}}{e_{0i}} \Big/ \frac{p_j}{p_0} \\
&= \frac{e_{ji}}{p_j} \frac{p_0}{e_{0i}}
\end{aligned}$$

First solve for π_{0i}.

$$\begin{gathered}
\text{Sum } J \text{ equations} \\
\sum_{j=1}^{J} \frac{\pi_{ji}}{\pi_{0i}} = \sum_{j=1}^{J} \frac{e_{ji}}{p_j} \frac{p_0}{e_{0i}} \\
\frac{\sum_{j=1}^{J} \pi_{ji}}{\pi_{0i}} = \sum_{j=1}^{J} \frac{e_{ji}}{p_j} \frac{p_0}{e_{0i}} \\
\text{By } \sum_{j=0}^{J} \pi_{ji} = 1 \\
\frac{1 - \pi_{0i}}{\pi_{0i}} = \sum_{j=1}^{J} \frac{e_{ji}}{p_j} \frac{p_0}{e_{0i}} \\
\frac{\pi_{0i}}{1 - \pi_{0i}} = \frac{1}{\sum_{j=1}^{J} \frac{e_{ji}}{p_j} \frac{p_0}{e_{0i}}}
\end{gathered}$$

$$
\begin{aligned}
\pi_{0i} &= \frac{\frac{1}{\sum_{j=1}^{J} \frac{e_{ji}}{p_j} \frac{p_0}{e_{0i}}}}{1 + \frac{1}{\sum_{j=1}^{J} \frac{e_{ji}}{p_j} \frac{p_0}{e_{0i}}}} \\
&= \frac{1}{1 + \sum_{j=1}^{J} \frac{e_{ji}}{p_j} \frac{p_0}{e_{0i}}} \\
&= \frac{\frac{e_{0i}}{p_0}}{\frac{e_{0i}}{p_0} + \sum_{j=1}^{J} \frac{e_{ji}}{p_j}} \\
&= \frac{\frac{e_{0i}}{p_0}}{\sum_{j=0}^{J} \frac{e_{ji}}{p_j}}
\end{aligned}
$$

Now solve for an arbitrary $j \in \{1, ..., J\}$.

$$
\begin{aligned}
\frac{\pi_{ji}}{\pi_{0i}} &= \frac{e_{ji}}{p_j} \frac{p_0}{e_{0i}} \\
\pi_{ji} &= \pi_{0i} \frac{e_{ji}}{p_j} \frac{p_0}{e_{0i}} \\
&= \pi_{0i} \frac{e_{ji}}{p_j} \frac{p_0}{e_{0i}} \\
&\qquad \text{Substitute } \pi_{0i} \\
&= \frac{\frac{e_{0i}}{p_0}}{\sum_{k=0}^{J} \frac{e_{ki}}{p_k}} \frac{e_{ji}}{p_j} \frac{p_0}{e_{0i}} \\
&= \frac{1}{\sum_{k=0}^{J} \frac{e_{ki}}{p_k}} \frac{e_{ji}}{p_j} \\
&= \frac{\frac{e_{ji}}{p_j}}{\sum_{k=0}^{J} \frac{e_{ki}}{p_k}}
\end{aligned}
$$

Taken together, for $j \in \{0, 1, ..., J\}$,

$$
\pi_{ji} = \frac{\frac{e_{ji}}{p_j}}{\sum_{k=0}^{J} \frac{e_{ki}}{p_k}}
$$

2.2 Rationale for region of empirical equipoise

By the proposed generalization, each subject has a *preference score vector* $\boldsymbol{\pi}_i$ with $J + 1$ elements π_{ji} where $j = 0, 1, ..., J$ and $\sum_{j=0}^{J+1} \pi_{ji} = 1$. Note the expectation of the corresponding *propensity score vector* $\mathbf{e}_i$ is the *treatment prevalence vector* $\mathbf{p}$ ($E[\mathbf{e}_i] = \mathbf{p}$).

# of Groups	Preference score space	Center of preference score space	Threshold
2	$[0,1]^2$	$\left(\frac{1}{2},\frac{1}{2}\right)^T$	0.30
3	$[0,1]^3$	$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)^T$	0.20
4	$[0,1]^4$	$\left(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\right)^T$	0.15
5	$[0,1]^5$	$\left(\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5}\right)^T$	0.12
6	$[0,1]^6$	$\left(\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6}\right)^T$	0.10
$\vdots$			
$J+1$	$[0,1]^{J+1}$	$\left(\frac{1}{J+1},\ldots,\frac{1}{J+1}\right)^T$	$\left(\frac{1}{J+1}\right)\left(\frac{3}{5}\right)$

An "average" individual with a PS vector agreeing with the treatment prevalence vector is given a preference score vector $\left(\frac{1}{J+1},\ldots,\frac{1}{J+1}\right)^T$. This is $\left(\frac{1}{2},\frac{1}{2}\right)^T$ in the two-group setting, $\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)^T$ in the three-group setting, and $\left(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4}\right)^T$ in the four-group setting.

Because of this change in the center of the preference score space, the threshold for defining the region for empirical equipoise assessment must adapt to the number of group. For example, the threshold of $\pi_{ji} > 0.3$ for all $j \in \{0,\ldots,J\}$ is not possible once there are four groups.

2.3 Proof that the generalized definition reduces to the original two-group definition

We can check this definition reduces to the original definition in the two-group setting as follows.

Preference score is recovered as follows.

$$\log\left(\frac{\pi_{1i}}{\pi_{0i}}\right) = \log\left(\frac{e_{1i}}{e_{0i}}\right) - \log\left(\frac{p_1}{p_0}\right)$$
$$\log\left(\frac{\pi_{1i}}{1-\pi_{1i}}\right) = \log\left(\frac{e_{1i}}{1-e_{1i}}\right) - \log\left(\frac{p_1}{1-p_1}\right)$$
$$\log\left(\frac{\pi_{i}}{1-\pi_{i}}\right) = \log\left(\frac{e_{i}}{1-e_{i}}\right) - \log\left(\frac{p}{1-p}\right)$$

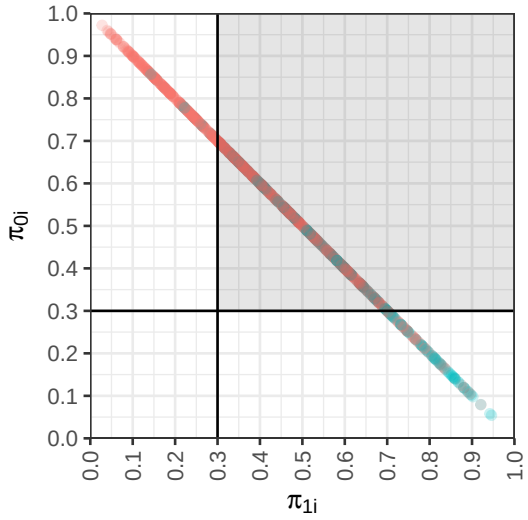
Let $I = \{1,...,n\}$ be the set of indices for n individuals in the entire cohort and $\alpha_{J,w}$ be the threshold proposed above. The index set for the individuals in the region of empirical equipoise is the following for the $J+1$ group setting.

$$I_{J,w} = \{i \in I : \pi_{ji} \geq \alpha_{J,w} \ \forall \ j \in \{0,...,J\}\}$$

We can show this expression reduces to the original two-group definition for $J = 1$ (two group setting).

$$\begin{aligned}
I_{1,w} &= \{i \in I : \pi_{ji} \geq \alpha_{J,w} \ \forall \ j \in \{0,1\}\} \\
&= \{i \in I : \pi_{0i} \geq \alpha_{J,w}, \pi_{1i} \geq \alpha_{J,w}\} \\
&\quad \text{Since } \pi_{0i} = 1 - \pi_{1i} \\
&= \{i \in I : 1 - \pi_{1i} \geq \alpha_{J,w}, \pi_{1i} \geq \alpha_{J,w}\} \\
&= \{i \in I : \pi_{1i} \leq 1 - \alpha_{J,w}, \pi_{1i} \geq \alpha_{J,w}\} \\
&= \{i \in I : \alpha_{J,w} \leq \pi_{1i} \leq 1 - \alpha_{J,w}\} \\
&= \{i \in I : \pi_{1i} \in [\alpha_{1,w}, 1 - \alpha_{1,w}]\} \\
&\quad \text{Note } \pi_{1i} = \pi_i \text{ (two-group preference score).} \\
&\quad \alpha_{1,w} = 0.3 \\
&= \{i \in I : \pi_i \in [0.3, 0.7]\} \\
&= \text{ original two-group definition}
\end{aligned}$$

If we visualize the two-group preference scores, we obtain a two-dimensional plot. However, because of the constraint that $\pi_{1i} + \pi_{0i} = 1$, all individuals (red group 0; blue group 1) appear on the diagonal line. That is, the information is one-dimensional, so we only need $\pi_i = \pi_{1i}$. With this visualization, we can see that individuals satisfy $\pi_{1i} \geq 0.3$ and $\pi_{0i} \geq 0.3$ (gray region) if and only if they satisfy $\pi_i = \pi_{1i} \in [0.3, 0.7]$.



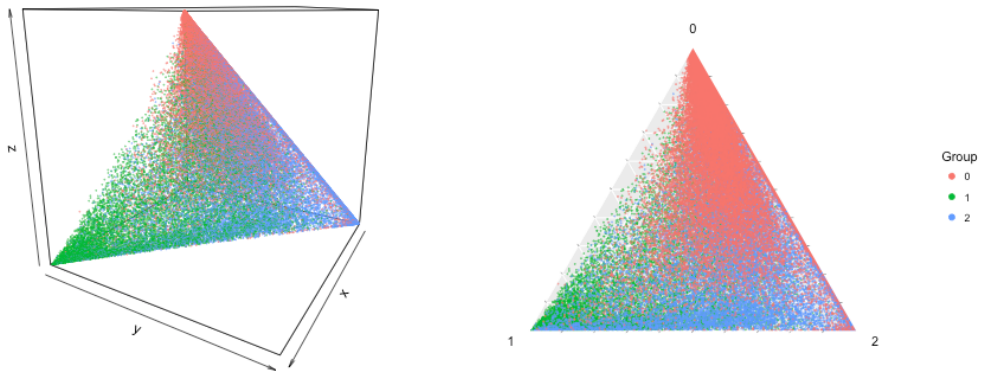
3 Empirical data demonstration

3.1 Visualization with a ternary plot

The generalized propensity score in the three-group setting is a vector of three elements $(e_{0i}, e_{1i}, e_{2i})^T$. The generalized preference score in the three-group setting is also a vector of three elements $(\pi_{0i}, \pi_{1i}, \pi_{2i})^T$. The following explanation is written in terms of the generalized propensity score, but the explanation is analogous for the generalized preference score.

As three dimensional data, individual subjects can be plotted in a three-dimensional cube $[0, 1]^3$ (left). The Z-axis represents e_{0i}, X-axis represents e_{1i}, and Y-axis represents e_{2i}. As seen in the three-dimensional plot (left), the points only occupy the diagonal triangular plane. This is because of the constraint $e_{0i} + e_{1i} + e_{2i} = 1$ for all i. In this case, we know what e_{2i} is as soon as we know e_{0i} and e_{1i}. That is, although the data are three-dimensional, the information carried is only two dimensional.

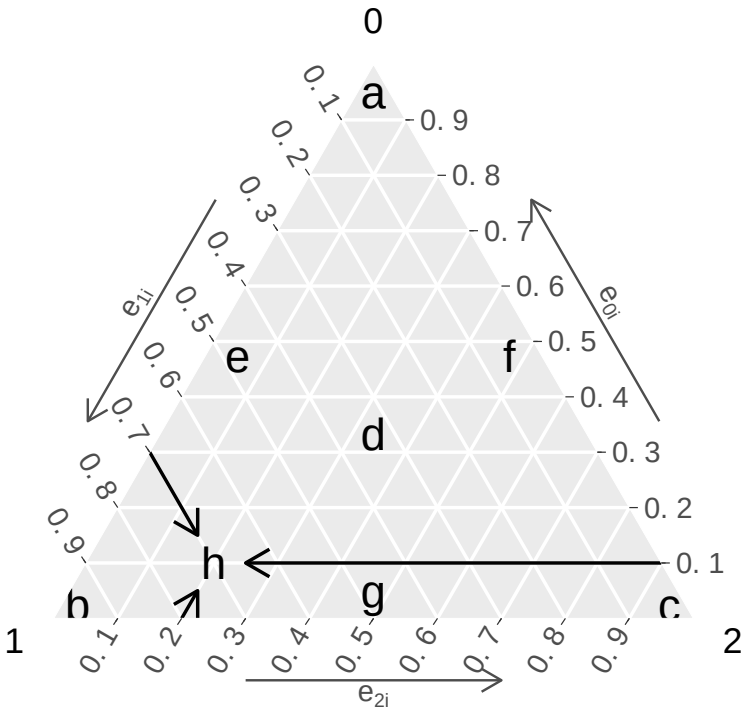
Therefore, we can take out this triangular plane in the left plot and represent as a two-dimensional plot (right). This two-dimensional representation is called a *ternary plot*. We used the `ggtern` R package for ternary plots [5].



The coordinate systems is explained here. The top corner of the triangle (a) is $\mathbf{e}_i = (1, 0, 0)$, *i.e.*, 100% probability of being in Group 0. The left lower corner (b) is $\mathbf{e}_i = (0, 1, 0)$ and the right lower corner (c) is $\mathbf{e}_i = (0, 0, 1)$. The mid-point in the triangle (d) is $\mathbf{e}_i = (1/3, 1/3, 1/3)$. That is, equal probability of being in any of the three groups. The mid points on the edges are: (e) $\mathbf{e}_i = (1/2, 1/2, 0)$, (f) $\mathbf{e}_i = (1/2, 0, 1/2)$, and (g) $\mathbf{e}_i = (0, 1/2, 1/2)$.

To look up point (h), all three axes have to be looked up. The e_{0i} axis is on the right edge. Use the horizontal guide lines because the labels (0.1, etc) are horizontal. Point (h) is at $e_{0i} = 0.1$. The e_{1i} axis is on the left edge. Use the guide lines going into the lower right direction as the labels indicate. Point (h) is at $e_{1i} = 0.7$. The e_{2i} axis is on the bottom edge. Use the guide lines going into the upper right direction as the labels indicate. Point (h) is at $e_{2i} = 0.2$. As a result, Point (h) is at $\mathbf{e}_i = (0.1, 0.7, 0.2)$.

We omitted the axis labels in the empirical examples since we did not need precise value lookup. The general intuition is that being far from a given corner, for example, the top corner labeled 0, means having a low probability of being in that group.



	e_{0i}	e_{1i}	e_{2i}
(a)	1	0	0
(b)	0	1	0
(c)	0	0	1
(d)	1/3	1/3	1/3
(e)	1/2	1/2	0
(f)	1/2	0	1/2
(g)	0	1/2	1/2
(h)	0.1	0.7	0.2

3.2 Non-selective non-steroidal anti-inflammatory drugs (nsNSAIDs) example

This dataset contained demographic and clinical including dispensing information on Medicare beneficiaries from Pennsylvania and New Jersey who qualified for pharmaceutical assistance programs for low-income older adults (January 1, 1999, through December 31, 2005) [6].

Individuals were required to have diagnoses for osteoarthritis or rheumatoid arthritis on two separate occasions and consistent use of health care services in the preceding 365 days. Those who had dispensing of analgesics within the preceding 180 days, those with malignancy, those using hospice services within the preceding 365 days, and those had simultaneous dispensing of multiple analgesics were secluded. The outcomes of interest of the original study included cardiovascular and gastrointestinal adverse events.

We chose three non-selective NSAIDs with different prevalence in the dataset for visual examination: naproxen, ibuprofen, and diclofenac. These non-selective NSAIDs were expected to have been used similarly in practice. This

example was used to illustrate the centering property of the generalized preference score in the presence of groups of different sizes. Generalized PSs were estimated with 38 predictor variables thought to be risk factors for any of several potential adverse effects of nsNSAIDs (**eTable 1** and **eFigure 1**).

3.3 Biological disease-modifying anti-rheumatic drugs (bDMARDs) example

This example was taken from more recent MarketScan data (2011-June 2015) of new users of biological disease-modifying anti-rheumatic drugs (DMARDs) [7, 8]. In the original studies, Kim *et al* [7] examined the tocilizumab vs tumor necrosis factor (TNF) inhibitor comparison and Kang *et al* [8] examined the abatacept TNF inhibitor comparison. Both studies used multiple data sources, but we focused on the MarketScan data for simplicity. Our three arms of interests were abatacept users, tocilizumab users, and TNF inhibitor users. Therefore, we re-extracted the datasets and combined such that we have three mutually exclusive groups.

Individuals were required to have two separate outpatient or one inpatient code for rheumatoid arthritis and initiation of the drugs of interest. The exclusion criteria were nursing home residents, patients with HIV/AIDS, patients with malignancy other than nonmelanoma skin cancer, and those with end-stage renal disease including use of dialysis or renal transplant. The outcome of interest of the original studies was composite cardiovascular events.

The most up-to-date recommendations list these three classes of bDMARDs as equally indicated [9, 10]. However, TNF inhibitors, by the virtue of being the first biological DMARDs to come on the market, were more often used first. On the other hand, tocilizumab and abatacept were market-approved more recently in the U.S. market, and thus, were more commonly used as subsequent biological DMARDs after failure of one or more biological DMARDs.

Therefore, first-line tocilizumab and abatacept users were expected to be somewhat atypical patients compared to first-line TNFi users, whereas users were expected to be more similar when using these agents as a second-line bDMARD. A second-line TNFi after one TNFi means that there was a switch from one specific agent to another within the five-member TNFi class (adalimumab, certolizumab pegol, etanercept, golimumab, infliximab).

4 Simulation: methodological details

4.1 Data generating mechanism

In all scenarios, our sample size was n = 6,000.

4.1.1 Covariate generation

Latent covariates Z_{1i} through Z_{7i} were generated from a multivariate normal distribution to induce a given level of correlation $\rho \in \{0, 0.1, 0.3, 0.5, 0.7, 0.9\}$.

$$\begin{bmatrix} Z_{1i} \\ Z_{2i} \\ Z_{3i} \\ Z_{4i} \\ Z_{5i} \\ Z_{6i} \\ Z_{7i} \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho^1 & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 \\ \rho^1 & 1 & \rho^1 & \rho^2 & \rho^3 & \rho^4 & \rho^5 \\ \rho^2 & \rho^1 & 1 & \rho^1 & \rho^2 & \rho^3 & \rho^4 \\ \rho^3 & \rho^2 & \rho^1 & 1 & \rho^1 & \rho^2 & \rho^3 \\ \rho^4 & \rho^3 & \rho^2 & \rho^1 & 1 & \rho^1 & \rho^2 \\ \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho^1 & 1 & \rho^1 \\ \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho^1 & 1 \end{bmatrix} \right)$$

This means each Z_{ji} was a standard normal marginally. The correlation of Z_{ji} and Z_{ki} for $j \neq k$ was $\rho^{|j-k|}$. These latent variables were then transformed as follows.

$$\begin{aligned} X_{1i} &:= Z_{1i} \\ X_{2i} &:= F^{-1}_{Pois,1}(\Phi(Z_{2i})) \\ X_{3i} &:= F^{-1}_{Bern,0.2}(\Phi(Z_{3i})) \\ X_{4i} &:= F^{-1}_{Bern,0.2}(\Phi(Z_{4i})) \\ X_{5i} &:= F^{-1}_{Bern,0.2}(\Phi(Z_{5i})) \\ X_{6i} &:= F^{-1}_{Bern,0.2}(\Phi(Z_{6i})) \\ X_{7i} &:= Z_{7i} \end{aligned}$$

$\Phi(\cdot)$ was the standard normal cumulative distribution function (`pnorm(x, mean = 0, sd = 1)` in `R`). $F^{-1}_{Pois,1}(\cdot)$ was the inverse distribution function for a Poisson distribution with a rate parameter of 1 (`qpois(p, lambda = 1)` in `R`). $F^{-1}_{Bern,0.2}(\cdot)$ was the inverse distribution function for a Bernoulli distribution with a success probability of 0.2 (`qbinom(p, size = 1, prob = 0.2)` in `R`). The first transformation gave a Uniform(0,1) variable, and the second transformation gave a random variable with the desired distribution. The correlation structure was preserved in X_{1i} through X_{7i} using this two-step covariate generation.

4.1.2 Treatment generation

Treatment A_i was assigned based on all covariates $\mathbf{X}_i = (X_{1i}, \ldots, X_{7i})^T$.

Linear predictors

$$\begin{cases} \eta_{A1i} = \log\left(\dfrac{P[A_i = 1|\mathbf{X}_i]}{P[A_i = 0|\mathbf{X}_i]}\right) = \alpha_{01} + \mathbf{X}_i^T \boldsymbol{\alpha}_{X1} \\ \eta_{A2i} = \log\left(\dfrac{P[A_i = 2|\mathbf{X}_i]}{P[A_i = 0|\mathbf{X}_i]}\right) = \alpha_{02} + \mathbf{X}_i^T \boldsymbol{\alpha}_{X2} \end{cases}$$

True propensity scores

$$\begin{cases} e_{0i} = P(A_i = 0|\mathbf{X}_i) = \dfrac{1}{1 + \exp(\eta_{A1i}) + \exp(\eta_{A2i})} \\ e_{1i} = P(A_i = 1|\mathbf{X}_i) = \dfrac{\exp(\eta_{A1i})}{1 + \exp(\eta_{A1i}) + \exp(\eta_{A2i})} \\ e_{2i} = P(A_i = 2|\mathbf{X}_i) = \dfrac{\exp(\eta_{A2i})}{1 + \exp(\eta_{A1i}) + \exp(\eta_{A2i})} \end{cases}$$

Treatment assignment

$$A_i \in \{0, 1, 2\} \sim \text{Multinomial}\left((e_{0i}, e_{1i}, e_{2i})^T, 1\right)$$

The treatment model parameter values are in the following table.

- The Size column is the treatment prevalence setting.
- The "RelX7" column corresponds to the "Relative treatment association of X7" in the figures, the strength of the treatment association of X_7 relative to X_1 through X_6.
- The "Equipoise" column corresponds to the "Level of equipoise" in the figures. "Perfect" indicates no covariate effect on treatments (randomized treatment). Increasing levels of covariate effects were introduced for "Good", "Moderate", and "Poor" as seen in the magnitude of coefficients.
- The alternating rows correspond to the first and second linear predictors (See Contrast column).
- Column 0 corresponds to the intercept coefficient. Columns 1 through 7 correspond to the coefficients for X_1 through X_7.

Number	Size	RelX7	Equipoise	Contrast	0	1	2	3	4	5	6	7
1	33:33:33	Zero	Perfect	1vs0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	33:33:33	Zero	Perfect	2vs0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	33:33:33	Zero	Good	1vs0	-0.34	0.25	0.25	0.25	0.25	0.25	0.25	0.00
3	33:33:33	Zero	Good	2vs0	-1.00	0.50	0.50	0.50	0.50	0.50	0.50	0.00
4	33:33:33	Zero	Moderate	1vs0	-0.40	0.50	0.50	0.50	0.50	0.50	0.50	0.00
4	33:33:33	Zero	Moderate	2vs0	-1.70	1.00	1.00	1.00	1.00	1.00	1.00	0.00
5	33:33:33	Zero	Poor	1vs0	-0.40	1.00	1.00	1.00	1.00	1.00	1.00	0.00

Continued

Number	Size	RelX7	Equipoise	Contrast	0	1	2	3	4	5	6	7
5	33:33:33	Zero	Poor	2vs0	-3.10	2.00	2.00	2.00	2.00	2.00	2.00	0.00
6	33:33:33	Half	Perfect	1vs0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	33:33:33	Half	Perfect	2vs0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	33:33:33	Half	Good	1vs0	-0.34	0.25	0.25	0.25	0.25	0.25	0.25	0.12
8	33:33:33	Half	Good	2vs0	-1.00	0.50	0.50	0.50	0.50	0.50	0.50	0.25
9	33:33:33	Half	Moderate	1vs0	-0.40	0.50	0.50	0.50	0.50	0.50	0.50	0.25
9	33:33:33	Half	Moderate	2vs0	-1.70	1.00	1.00	1.00	1.00	1.00	1.00	0.50
10	33:33:33	Half	Poor	1vs0	-0.40	1.00	1.00	1.00	1.00	1.00	1.00	0.50
10	33:33:33	Half	Poor	2vs0	-3.10	2.00	2.00	2.00	2.00	2.00	2.00	1.00
11	33:33:33	Same	Perfect	1vs0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11	33:33:33	Same	Perfect	2vs0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13	33:33:33	Same	Good	1vs0	-0.34	0.25	0.25	0.25	0.25	0.25	0.25	0.25
13	33:33:33	Same	Good	2vs0	-1.00	0.50	0.50	0.50	0.50	0.50	0.50	0.50
14	33:33:33	Same	Moderate	1vs0	-0.40	0.50	0.50	0.50	0.50	0.50	0.50	0.50
14	33:33:33	Same	Moderate	2vs0	-1.70	1.00	1.00	1.00	1.00	1.00	1.00	1.00
15	33:33:33	Same	Poor	1vs0	-0.40	1.00	1.00	1.00	1.00	1.00	1.00	1.00
15	33:33:33	Same	Poor	2vs0	-3.10	2.00	2.00	2.00	2.00	2.00	2.00	2.00
16	33:33:33	Twice	Perfect	1vs0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
16	33:33:33	Twice	Perfect	2vs0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18	33:33:33	Twice	Good	1vs0	-0.34	0.25	0.25	0.25	0.25	0.25	0.25	0.50
18	33:33:33	Twice	Good	2vs0	-1.00	0.50	0.50	0.50	0.50	0.50	0.50	1.00
19	33:33:33	Twice	Moderate	1vs0	-0.40	0.50	0.50	0.50	0.50	0.50	0.50	1.00
19	33:33:33	Twice	Moderate	2vs0	-1.70	1.00	1.00	1.00	1.00	1.00	1.00	2.00
20	33:33:33	Twice	Poor	1vs0	-0.40	1.00	1.00	1.00	1.00	1.00	1.00	2.00
20	33:33:33	Twice	Poor	2vs0	-3.10	2.00	2.00	2.00	2.00	2.00	2.00	4.00
21	10:45:45	Zero	Perfect	1vs0	1.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21	10:45:45	Zero	Perfect	2vs0	1.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
23	10:45:45	Zero	Good	1vs0	0.90	0.25	0.25	0.25	0.25	0.25	0.25	0.00
23	10:45:45	Zero	Good	2vs0	0.40	0.50	0.50	0.50	0.50	0.50	0.50	0.00
24	10:45:45	Zero	Moderate	1vs0	1.50	0.50	0.50	0.50	0.50	0.50	0.50	0.00
24	10:45:45	Zero	Moderate	2vs0	0.50	1.00	1.00	1.00	1.00	1.00	1.00	0.00
25	10:45:45	Zero	Poor	1vs0	1.70	1.00	1.00	1.00	1.00	1.00	1.00	0.00
25	10:45:45	Zero	Poor	2vs0	-0.30	2.00	2.00	2.00	2.00	2.00	2.00	0.00
26	10:45:45	Half	Perfect	1vs0	1.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
26	10:45:45	Half	Perfect	2vs0	1.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
28	10:45:45	Half	Good	1vs0	0.90	0.25	0.25	0.25	0.25	0.25	0.25	0.12
28	10:45:45	Half	Good	2vs0	0.40	0.50	0.50	0.50	0.50	0.50	0.50	0.25
29	10:45:45	Half	Moderate	1vs0	1.50	0.50	0.50	0.50	0.50	0.50	0.50	0.25
29	10:45:45	Half	Moderate	2vs0	0.50	1.00	1.00	1.00	1.00	1.00	1.00	0.50
30	10:45:45	Half	Poor	1vs0	1.70	1.00	1.00	1.00	1.00	1.00	1.00	0.50
30	10:45:45	Half	Poor	2vs0	-0.30	2.00	2.00	2.00	2.00	2.00	2.00	1.00
31	10:45:45	Same	Perfect	1vs0	1.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
31	10:45:45	Same	Perfect	2vs0	1.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
33	10:45:45	Same	Good	1vs0	0.90	0.25	0.25	0.25	0.25	0.25	0.25	0.25
33	10:45:45	Same	Good	2vs0	0.40	0.50	0.50	0.50	0.50	0.50	0.50	0.50
34	10:45:45	Same	Moderate	1vs0	1.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
34	10:45:45	Same	Moderate	2vs0	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00
35	10:45:45	Same	Poor	1vs0	1.70	1.00	1.00	1.00	1.00	1.00	1.00	1.00
35	10:45:45	Same	Poor	2vs0	-0.30	2.00	2.00	2.00	2.00	2.00	2.00	2.00
36	10:45:45	Twice	Perfect	1vs0	1.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
36	10:45:45	Twice	Perfect	2vs0	1.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
38	10:45:45	Twice	Good	1vs0	0.90	0.25	0.25	0.25	0.25	0.25	0.25	0.50
38	10:45:45	Twice	Good	2vs0	0.40	0.50	0.50	0.50	0.50	0.50	0.50	1.00
39	10:45:45	Twice	Moderate	1vs0	1.50	0.50	0.50	0.50	0.50	0.50	0.50	1.00
39	10:45:45	Twice	Moderate	2vs0	0.50	1.00	1.00	1.00	1.00	1.00	1.00	2.00
40	10:45:45	Twice	Poor	1vs0	1.70	1.00	1.00	1.00	1.00	1.00	1.00	2.00
40	10:45:45	Twice	Poor	2vs0	-0.30	2.00	2.00	2.00	2.00	2.00	2.00	4.00
41	10:10:80	Zero	Perfect	1vs0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
41	10:10:80	Zero	Perfect	2vs0	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
43	10:10:80	Zero	Good	1vs0	0.00	0.25	0.25	0.25	0.25	0.25	0.25	0.00

Continued

Number	Size	RelX7	Equipoise	Contrast	0	1	2	3	4	5	6	7
43	10:10:80	Zero	Good	2vs0	1.70	0.50	0.50	0.50	0.50	0.50	0.50	0.00
44	10:10:80	Zero	Moderate	1vs0	0.10	0.50	0.50	0.50	0.50	0.50	0.50	0.00
44	10:10:80	Zero	Moderate	2vs0	1.60	1.00	1.00	1.00	1.00	1.00	1.00	0.00
45	10:10:80	Zero	Poor	1vs0	0.80	1.00	1.00	1.00	1.00	1.00	1.00	0.00
45	10:10:80	Zero	Poor	2vs0	2.00	2.00	2.00	2.00	2.00	2.00	2.00	0.00
46	10:10:80	Half	Perfect	1vs0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
46	10:10:80	Half	Perfect	2vs0	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
48	10:10:80	Half	Good	1vs0	0.00	0.25	0.25	0.25	0.25	0.25	0.25	0.12
48	10:10:80	Half	Good	2vs0	1.70	0.50	0.50	0.50	0.50	0.50	0.50	0.25
49	10:10:80	Half	Moderate	1vs0	0.10	0.50	0.50	0.50	0.50	0.50	0.50	0.25
49	10:10:80	Half	Moderate	2vs0	1.60	1.00	1.00	1.00	1.00	1.00	1.00	0.50
50	10:10:80	Half	Poor	1vs0	0.80	1.00	1.00	1.00	1.00	1.00	1.00	0.50
50	10:10:80	Half	Poor	2vs0	2.00	2.00	2.00	2.00	2.00	2.00	2.00	1.00
51	10:10:80	Same	Perfect	1vs0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
51	10:10:80	Same	Perfect	2vs0	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
53	10:10:80	Same	Good	1vs0	0.00	0.25	0.25	0.25	0.25	0.25	0.25	0.25
53	10:10:80	Same	Good	2vs0	1.70	0.50	0.50	0.50	0.50	0.50	0.50	0.50
54	10:10:80	Same	Moderate	1vs0	0.10	0.50	0.50	0.50	0.50	0.50	0.50	0.50
54	10:10:80	Same	Moderate	2vs0	1.60	1.00	1.00	1.00	1.00	1.00	1.00	1.00
55	10:10:80	Same	Poor	1vs0	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00
55	10:10:80	Same	Poor	2vs0	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
56	10:10:80	Twice	Perfect	1vs0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
56	10:10:80	Twice	Perfect	2vs0	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
58	10:10:80	Twice	Good	1vs0	0.00	0.25	0.25	0.25	0.25	0.25	0.25	0.50
58	10:10:80	Twice	Good	2vs0	1.70	0.50	0.50	0.50	0.50	0.50	0.50	1.00
59	10:10:80	Twice	Moderate	1vs0	0.10	0.50	0.50	0.50	0.50	0.50	0.50	1.00
59	10:10:80	Twice	Moderate	2vs0	1.60	1.00	1.00	1.00	1.00	1.00	1.00	2.00
60	10:10:80	Twice	Poor	1vs0	0.80	1.00	1.00	1.00	1.00	1.00	1.00	2.00
60	10:10:80	Twice	Poor	2vs0	2.00	2.00	2.00	2.00	2.00	2.00	2.00	4.00

4.1.3 Outcome generation

The linear predictor (log rate) for the Poisson count outcome was assigned based on all covariates and treatment. The log link was used to avoid the issue of non-collapsibility of the logit link [11].

$$\begin{aligned}\eta_{Yi} &= \beta_0 + \beta_{A1} I(A_i = 1) + \beta_{A2} I(A_i = 2) \\ &\quad + \mathbf{X}_i^T \boldsymbol{\beta}_X + I(A_i = 1)\mathbf{X}_i^T \boldsymbol{\beta}_{XA1} + I(A_i = 2)\mathbf{X}_i^T \boldsymbol{\beta}_{XA2}\end{aligned}$$

$$Y_i \sim \text{Poisson}\left(\exp(\eta_{Yi})\right)$$

Additionally, the following counterfactual log rates were kept for use in calculating the marginal causal effects.

$$\begin{aligned}\eta_{Y_i^0} &= \beta_0 + \mathbf{X}_i^T \boldsymbol{\beta}_X \\ \eta_{Y_i^1} &= \beta_0 + \beta_{A1} + \mathbf{X}_i^T \boldsymbol{\beta}_X + \mathbf{X}_i^T \boldsymbol{\beta}_{XA1} \\ \eta_{Y_i^2} &= \beta_0 + \beta_{A2} + \mathbf{X}_i^T \boldsymbol{\beta}_X + \mathbf{X}_i^T \boldsymbol{\beta}_{XA2}\end{aligned}$$

The outcome model parameter values were the following (RR: rate ratio).

$$\beta_0 = \log(0.20) \quad \text{Baseline rate}$$

$$(\beta_{A1}, \beta_{A2}) = (\log(1.0), \log(1.0)) \quad \text{Null main effects}$$

$$\boldsymbol{\beta}_X^T = \begin{cases} (\log(1.2), \log(1.2), \log(1.2), \log(1.2), \log(1.2), \log(1.2), \log(1.2)) & X_7 - Y \text{ RR } 1.2 \\ (\log(1.2), \log(1.2), \log(1.2), \log(1.2), \log(1.2), \log(1.2), \log(1.5)) & X_7 - Y \text{ RR } 1.5 \\ (\log(1.2), \log(1.2), \log(1.2), \log(1.2), \log(1.2), \log(1.2), \log(2.0)) & X_7 - Y \text{ RR } 2.0 \end{cases}$$

$$\begin{bmatrix} \boldsymbol{\beta}_{XA1}^T \\ \boldsymbol{\beta}_{XA2}^T \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{No effect modification}$$

4.2 Estimands of interest

Four outcome analyses were conducted. The first was the unadjusted analysis. The other three were weighted analyses with inverse probability of treatment weights (IPTW) [12], matching weights (MW)[13, 14], and overlap weights [15, 16, 17].

$$IPTW_i = \frac{1}{\sum_{j=0}^{2} I(A_i = j) e_{ji}}$$

$$MW_i = \frac{\min(e_{0i}, e_{1i}, e_{2i})}{\sum_{j=0}^{2} I(A_i = j) e_{ji}}$$

$$OW_i = \frac{\frac{1}{\frac{1}{e_{0i}} + \frac{1}{e_{1i}} + \frac{1}{e_{2i}}}}{\sum_{j=0}^{2} I(A_i = j) e_{ji}}$$

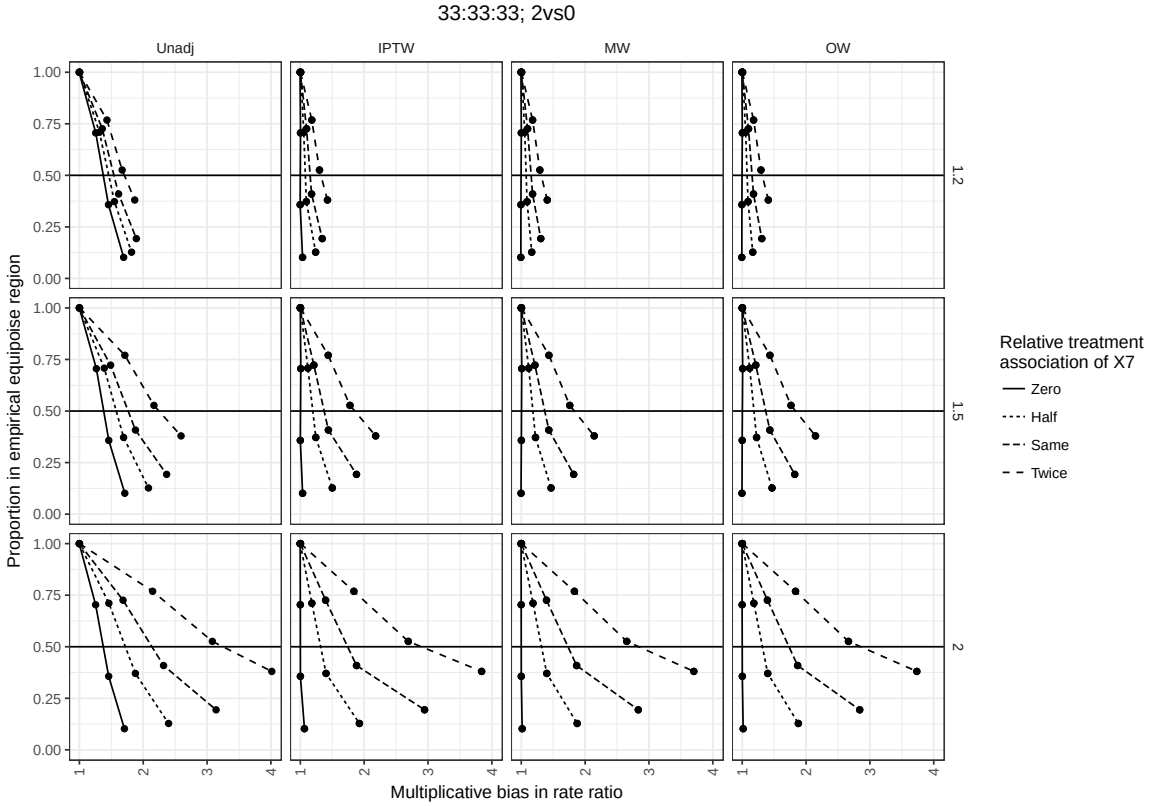
where $I(\cdot)$ is an indicator function that is 1 if the expression inside holds and 0 if not.

5 Simulation: additional results

5.1 Additional results

5.1.1 Group 2 vs. 0 contrast

For the group 2 vs. 0 contrast (Figure 4), which was designed to have more different covariate distributions, greater levels of residual bias were observed than in Figure 3.

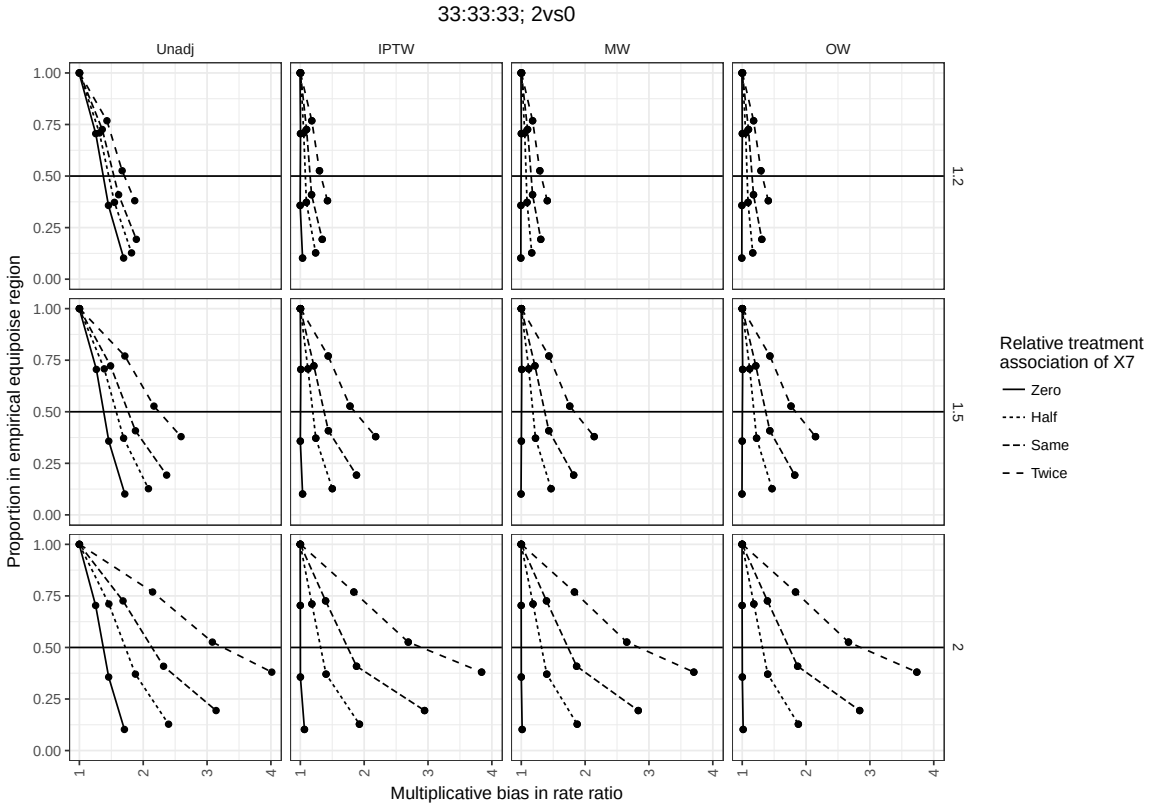


The columns of panels denote different confounding adjustment methods. The rows of panels denote different levels of associations between X7 (unmeasured covariate) and outcome. A rate ratio of 1.2 was the same strength of association as

the measured covariates, whereas only X7 had a stronger outcome association at a rate ratio of 1.5 and 2.0. In each panel, the X-axis represents the multiplicative bias in RR estimates, whereas the Y-axis represents the average proportion of the simulated cohorts within the region of empirical equipoise (overall proportion). The line types denote different levels of associations between X7 and treatment relative to the associations between measured variables and treatment.
Abbreviations: Unadj.: unadjusted; IPTW: inverse probability of treatment weights; MW: matching weights; OW: overlap weights.

5.1.2 Group 2 vs. 1 contrast

The group 2 vs. 1 contrast (Figure 5) gave similar results to the group 1 vs. 0 contrast.



The columns of panels denote different confounding adjustment methods. The rows of panels denote different levels of associations between X7 (unmeasured covariate) and outcome. A rate ratio of 1.2 was the same strength of association as the measured covariates, whereas only X7 had a stronger outcome association at a rate ratio of 1.5 and 2.0. In each panel, the X-axis represents the multiplicative bias in RR estimates, whereas the Y-axis represents the average proportion of the simulated cohorts within the region of empirical equipoise (overall proportion). The line types denote different levels of associations between X7 and treatment relative to the associations between measured variables and treatment.
Abbreviations: Unadj.: unadjusted; IPTW: inverse probability of treatment weights; MW: matching weights; OW: overlap weights.

5.1.3 Consideration of 75% threshold

In all contrasts (Figures 3 and above two figures), using a threshold of 75% instead of 50% would lead to a smaller range of biases although this comes at the cost of disregarding study design where the unmeasured variable indeed had weaker associations than measured ones.

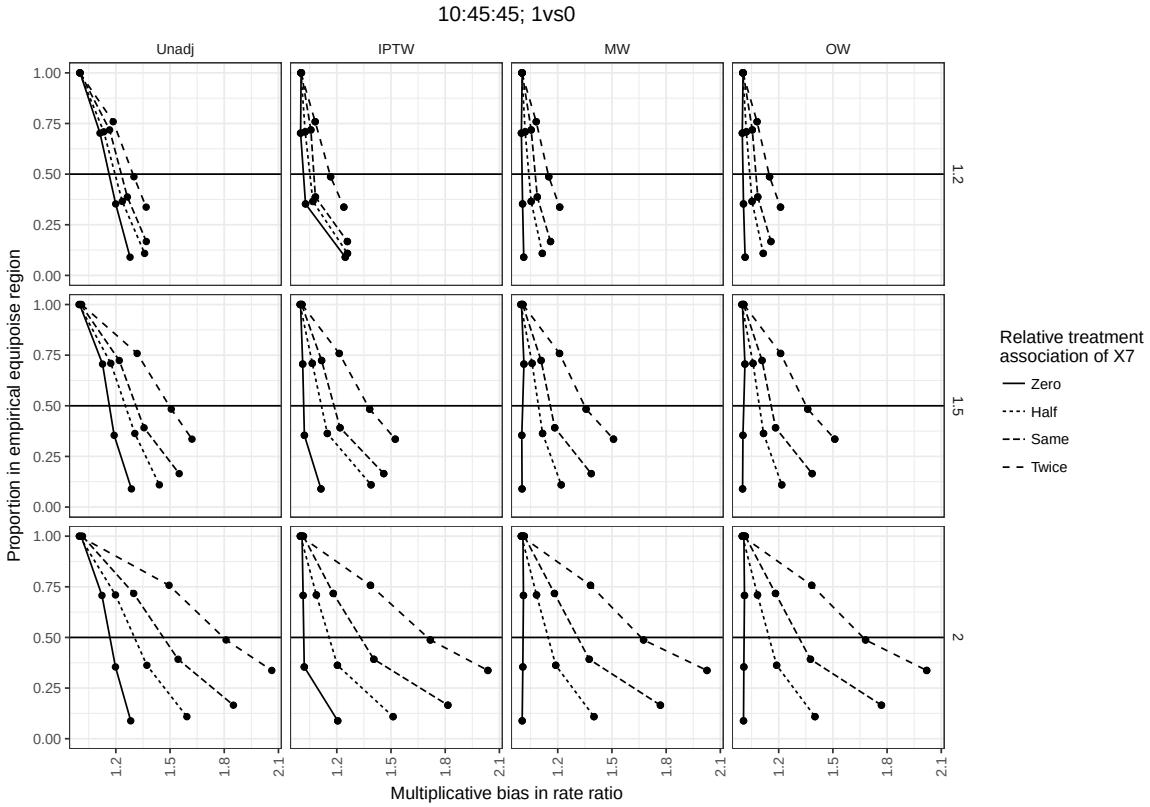
5.1.4 Summary of further results

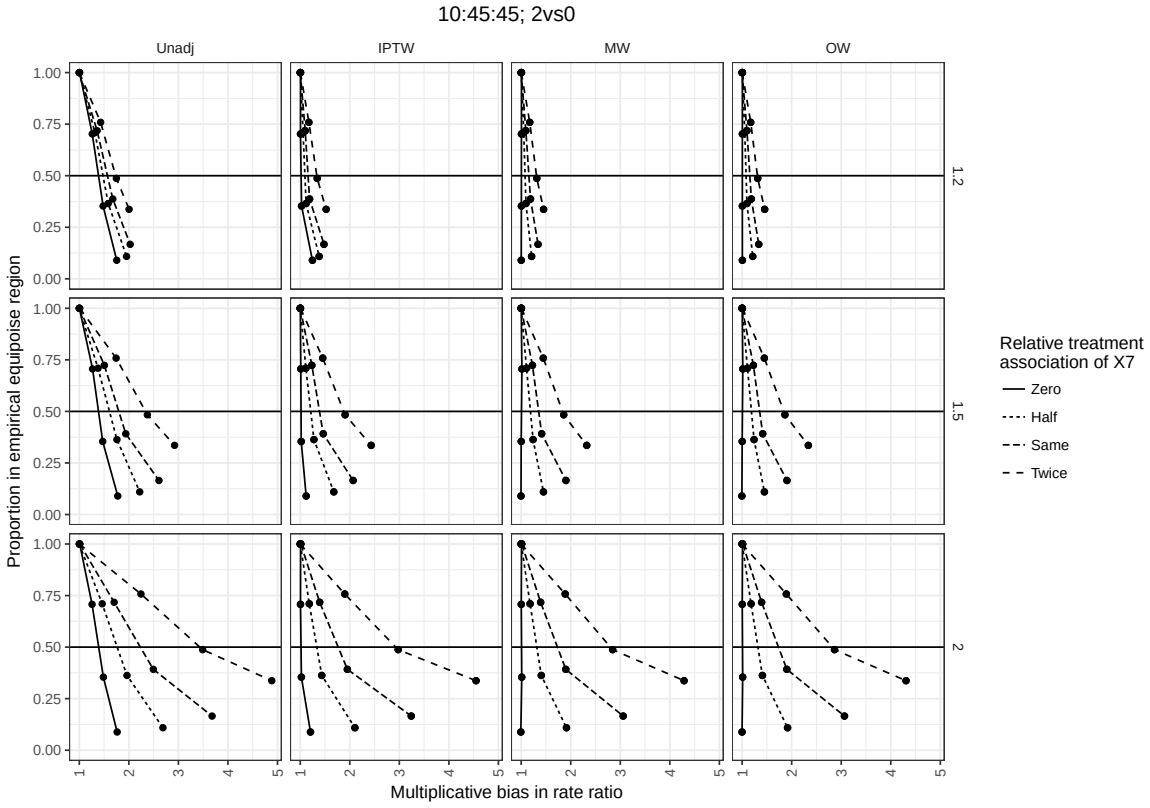
The results were similar when we varied treatment prevalence (eAppendix 5.2) and when we switched the assessment metric to the group-specific proportion (eAppendix 5.3). Also, the results were invariant with increasing correlation among covariates except in the very extreme setting with $\rho = 0.9$, in which the residual confounding was reduced by surrogacy via highly correlated measured variables. (eAppendix 5.4).

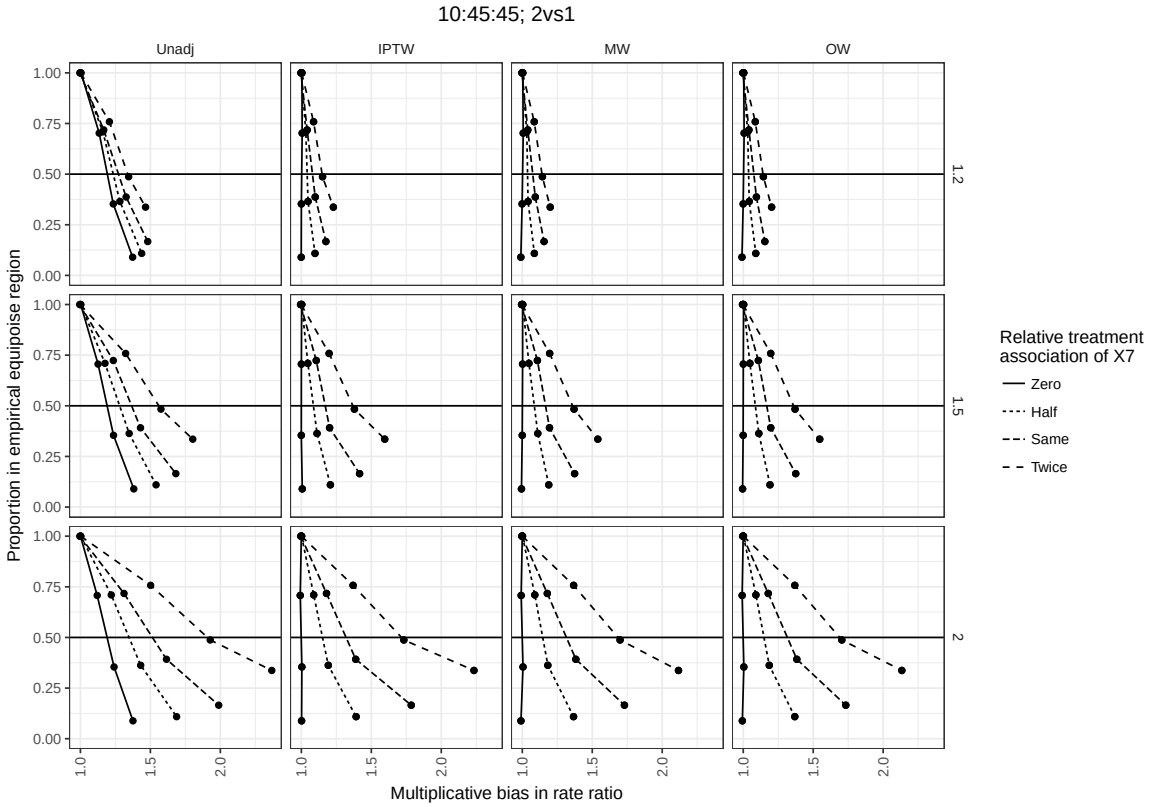
5.2 Overall proportion as the summary measure of empirical equipoise

In the following additional results, the index was the proportion of the overall cohort that fell into the proposed empirical equipoise region.

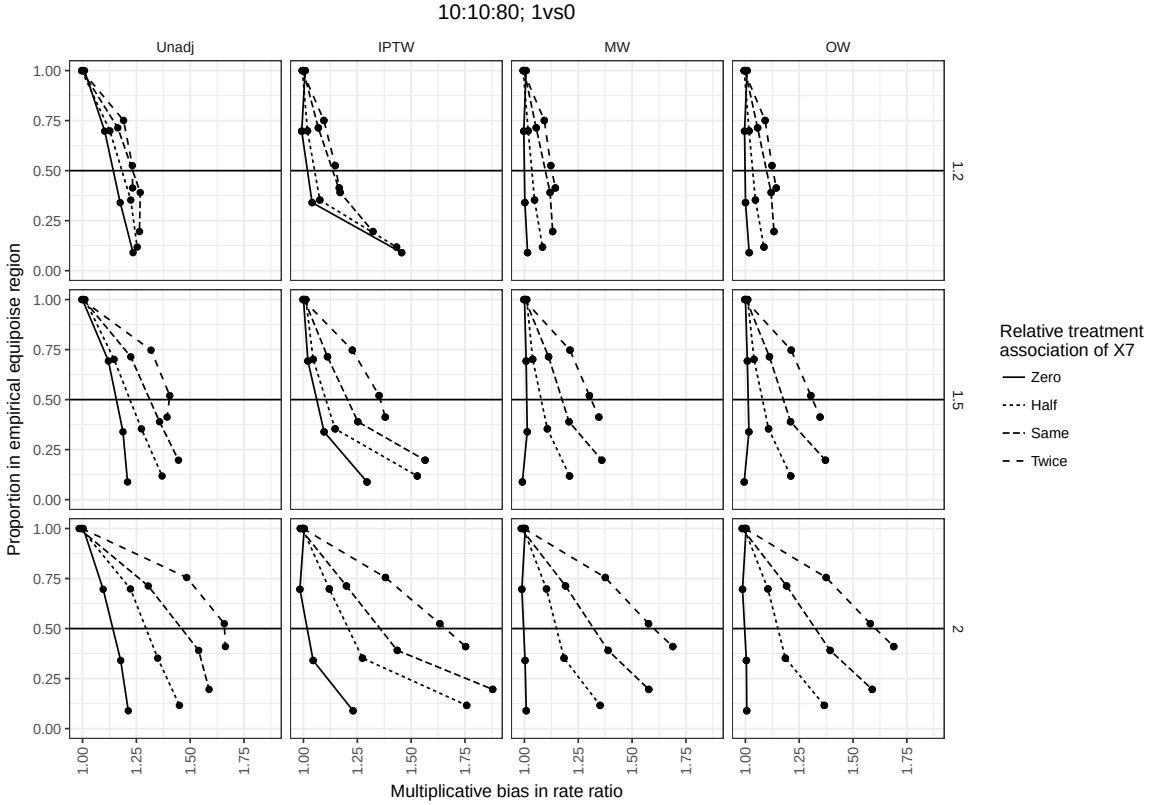
5.2.1 Unequal group sizes 10:45:45

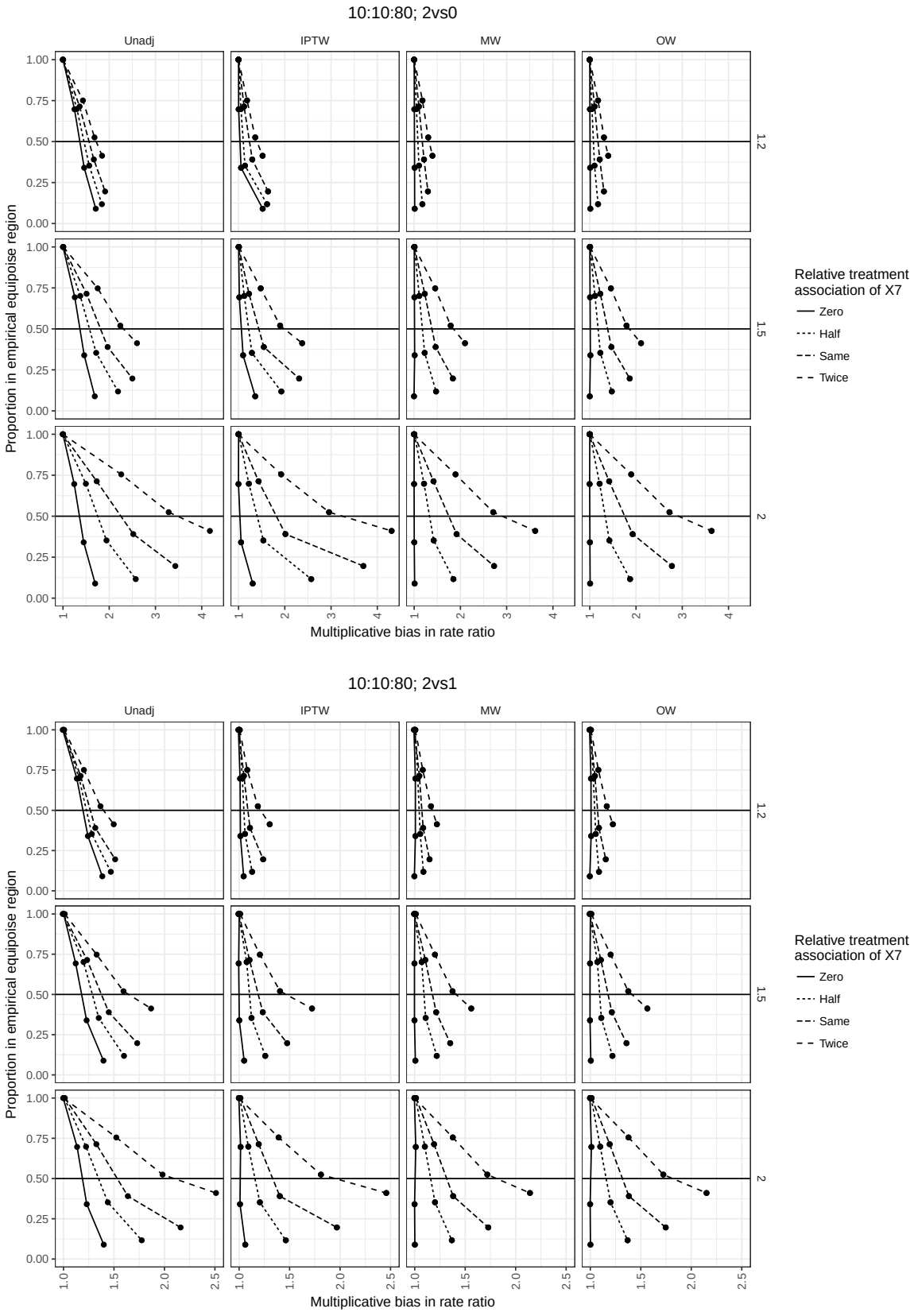






5.2.2 Unequal group sizes 10:10:80

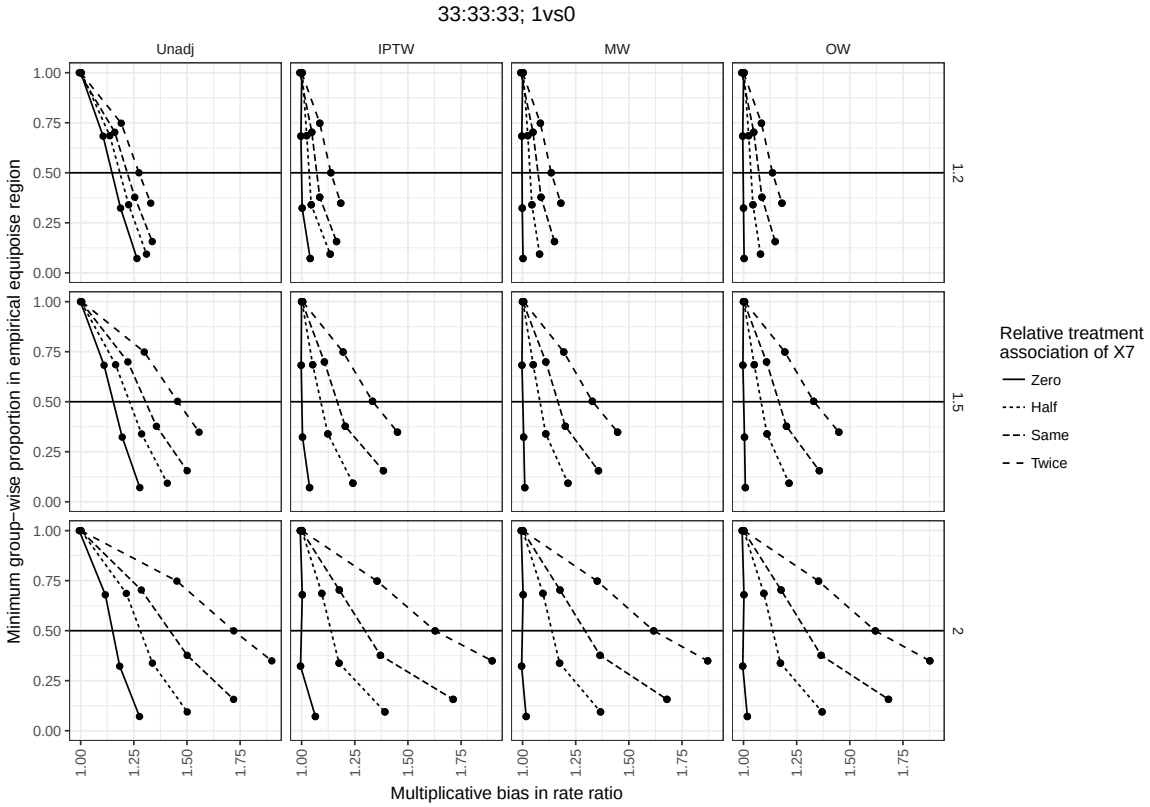


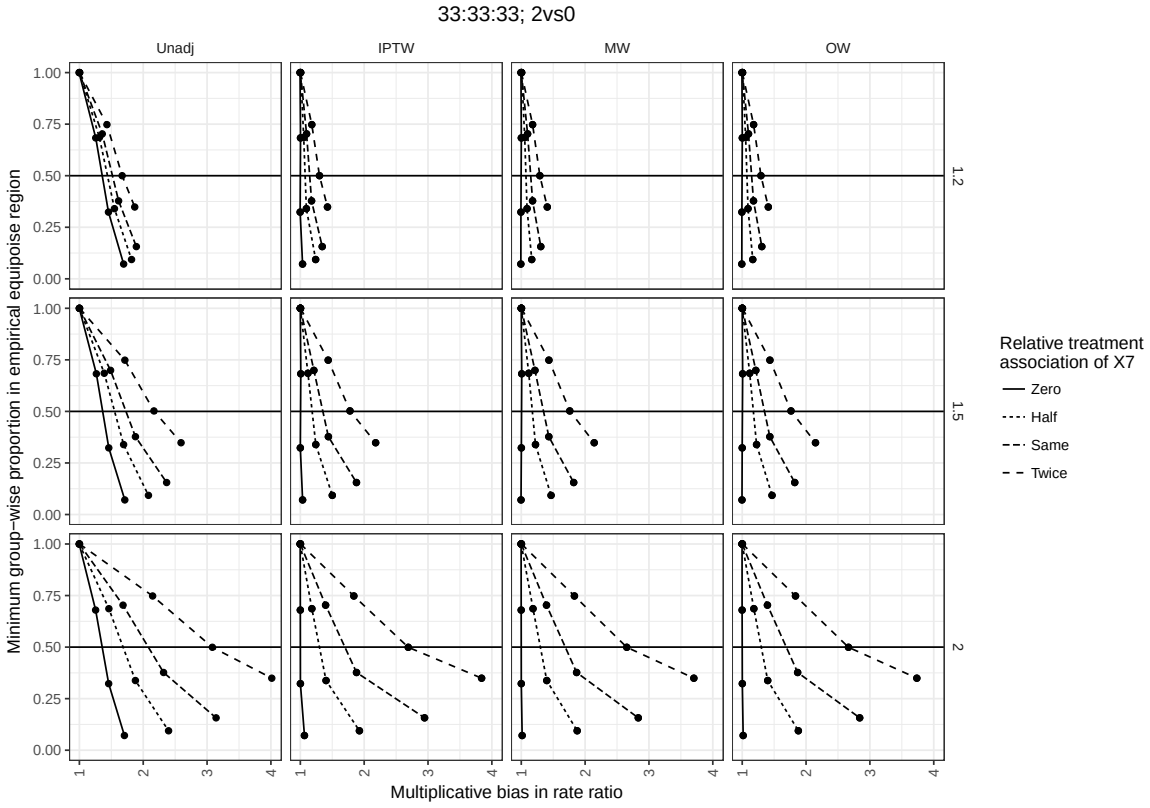


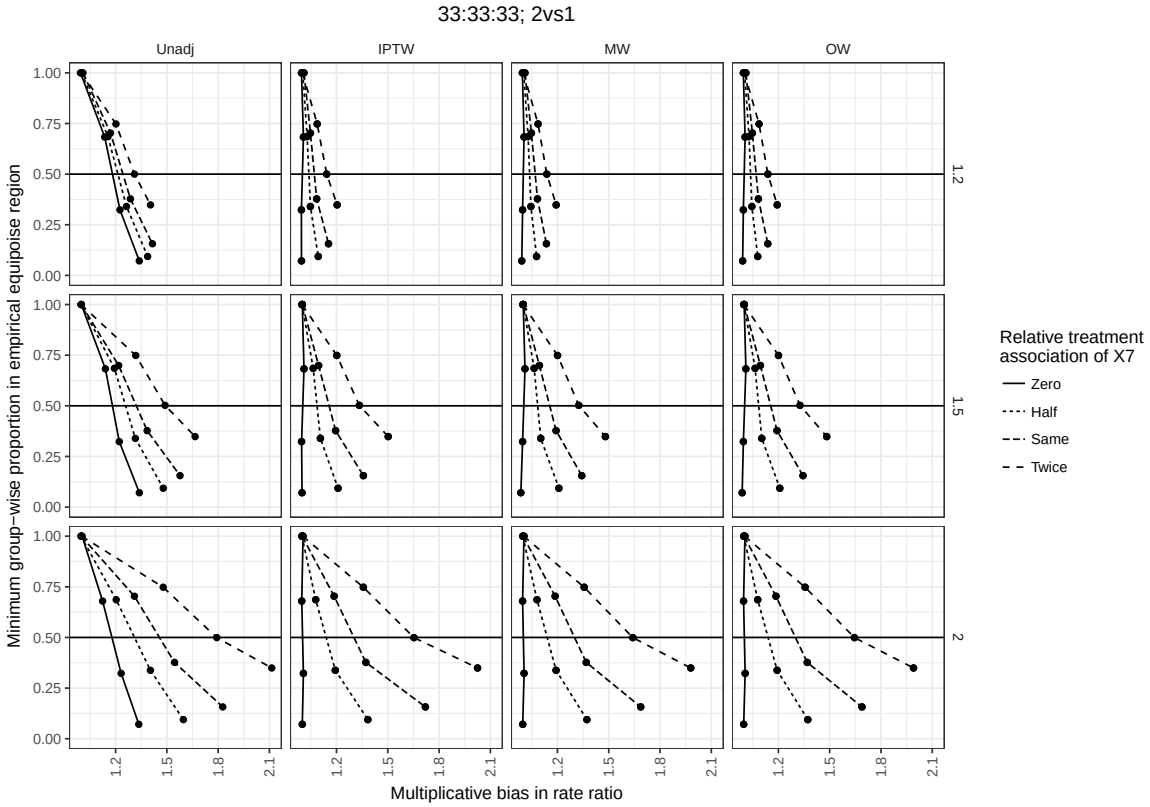
5.3 Minimum group-wise proportion as the summary measure of empirical equipoise

In the following additional results, the index was the minimum of the group-wise proportions of the treatment groups that fell into the proposed empirical equipoise region. The results were essentially the same as the

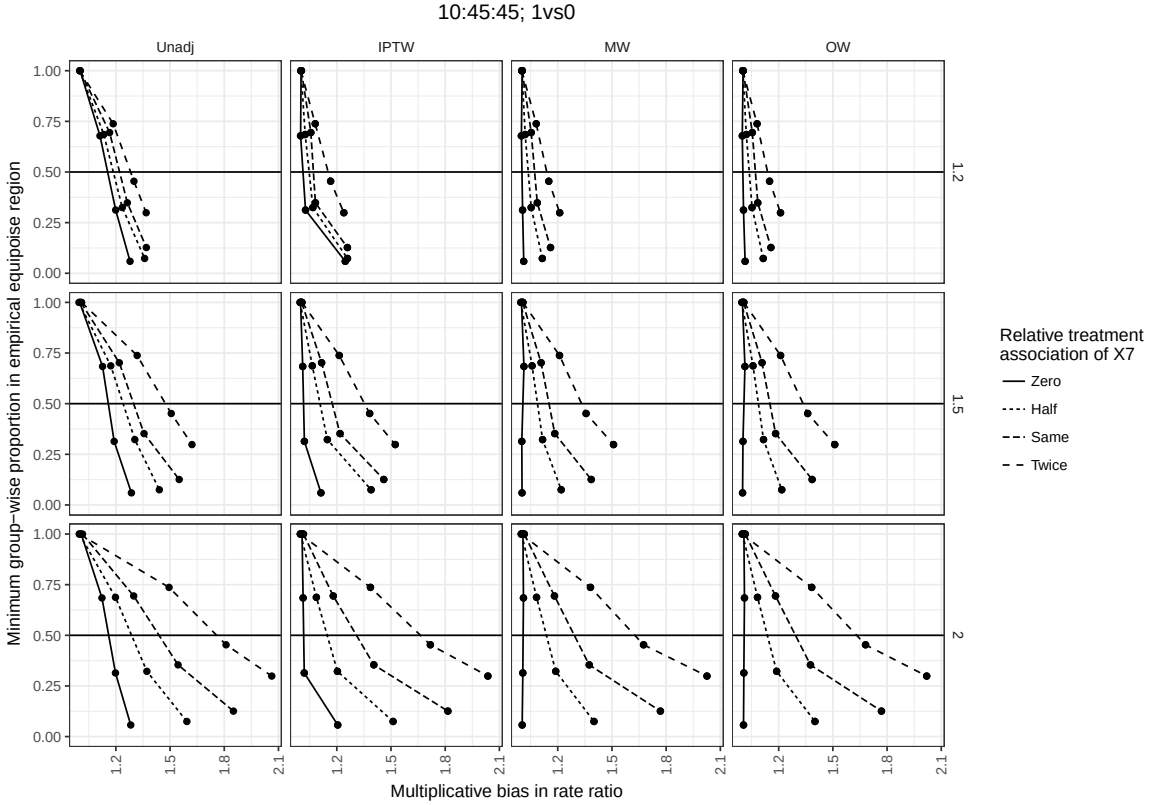
5.3.1 Equal group sizes 33:33:33

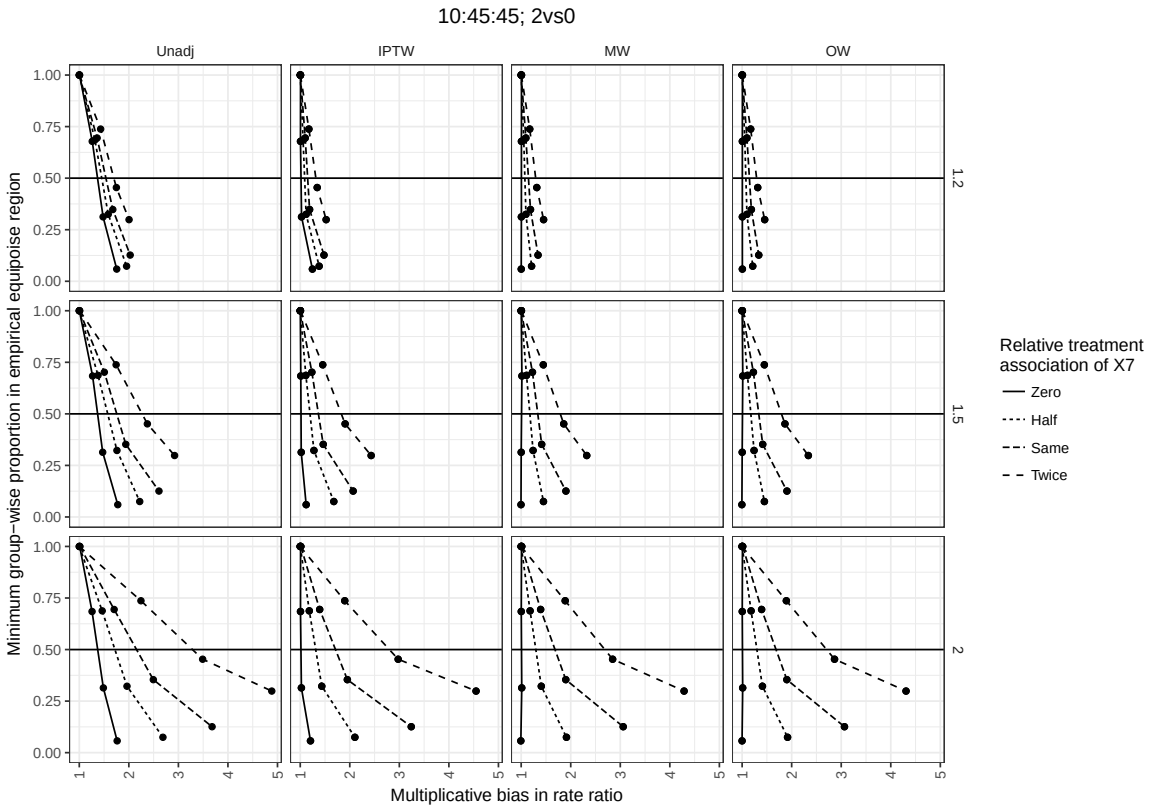


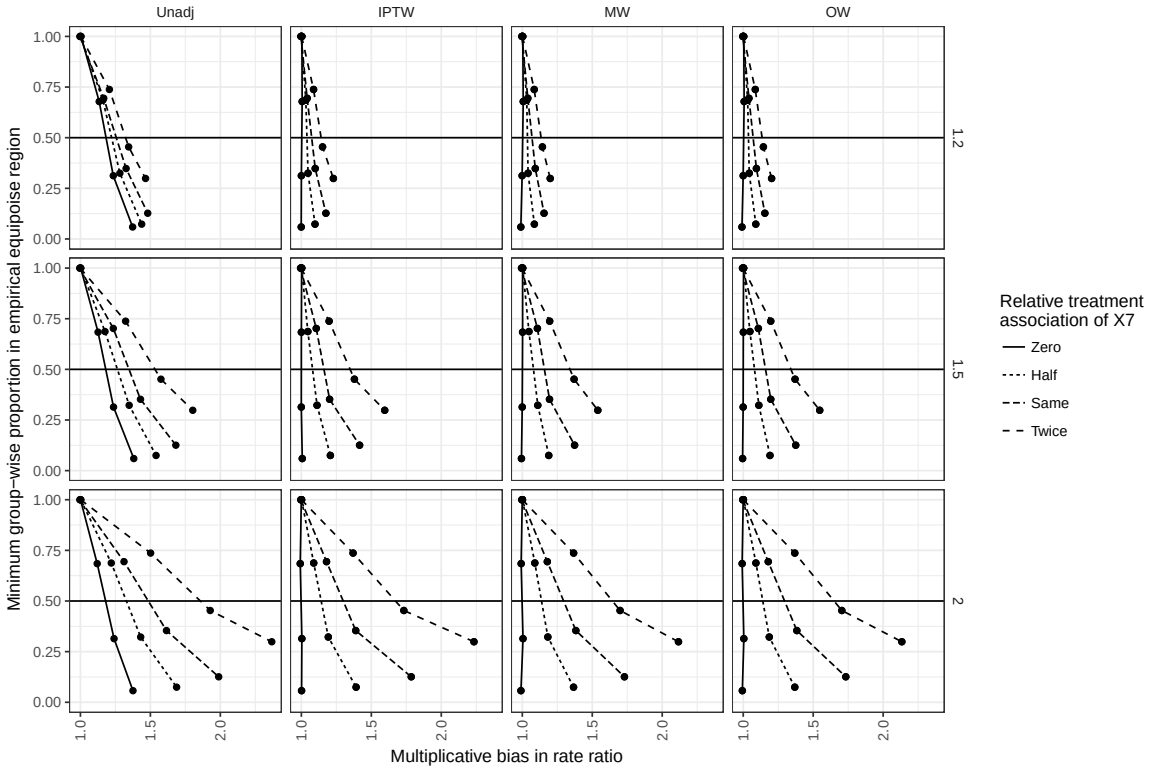




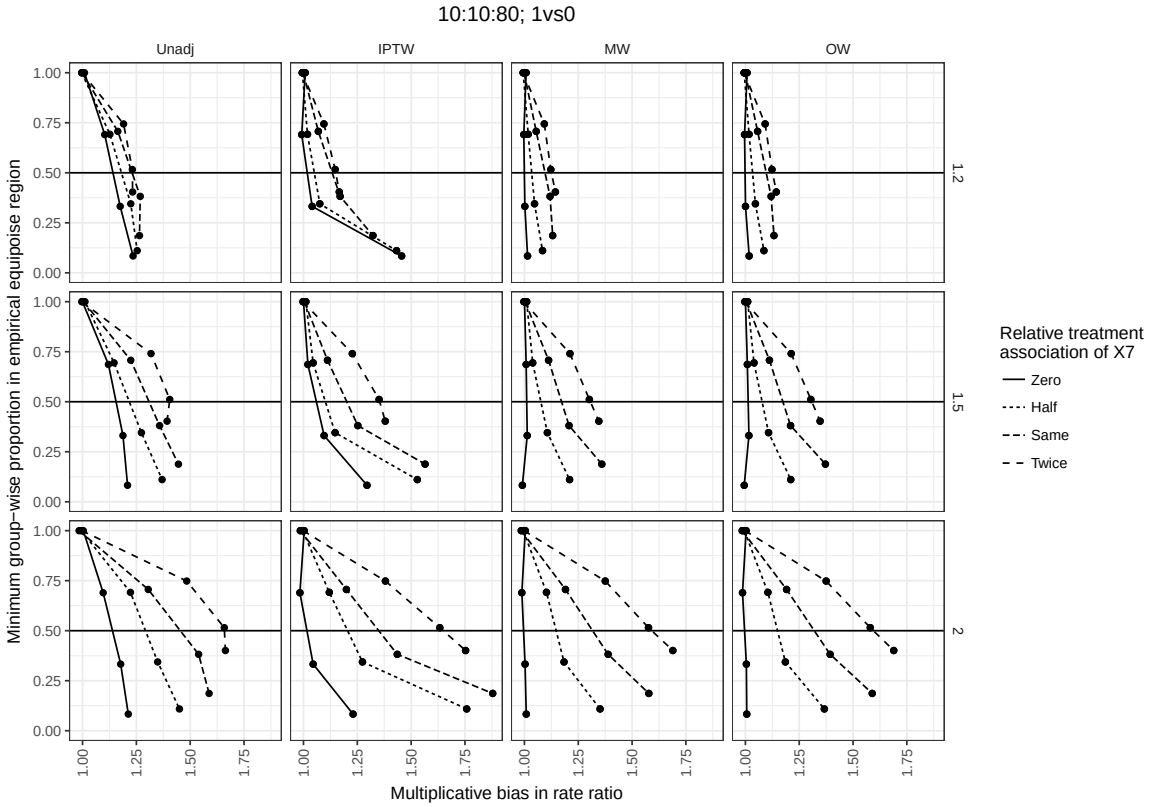
5.3.2 Unequal group sizes 10:45:45

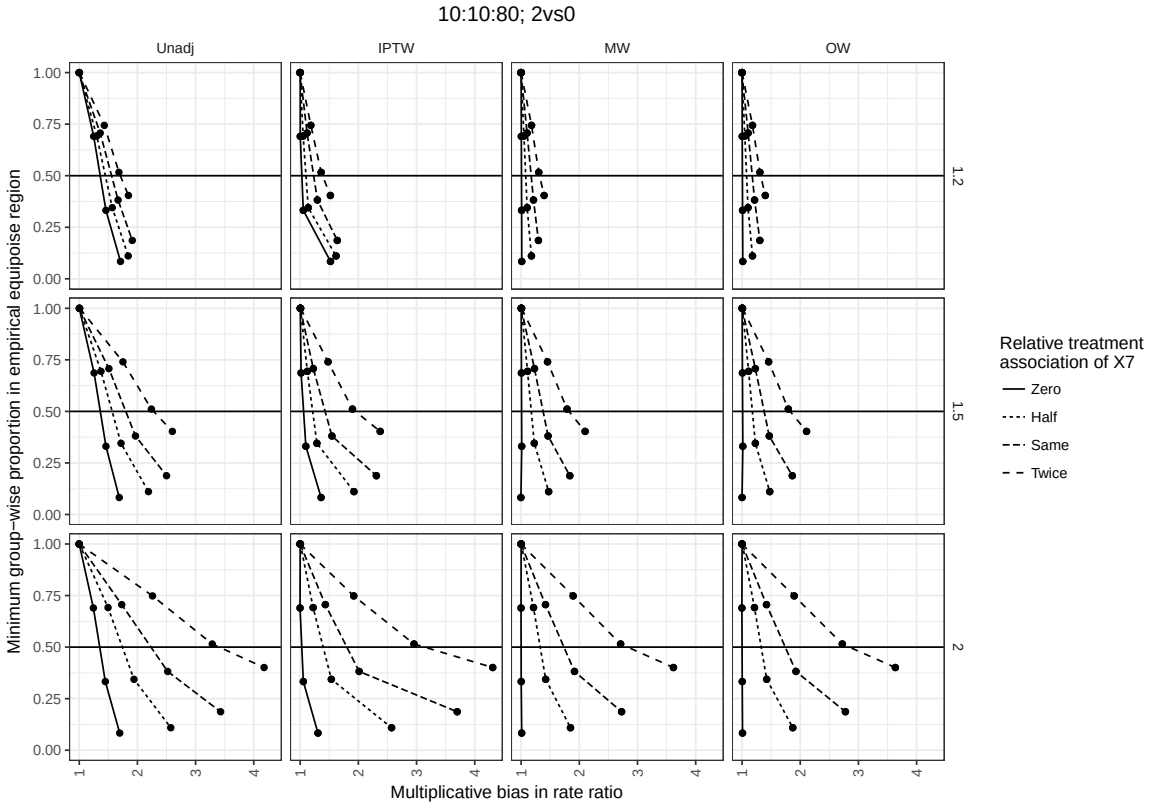


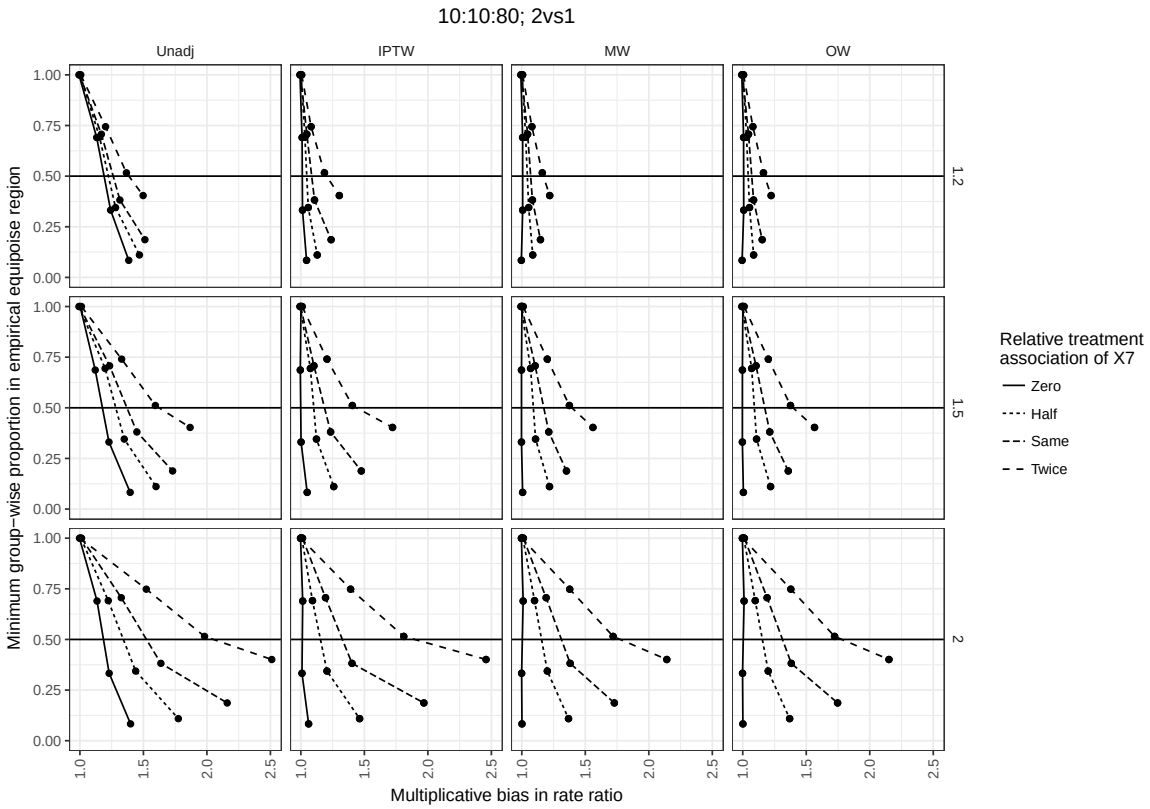

10:45:45; 2vs0
Unadj
IPTW
MW
OW
1.2
1.5
2
Minimum group−wise proportion in empirical equipoise region
Multiplicative bias in rate ratio
Relative treatment association of X7
Zero
Half
Same
Twice


10:45:45; 2vs1
Unadj
IPTW
MW
OW
1.2
1.5
2
Minimum group−wise proportion in empirical equipoise region
Multiplicative bias in rate ratio
Relative treatment association of X7
Zero
Half
Same
Twice

5.3.3 Unequal group sizes 10:10:80



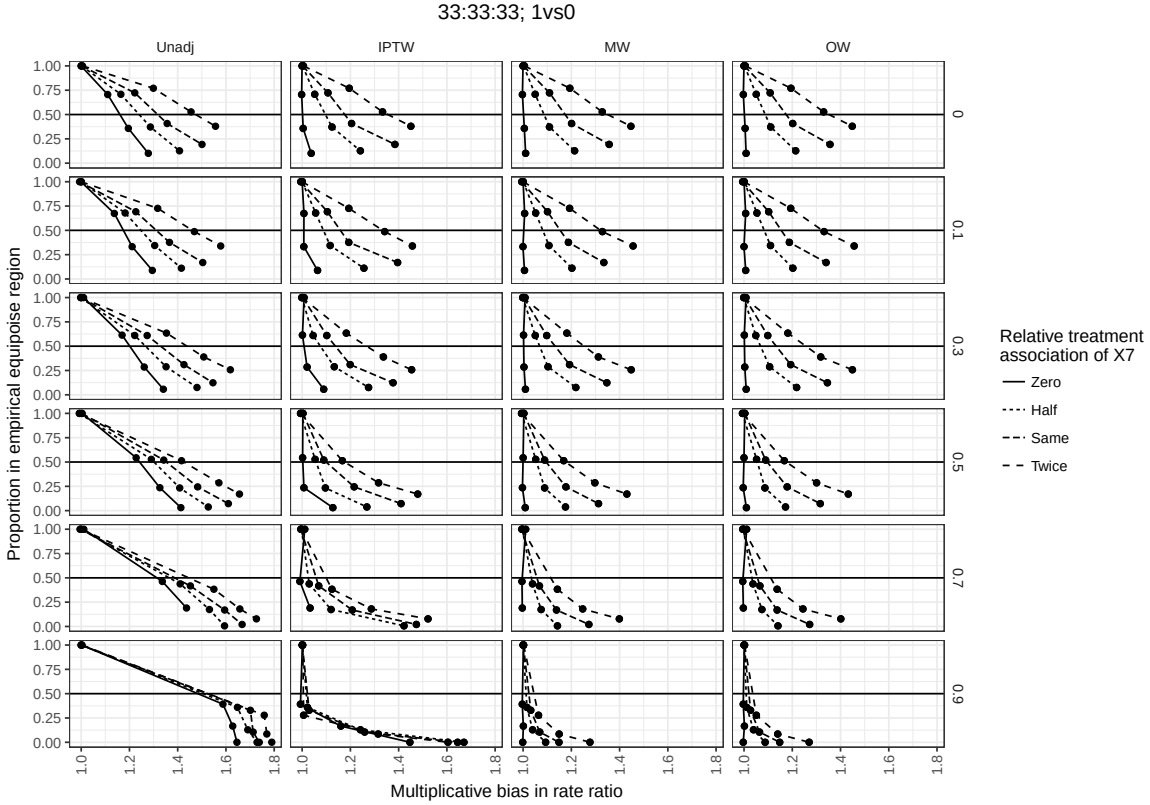


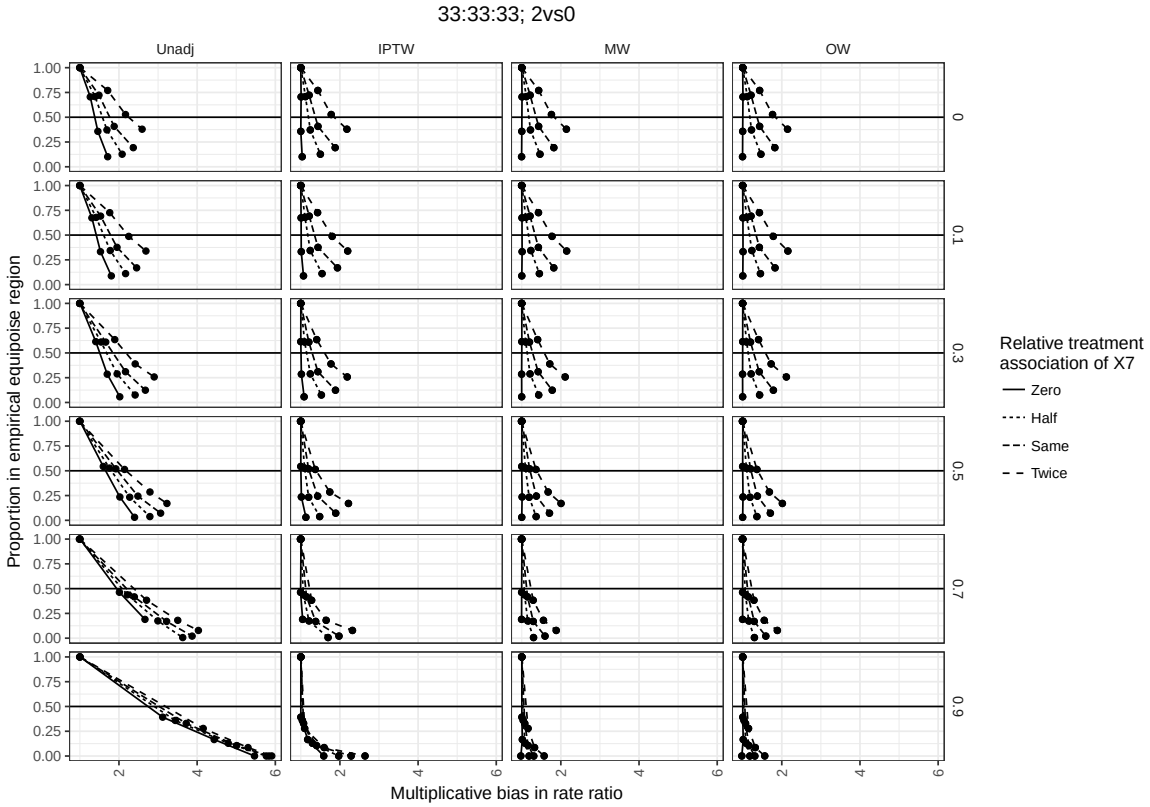


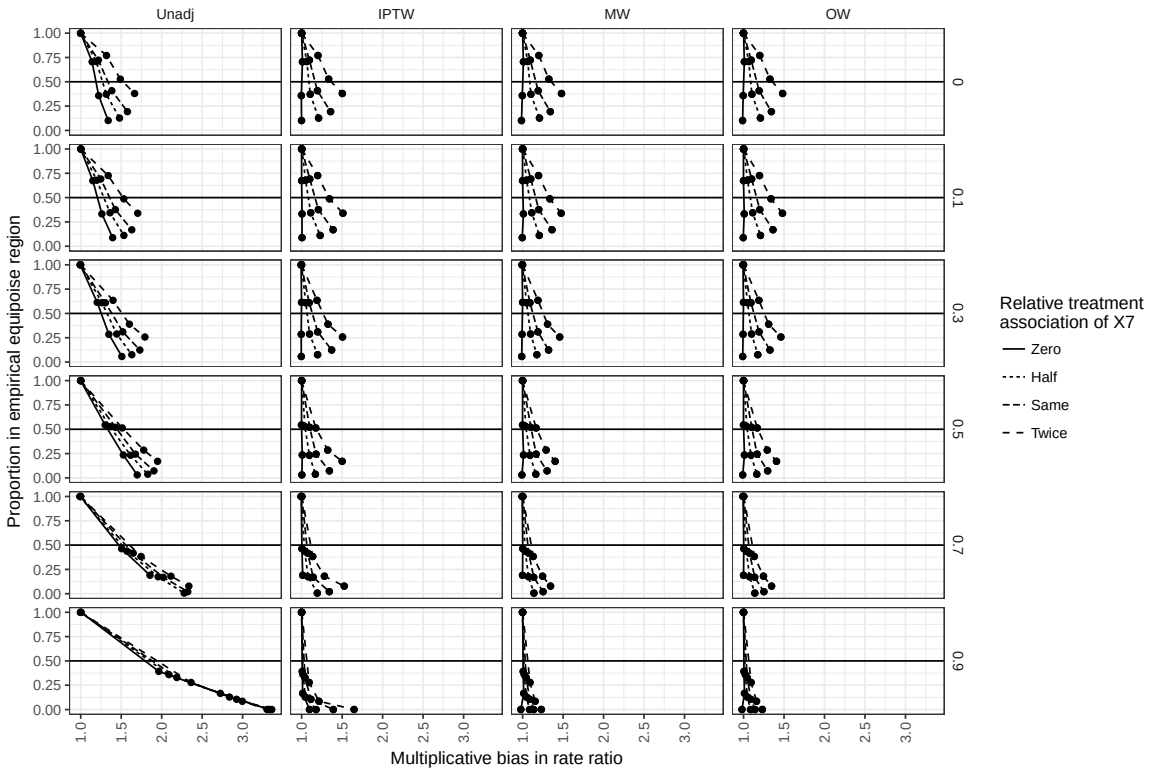
5.4 Different correlation structures

Here the correlation among covariates was varied from 0 to 0.9 (rows of the panels). The RR for the unmeasured variable was kept at 1.5.

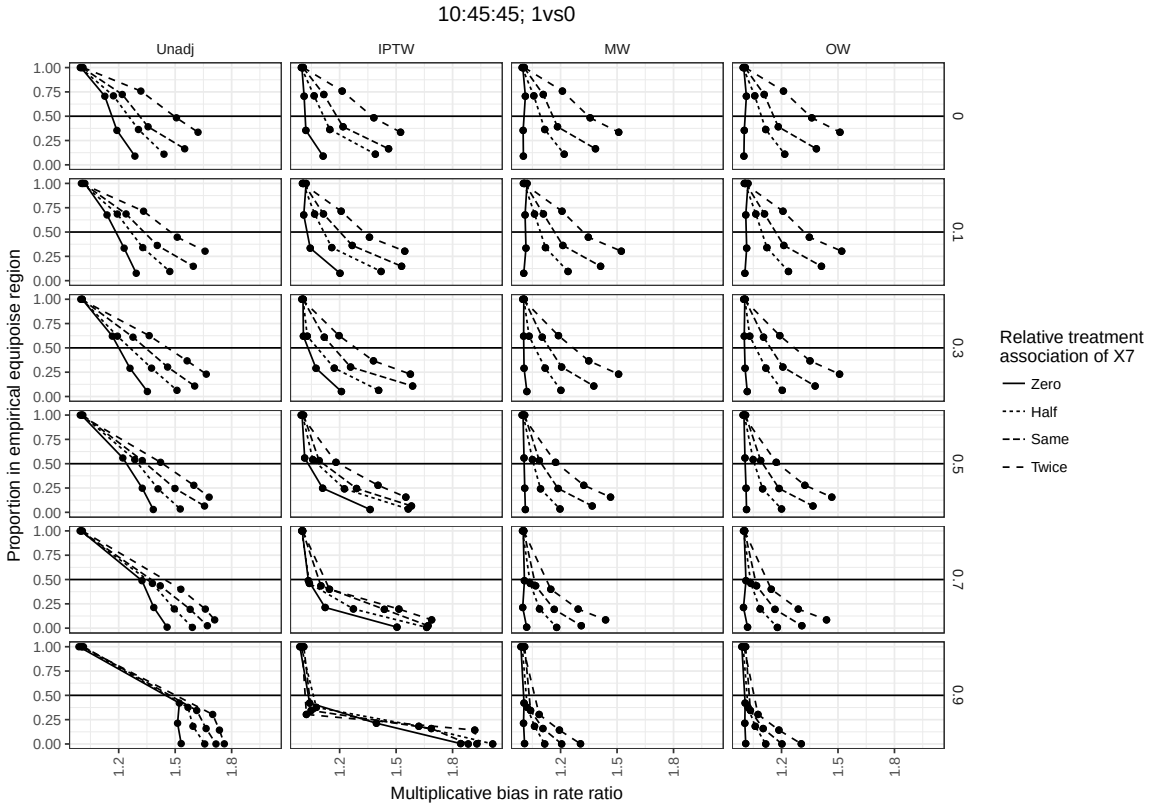
5.4.1 Equal group sizes 33:33:33

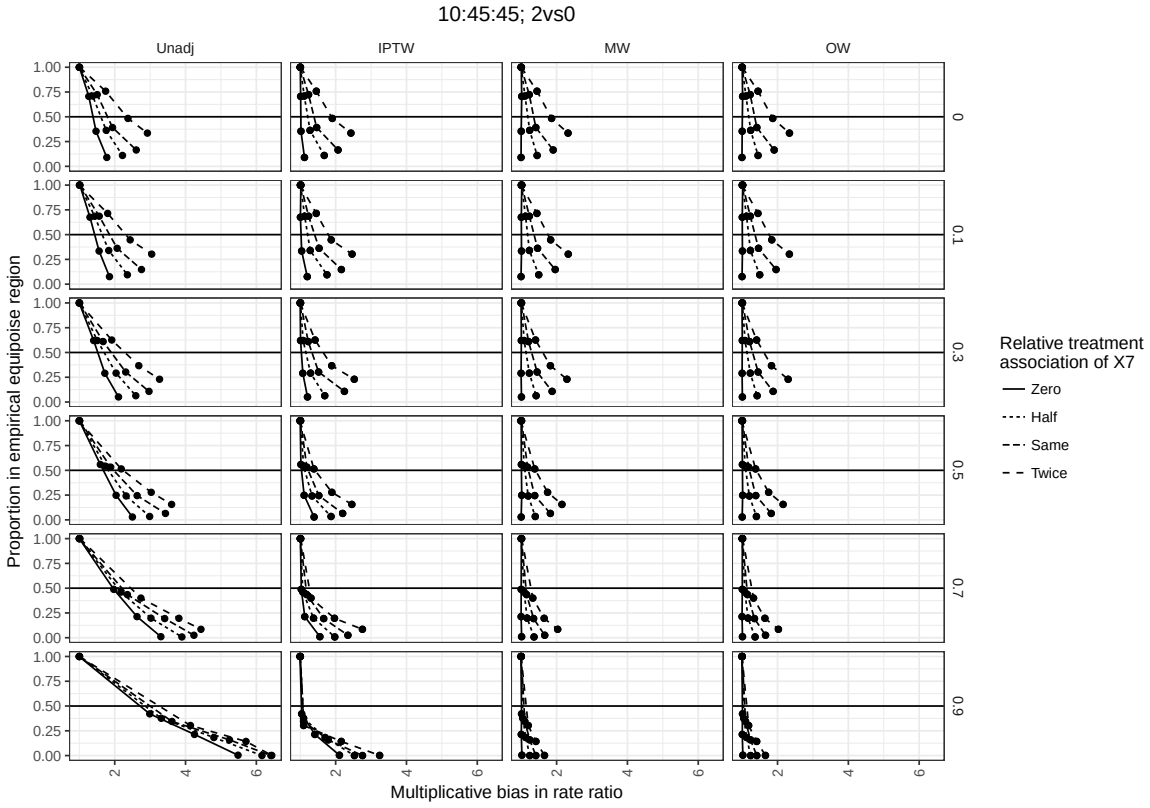


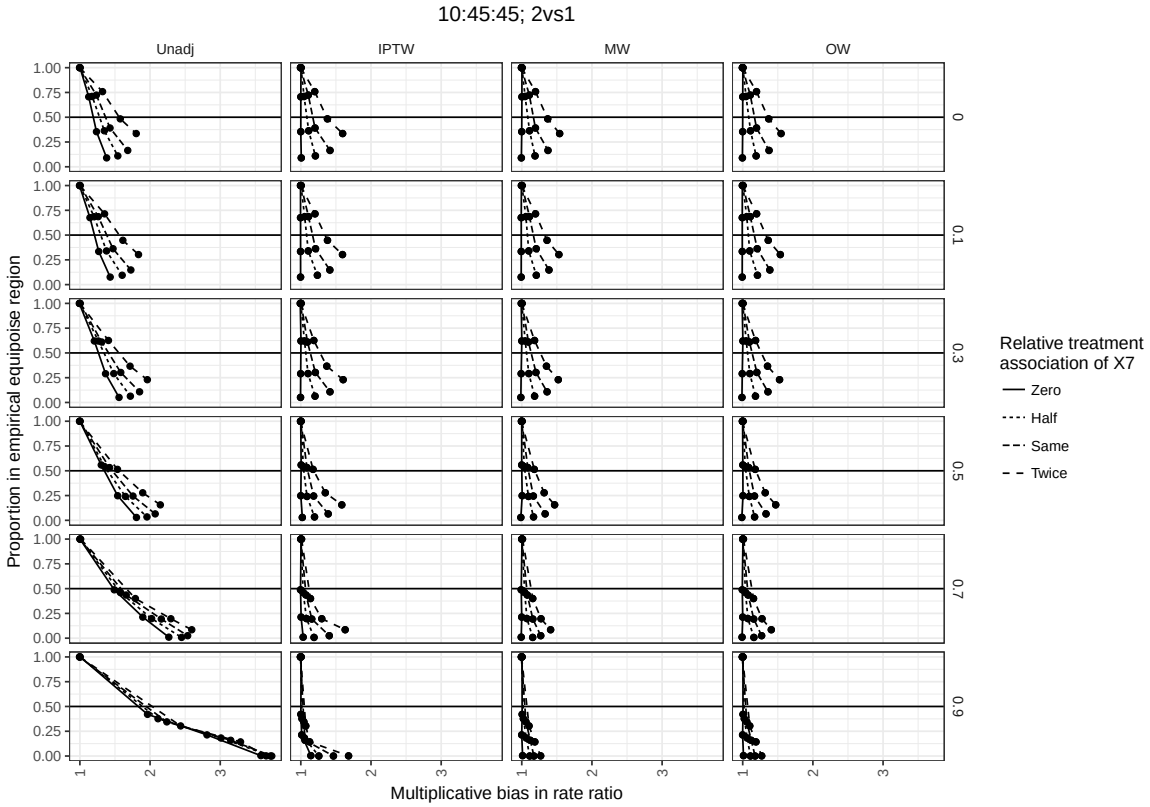

33:33:33; 2vs0
Unadj
IPTW
MW
OW
Proportion in empirical equipoise region
Multiplicative bias in rate ratio
Relative treatment association of X7
Zero
Half
Same
Twice


33:33:33; 2vs1
Unadj
IPTW
MW
OW
Proportion in empirical equipoise region
Multiplicative bias in rate ratio
Relative treatment association of X7
Zero
Half
Same
Twice

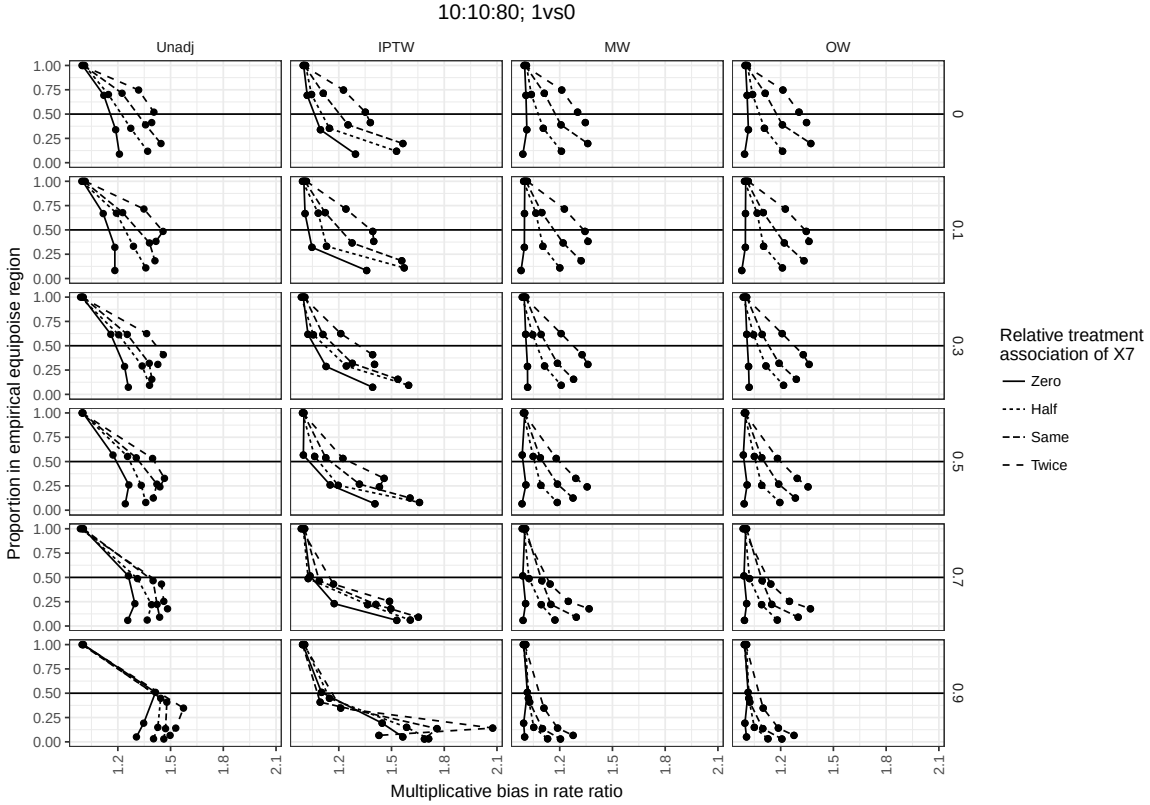
5.4.2 Unequal group sizes 10:45:45

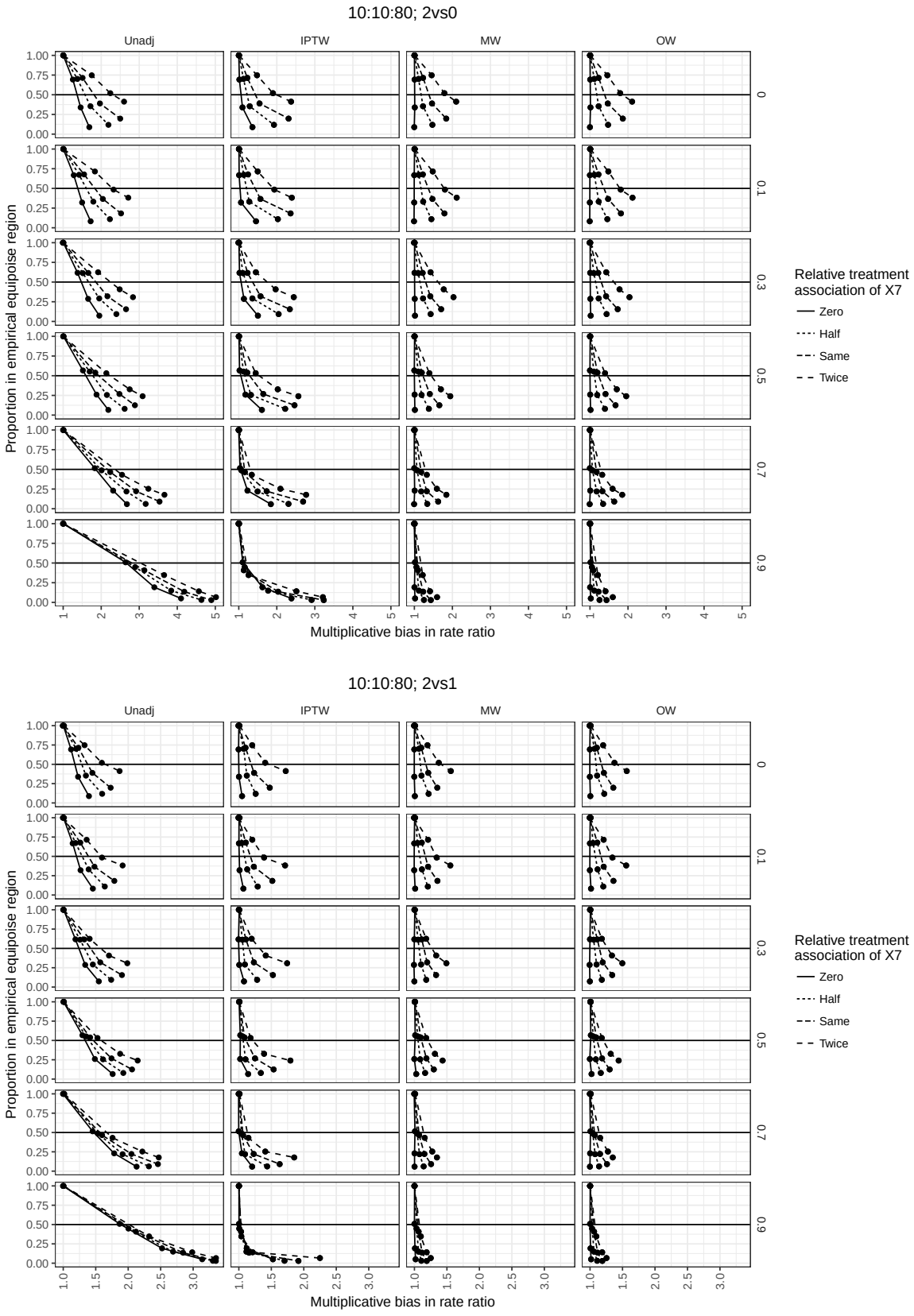






5.4.3 Unequal group sizes 10:10:80





6 Bibliography

[1] A. M. Walker, A. R. Patrick, M. S. Lauer, M. C. Hornbrook, M. G. Marin, R. Platt, V. L. Roger, P. Stang, and S. Schneeweiss, "A tool for assessing the feasibility of comparative effectiveness research," *Comp Eff Res*, vol. 2013, pp. 11–20, Jan. 2013.

[2] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*. Wiley series in probability and statistics, Hoboken, New Jersey: Wiley, third edition ed., 2013.

[3] R. L. Prentice and R. Pyke, "Logistic Disease Incidence Models and Case-Control Studies," *Biometrika*, vol. 66, no. 3, pp. 403–411, 1979.

[4] A. J. Scott and C. J. Wild, "Fitting Logistic Models Under Case-Control or Choice Based Sampling," *J Royal Stat Soc*, vol. 48, no. 2, pp. 170–182, 1986.

[5] N. Hamilton, "Ggtern: An Extension to 'ggplot2', for the Creation of Ternary Diagrams." https://CRAN.R-project.org/package=ggtern, Nov. 2018.

[6] D. H. Solomon, J. A. Rassen, R. J. Glynn, J. Lee, R. Levin, and S. Schneeweiss, "The comparative safety of analgesics in older adults with arthritis," *Arch. Intern. Med.*, vol. 170, pp. 1968–1976, Dec. 2010.

[7] S. C. Kim, D. H. Solomon, J. R. Rogers, S. Gale, M. Klearman, K. Sarsour, and S. Schneeweiss, "Cardiovascular Safety of Tocilizumab Versus Tumor Necrosis Factor Inhibitors in Patients With Rheumatoid Arthritis: A Multi-Database Cohort Study," *Arthritis & Rheumatology (Hoboken, N.J.)*, vol. 69, pp. 1154–1164, June 2017.

[8] E. H. Kang, Y. Jin, G. Brill, J. Lewey, E. Patorno, R. J. Desai, and S. C. Kim, "Comparative Cardiovascular Risk of Abatacept and Tumor Necrosis Factor Inhibitors in Patients With Rheumatoid Arthritis With and Without Diabetes Mellitus: A Multidatabase Cohort Study," *J Am Heart Assoc*, vol. 7, Jan. 2018.

[9] J. A. Singh, K. G. Saag, S. L. Bridges, E. A. Akl, R. R. Bannuru, M. C. Sullivan, E. Vaysbrot, C. McNaughton, M. Osani, R. H. Shmerling, J. R. Curtis, D. E. Furst, D. Parks, A. Kavanaugh, J. O'Dell, C. King, A. Leong, E. L. Matteson, J. T. Schousboe, B. Drevlow, S. Ginsberg, J. Grober, E. W. St Clair, E. Tindall, A. S. Miller, T. McAlindon, and American College of Rheumatology, "2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis," *Arthritis Care Res (Hoboken)*, vol. 68, pp. 1–25, Jan. 2016.

[10] J. S. Smolen, R. Landewé, J. Bijlsma, G. Burmester, K. Chatzidionysiou, M. Dougados, J. Nam, S. Ramiro, M. Voshaar, R. van Vollenhoven, D. Aletaha, M. Aringer, M. Boers, C. D. Buckley, F. Buttgereit, V. Bykerk, M. Cardiel, B. Combe, M. Cutolo, Y. van Eijk-Hustings, P. Emery, A. Finckh, C. Gabay, J. Gomez-Reino, L. Gossec, J.-E. Gottenberg, J. M. W. Hazes, T. Huizinga, M. Jani, D. Karateev, M. Kouloumas, T. Kvien, Z. Li, X. Mariette, I. McInnes, E. Mysler, P. Nash, K. Pavelka, G. Poór, C. Richez, P. van Riel, A. Rubbert-Roth, K. Saag, J. da Silva, T. Stamm, T. Takeuchi, R. Westhovens, M. de Wit, and D. van der Heijde, "EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update," *Ann. Rheum. Dis.*, vol. 76, pp. 960–977, June 2017.

[11] S. Greenland, J. M. Robins, and J. Pearl, "Confounding and Collapsibility in Causal Inference," *Statistical Science*, vol. 14, no. 1, pp. 29–46, 1999.

[12] J. M. Robins, M. A. Hernán, and B. Brumback, "Marginal structural models and causal inference in epidemiology," *Epidemiology*, vol. 11, pp. 550–560, Sept. 2000.

[13] L. Li and T. Greene, "A weighting analogue to pair matching in propensity score analysis," *Int J Biostat*, vol. 9, no. 2, pp. 215–234, 2013.

[14] K. Yoshida, S. Hernandez-Diaz, D. H. Solomon, J. W. Jackson, J. J. Gagne, R. J. Glynn, and J. M. Franklin, "Matching Weights to Simultaneously Compare Three Treatment Groups: Comparison to Three-way Matching," *Epidemiology*, vol. 28, pp. 387–395, May 2017.

[15] F. Li, K. L. Morgan, and A. M. Zaslavsky, "Balancing Covariates via Propensity Score Weighting," *Journal of the American Statistical Association*, vol. 113, pp. 390–400, Jan. 2018.

[16] F. Li, L. E. Thomas, and F. Li, "Addressing Extreme Propensity Scores via the Overlap Weights [available online ahead of print September 5, 2018]," *Am. J. Epidemiol.*, vol. doi: 10.1093/aje/kwy201, Sept. 2018.

[17] F. Li and F. Li, "Propensity Score Weighting for Causal Inference with Multi-valued Treatments," *arXiv:1808.05339 [stat]*, Aug. 2018.