Testing different models for *strong* confouding effects (fitting 2000 times)

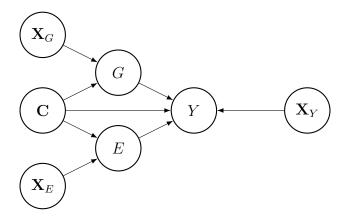


Figure 1: Directed Acyclic Graphs for causal interactions under two treatments G and E.

Let G and E denote the two binary treatments and Y be the binary outcome. Let \mathbf{X}_G , \mathbf{X}_E be the set of instrumental variables for G and E, respectively. Let \mathbf{X}_Y be the outcome-only covariates. Let \mathbf{C} be the set of common confounders to both treatments.

Data generation

In this study, we will assume the outcome Y is binary and there are two treatments G and E. Let $\mathbf{X}_G = \{X_1, X_2\}$, $\mathbf{X}_E = \{X_3, X_4\}$, $\mathbf{X}_Y = \{X_5, X_6\}$ and $\mathbf{C} = \{X_7, X_8\}$. Suppose $\mathbf{X} = \{X_1, \dots, X_8\}$ is generated to follow a multivariate normal distribution such that $\mathbf{X} \sim \mathcal{N}_8(\mathbf{0}, \mathbf{I}_8)$. To generate the treatment assignment probabilities, we firstly generate probabilities for each treatment individually:

$$\pi_{G=1} = P(G = 1 \mid \mathbf{Z}_G) = \frac{\exp\{f(\mathbf{Z}_G)\}}{1 + \exp\{f(\mathbf{Z}_G)\}}$$
$$\pi_{E=1} = P(E_i = 1 \mid \mathbf{Z}_E) = \frac{\exp\{f(\mathbf{Z}_E)\}}{1 + \exp\{f(\mathbf{Z}_E)\}}$$

where $f(\mathbf{Z}_G)$ and $f(\mathbf{Z}_E)$ are the function of \mathbf{Z}_G and \mathbf{Z}_E such that $\mathbf{Z}_G = \{X_1, X_2, X_7, X_8\}$ and $\mathbf{Z}_E = \{X_3, X_4, X_7, X_8\}$. We now suppose

$$f(\mathbf{Z}_G) = 0.8X_1 + 0.5X_2 + 0.3X_7 + 0.5X_8$$

$$f(\mathbf{Z}_E) = -0.5X_3 + 0.2X_4 + 0.3X_7 + 0.5X_8$$

Note that here it is unnecessary to add the intercept as the expectation for each covariate is 0, making the marginal probabilities $\pi_{G=1}$ and $\pi_{E=1}$ are approximately 0.5. With the above setups, the probability of assignment for each combination of treatments can be computed as

$$\theta_{00} = p_{G=0,E=0} = (1 - \pi_{G=1})(1 - \pi_{E=1})$$

$$\theta_{10} = p_{G=1,E=0} = \pi_{G=1}(1 - \pi_{E=1})$$

$$\theta_{01} = p_{G=0,E=1} = (1 - \pi_{G=1})\pi_{E=1}$$

$$\theta_{11} = p_{G=1,E=1} = \pi_{G=1}\pi_{E=1}$$

For each observation, the probability of receiving each pair of treatment follows a multinomial distribution such that $T_i \mid \mathbf{X} \sim \text{Multinomial}(1, \theta_{00}, \theta_{10}, \theta_{01}, \theta_{11})$, where $T_i \in \{(0,0), (1,0), (0,1), (1,1)\}$. In addition, denote $\mathbf{Z}_Y = \{\mathbf{X}_Y, \mathbf{C}\} = \{X_5, X_6, X_7, X_8\}$, the outcome is generated as

$$Pr(Y = 1 \mid \mathbf{Z}_Y, G, E) = logit^{-1}(\beta_0 + 0.8X_5 + 0.5X_6 - 0.6X_7 + 0.4X_8 + 0.3G + 0.4E + 0.8GE)$$

The true value of conditional RERI in terms of odds ratio can be computed as $\exp\{0.3+0.4+0.8\} - \exp\{0.3\} - \exp\{0.4\} + 1 = 2.64$.

Using control only

 $\beta_0 = -6$: prevalence ≈ 0.01

Model	eta_3	\widehat{SE}_{β_3}	coverage	Abs. Bias	Bias(%)
Default LOM	4.401(13.803)	6.084	0.958	1.834	71.45
CLOM(X5X6)	4.546(12.550)	6.078	0.960	1.906	72.20
CLOM(X7X8)	4.598(13.890)	6.128	0.948	1.958	74.17
CLOM(X5678)	4.436(10.610)	5.915	0.956	1.796	68.03
MSLOM(X7X8)	4.096(9.955)	4.583	0.944	1.529	59.56
MSLOM(X1278)	4.260(11.728)	5.605	0.936	1.693	65.95
MSLOM(X3478)	4.287(11.649)	4.787	0.944	1.720	67.00
MSLOM(X5678)	4.197(9.790)	5.023	0.942	1.630	63.50
MSLOM(X123478)	4.484(13.921)	5.804	0.940	1.917	74.68
MSLOM(all)	4.644(13.849)	7.204	0.939	2.077	80.91
DR	4.689(13.111)	3.822	0.769	2.049	77.61

 $RERI_{OR}^{true} = 2.64$; Empirical: $RERI_{OR} = 2.567$, $RERI_{RR} = 2.501$

 $\beta_0 = -5.5$: **prevalence** ≈ 0.018

Model	β_3	\widehat{SE}_{β_3}	coverage	Abs. Bias	Bias(%)
Default LOM	3.012(4.289)	2.085	0.949	0.519	20.82
CLOM(X5X6)	3.291(7.281)	2.258	0.946	0.651	24.66
CLOM(X7X8)	3.012(3.961)	2.163	0.935	0.372	14.09
CLOM(X5678)	3.301(5.071)	2.411	0.953	0.661	25.04
MSLOM(X7X8)	3.036(4.530)	2.168	0.946	0.543	21.78
MSLOM(X1278)	3.144(5.241)	2.533	0.945	0.651	26.11
MSLOM(X3478)	3.122(4.965)	2.337	0.944	0.629	25.23
MSLOM(X5678)	3.110(4.384)	2.192	0.950	0.617	24.75
MSLOM(X123478)	3.219(5.037)	2.669	0.944	0.726	29.12
MSLOM(all)	3.300(4.955)	2.717	0.946	0.807	32.37
DR	3.407(5.982)	1.295	0.709	0.767	29.05

 $RERI_{OR}^{true} = 2.64$; Empirical: $RERI_{OR} = 2.493$, $RERI_{RR} = 2.391$

 $\beta_0 = -5$: prevalence ≈ 0.03

Model	β_3	\widehat{SE}_{β_3}	coverage	Abs. Bias	Bias(%)
Default LOM	2.548(1.371)	1.195	0.948	0.148	6.17
CLOM(X5X6)	2.763(1.503)	1.289	0.939	0.123	4.66
CLOM(X7X8)	2.586(1.440)	1.268	0.918	0.054	2.05
CLOM(X5678)	2.891(1.643)	1.414	0.950	0.251	9.51
MSLOM(X7X8)	2.568(1.698)	1.337	0.947	0.168	7.00
MSLOM(X1278)	2.628(2.097)	1.573	0.945	0.228	9.50
MSLOM(X3478)	2.598(1.790)	1.414	0.950	0.198	8.25
MSLOM(X5678)	2.669(1.740)	1.383	0.953	0.269	11.21
MSLOM(X123478)	2.680(2.330)	1.676	0.946	0.280	11.67
MSLOM(all)	2.781(2.272)	1.726	0.952	0.381	15.88
DR	2.946(2.095)	0.717	0.677	0.306	11.59

 $RERI_{OR}^{true} = 2.64$; Empirical: $RERI_{OR} = 2.400$, $RERI_{RR} = 2.246$

 $\beta_0 = -4.5$: prevalence ≈ 0.05

Model	β_3	\widehat{SE}_{β_3}	coverage	Abs. Bias	Bias(%)
Default LOM	2.362(0.880)	0.859	0.962	0.075	3.28
CLOM(X5X6)	2.628(0.970)	0.948	0.939	0.012	0.45
CLOM(X7X8)	2.422(0.946)	0.923	0.902	0.218	8.26
CLOM(X5678)	2.791(1.088)	1.059	0.956	0.151	5.72
MSLOM(X7X8)	2.378(0.993)	0.961	0.951	0.091	3.98
MSLOM(X1278)	2.420(1.262)	1.132	0.947	0.133	5.82
MSLOM(X3478)	2.398(1.052)	1.016	0.954	0.111	4.85
MSLOM(X5678)	2.510(1.058)	1.009	0.962	0.223	9.75
MSLOM(X123478)	2.445(1.333)	1.198	0.947	0.158	6.91
MSLOM(all)	2.582(1.417)	1.259	0.954	0.295	12.90
DR	2.820(1.184)	0.520	0.647	0.180	6.82

 $RERI_{OR}^{true} = 2.64$; Empirical: $RERI_{OR} = 2.287$, $RERI_{RR} = 2.063$

 $\beta_0 = -4$: prevalence ≈ 0.08

Model	β_3	\widehat{SE}_{β_3}	coverage	Abs. Bias	Bias(%)
Default LOM	2.182(0.661)	0.650	0.957	0.019	0.88
CLOM(X5X6)	2.508(0.744)	0.738	0.931	0.132	5.00
CLOM(X7X8)	2.277(0.724)	0.709	0.867	0.363	13.75
CLOM(X5678)	2.719(0.854)	0.841	0.949	0.079	2.99
MSLOM(X7X8)	2.213(0.748)	0.730	0.954	0.050	2.31
MSLOM(X1278)	2.235(0.897)	0.861	0.947	0.072	3.33
MSLOM(X3478)	2.231(0.790)	0.772	0.958	0.068	3.14
MSLOM(X5678)	2.378(0.810)	0.779	0.968	0.215	9.94
MSLOM(X123478)	2.251(0.945)	0.909	0.952	0.088	4.07
MSLOM(all)	2.420(1.016)	0.971	0.964	0.257	11.88
DR	2.732(0.910)	0.406	0.637	0.092	3.48

 $RERI_{OR}^{true} = 2.64$; Empirical: $RERI_{OR} = 2.163$, $RERI_{RR} = 1.849$

 $\beta_0 = -3$: prevalence ≈ 0.18

Model	β_3	\widehat{SE}_{β_3}	coverage	Abs. Bias	Bias(%)
Default LOM	1.915(0.415)	0.420	0.959	0.000	0.00
CLOM(X5X6)	2.367(0.500)	0.507	0.889	0.273	10.34
CLOM(X7X8)	2.082(0.470)	0.473	0.705	0.558	21.14
CLOM(X5678)	2.683(0.602)	0.604	0.956	0.043	1.63
MSLOM(X7X8)	1.986(0.490)	0.482	0.953	0.071	3.71
MSLOM(X1278)	2.005(0.597)	0.572	0.955	0.090	4.70
MSLOM(X3478)	1.995(0.522)	0.510	0.958	0.080	4.18
MSLOM(X5678)	2.245(0.566)	0.538	0.945	0.330	17.23
MSLOM(X123478)	2.011(0.626)	0.602	0.960	0.096	5.01
MSLOM(all)	2.275(0.716)	0.674	0.960	0.360	18.80
DR	2.690(0.653)	0.285	0.636	0.050	1.89

 $RERI_{OR}^{true} = 2.64$; Empirical: $RERI_{OR} = 1.915$, $RERI_{RR} = 1.362$

 $\beta_0 = -2$: prevalence ≈ 0.4

Model	eta_3	\widehat{SE}_{β_3}	coverage	Abs. Bias	Bias(%)
Default LOM	1.730(0.325)	0.317	0.950	0.003	0.17
CLOM(X5X6)	2.277(0.418)	0.405	0.810	0.363	13.75
CLOM(X7X8)	1.949(0.377)	0.367	0.503	0.691	26.17
CLOM(X5678)	2.678(0.515)	0.497	0.948	0.038	1.44
MSLOM(X7X8)	1.838(0.397)	0.377	0.944	0.111	6.43
MSLOM(X1278)	1.848(0.476)	0.448	0.953	0.121	7.01
MSLOM(X3478)	1.839(0.418)	0.399	0.947	0.112	6.49
MSLOM(X5678)	2.206(0.498)	0.442	0.835	0.479	27.74
MSLOM(X123478)	1.850(0.498)	0.473	0.954	0.123	7.12
MSLOM(all)	2.224(0.621)	0.559	0.898	0.497	28.78
DR	2.684(0.563)	0.228	0.594	0.044	1.67

 $RERI_{OR}^{true} = 2.64$; Empirical: $RERI_{OR} = 1.727$, $RERI_{RR} = 0.888$

 $\beta_0 = -1$: prevalence ≈ 0.6

Model	β_3	\widehat{SE}_{β_3}	coverage	Abs. Bias	Bias(%)
Default LOM	1.655(0.286)	0.289	0.954	0.005	0.30
CLOM(X5X6)	2.233(0.373)	0.377	0.766	0.407	15.42
CLOM(X7X8)	1.892(0.331)	0.337	0.397	0.748	28.33
CLOM(X5678)	2.673(0.461)	0.467	0.953	0.033	1.25
MSLOM(X7X8)	1.802(0.366)	0.356	0.948	0.152	9.21
MSLOM(X1278)	1.803(0.451)	0.427	0.952	0.153	9.27
MSLOM(X3478)	1.805(0.392)	0.379	0.942	0.155	9.39
MSLOM(X5678)	2.276(0.484)	0.436	0.745	0.626	37.94
MSLOM(X123478)	1.806(0.477)	0.453	0.946	0.156	9.45
MSLOM(all)	2.288(0.633)	0.559	0.835	0.638	38.67
DR	2.687(0.521)	0.205	0.562	0.047	1.78

 $RERI_{OR}^{true} = 2.64$; Empirical: $RERI_{OR} = 1.650$, $RERI_{RR} = 0.511$

Using both control and case

 $\beta_0 = -6$: prevalence ≈ 0.01

Model	β_3	\widehat{SE}_{β_3}	coverage	Abs. Bias	Bias(%)
Default LOM	4.401(13.803)	6.084	0.958	1.834	71.45
CLOM(X5X6)	4.546(12.550)	6.078	0.960	1.906	72.20
CLOM(X7X8)	4.598(13.890)	6.128	0.948	1.958	74.17
CLOM(X5678)	4.436(10.610)	5.915	0.956	1.796	68.03
MSLOM(X7X8)	4.062(9.855)	5.017	0.943	1.495	58.24
MSLOM(X1278)	4.237(11.653)	6.688	0.936	1.670	65.06
MSLOM(X3478)	4.234(11.277)	5.244	0.943	1.667	64.94
MSLOM(X5678)	4.204(11.796)	4.866	0.942	1.637	63.77
MSLOM(X123478)	4.293(12.278)	9.428	0.940	1.726	67.24
MSLOM(all)	4.421(14.160)	11.028	0.935	1.854	72.22
DR	4.843(13.981)	3.942	0.773	2.203	83.45

 $RERI_{OR}^{true} = 2.64$; Empirical: $RERI_{OR} = 2.567$, $RERI_{RR} = 2.501$

 $\beta_0 = -5.5$: **prevalence** ≈ 0.018

Model	eta_3	\widehat{SE}_{β_3}	coverage	Abs. Bias	Bias(%)
Default LOM	3.012(4.289)	2.085	0.949	0.519	20.82
CLOM(X5X6)	3.291(7.281)	2.258	0.946	0.651	24.66
CLOM(X7X8)	3.012(3.961)	2.163	0.935	0.372	14.09
CLOM(X5678)	3.301(5.071)	2.411	0.953	0.661	25.04
MSLOM(X7X8)	3.002(4.774)	2.160	0.946	0.509	20.42
MSLOM(X1278)	3.110(5.442)	2.481	0.943	0.617	24.75
MSLOM(X3478)	3.052(4.413)	2.278	0.943	0.559	22.42
MSLOM(X5678)	2.966(4.011)	2.135	0.947	0.473	18.97
MSLOM(X123478)	3.197(5.542)	2.645	0.942	0.704	28.24
MSLOM(all)	3.158(4.855)	2.621	0.942	0.665	26.67
DR	3.428(5.941)	1.329	0.712	0.788	29.85

 $RERI_{OR}^{true} = 2.64$; Empirical: $RERI_{OR} = 2.493$, $RERI_{RR} = 2.391$

 $\beta_0 = -5$: prevalence ≈ 0.03

Model	β_3	\widehat{SE}_{β_3}	coverage	Abs. Bias	Bias(%)
Default LOM	2.548(1.371)	1.195	0.948	0.148	6.17
CLOM(X5X6)	2.763(1.503)	1.289	0.939	0.123	4.66
CLOM(X7X8)	2.586(1.440)	1.268	0.918	0.054	2.05
CLOM(X5678)	2.891(1.643)	1.414	0.950	0.251	9.51
MSLOM(X7X8)	2.509(1.639)	1.307	0.945	0.109	4.54
MSLOM(X1278)	2.569(2.045)	1.537	0.942	0.169	7.04
MSLOM(X3478)	2.541(1.818)	1.387	0.947	0.141	5.88
MSLOM(X5678)	2.507(1.643)	1.309	0.943	0.107	4.46
MSLOM(X123478)	2.615(2.187)	1.630	0.944	0.215	8.96
MSLOM(all)	2.613(2.276)	1.636	0.944	0.213	8.88
DR	2.952(2.166)	0.723	0.683	0.312	11.82

 $RERI_{OR}^{true} = 2.64$; Empirical: $RERI_{OR} = 2.400$, $RERI_{RR} = 2.246$

 $\beta_0 = -4.5$: prevalence ≈ 0.05

Model	β_3	\widehat{SE}_{β_3}	coverage	Abs. Bias	Bias(%)
Default LOM	2.362(0.880)	0.859	0.962	0.075	3.28
CLOM(X5X6)	2.628(0.970)	0.948	0.939	0.012	0.45
CLOM(X7X8)	2.422(0.946)	0.923	0.902	0.218	8.26
CLOM(X5678)	2.791(1.088)	1.059	0.956	0.151	5.72
MSLOM(X7X8)	2.307(0.952)	0.933	0.945	0.020	0.87
MSLOM(X1278)	2.345(1.206)	1.097	0.945	0.058	2.54
MSLOM(X3478)	2.325(1.007)	0.985	0.950	0.038	1.66
MSLOM(X5678)	2.303(0.939)	0.933	0.947	0.016	0.70
MSLOM(X123478)	2.369(1.273)	1.158	0.941	0.082	3.59
MSLOM(all)	2.364(1.258)	1.157	0.941	0.077	3.37
DR	2.824(1.184)	0.525	0.650	0.184	6.97

 $RERI_{OR}^{true} = 2.64$; Empirical: $RERI_{OR} = 2.287$, $RERI_{RR} = 2.063$

 $\beta_0 = -4$: prevalence ≈ 0.08

Model	eta_3	\widehat{SE}_{β_3}	coverage	Abs. Bias	Bias(%)
Default LOM	2.182(0.661)	0.650	0.957	0.019	0.88
CLOM(X5X6)	2.508(0.744)	0.738	0.931	0.132	5.00
CLOM(X7X8)	2.277(0.724)	0.709	0.867	0.363	13.75
CLOM(X5678)	2.719(0.854)	0.841	0.949	0.079	2.99
MSLOM(X7X8)	2.124(0.707)	0.702	0.948	0.039	1.80
MSLOM(X1278)	2.143(0.846)	0.825	0.940	0.020	0.92
MSLOM(X3478)	2.141(0.746)	0.741	0.949	0.022	1.02
MSLOM(X5678)	2.121(0.694)	0.702	0.953	0.042	1.94
MSLOM(X123478)	2.158(0.890)	0.869	0.946	0.005	0.23
MSLOM(all)	2.155(0.876)	0.868	0.945	0.008	0.37
DR	2.735(0.908)	0.411	0.645	0.095	3.60

 $RERI_{OR}^{true} = 2.64$; Empirical: $RERI_{OR} = 2.163$, $RERI_{RR} = 1.849$

 $\beta_0 = -3$: prevalence ≈ 0.18

Model	β_3	\widehat{SE}_{β_3}	coverage	Abs. Bias	Bias(%)
Default LOM	1.915(0.415)	0.420	0.959	0.000	0.00
CLOM(X5X6)	2.367(0.500)	0.507	0.889	0.273	10.34
CLOM(X7X8)	2.082(0.470)	0.473	0.705	0.558	21.14
CLOM(X5678)	2.683(0.602)	0.604	0.956	0.043	1.63
MSLOM(X7X8)	1.860(0.445)	0.453	0.946	0.055	2.87
MSLOM(X1278)	1.875(0.538)	0.531	0.942	0.040	2.09
MSLOM(X3478)	1.868(0.474)	0.477	0.949	0.047	2.45
MSLOM(X5678)	1.857(0.434)	0.452	0.952	0.058	3.03
MSLOM(X123478)	1.881(0.564)	0.557	0.948	0.034	1.78
MSLOM(all)	1.877(0.553)	0.557	0.951	0.038	1.98
DR	2.691(0.648)	0.294	0.648	0.051	1.93

 $RERI_{OR}^{true} = 2.64$; Empirical: $RERI_{OR} = 1.915$, $RERI_{RR} = 1.362$

 $\beta_0 = -2$: prevalence ≈ 0.4

Model	eta_3	\widehat{SE}_{β_3}	coverage	Abs. Bias	Bias(%)
Default LOM	1.730(0.325)	0.317	0.950	0.003	0.17
CLOM(X5X6)	2.277(0.418)	0.405	0.810	0.363	13.75
CLOM(X7X8)	1.949(0.377)	0.367	0.503	0.691	26.17
CLOM(X5678)	2.678(0.515)	0.497	0.948	0.038	1.44
MSLOM(X7X8)	1.677(0.346)	0.344	0.939	0.050	2.90
MSLOM(X1278)	1.684(0.409)	0.403	0.947	0.043	2.49
MSLOM(X3478)	1.677(0.363)	0.362	0.944	0.050	2.90
MSLOM(X5678)	1.676(0.336)	0.344	0.948	0.051	2.95
MSLOM(X123478)	1.684(0.426)	0.422	0.949	0.043	2.49
MSLOM(all)	1.682(0.418)	0.423	0.953	0.045	2.61
DR	2.684(0.554)	0.243	0.638	0.044	1.67

 $RERI_{OR}^{true} = 2.64$; Empirical: $RERI_{OR} = 1.727$, $RERI_{RR} = 0.888$

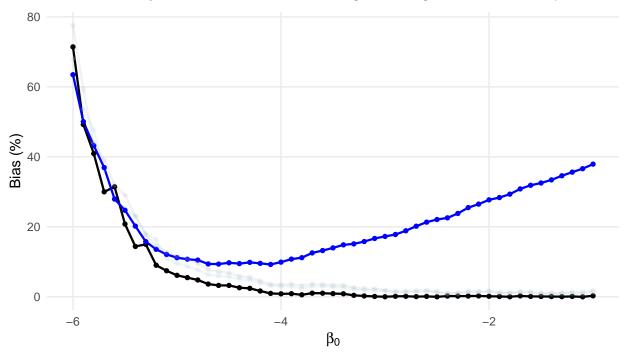
 $\beta_0 = -1$: prevalence ≈ 0.6

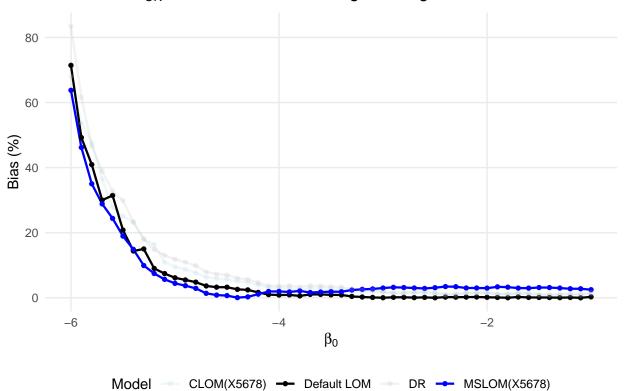
Model	β_3	\widehat{SE}_{β_3}	coverage	Abs. Bias	Bias(%)
Default LOM	1.655(0.286)	0.289	0.954	0.005	0.30
CLOM(X5X6)	2.233(0.373)	0.377	0.766	0.407	15.42
CLOM(X7X8)	1.892(0.331)	0.337	0.397	0.748	28.33
CLOM(X5678)	2.673(0.461)	0.467	0.953	0.033	1.25
MSLOM(X7X8)	1.611(0.307)	0.316	0.952	0.039	2.36
MSLOM(X1278)	1.606(0.367)	0.369	0.947	0.044	2.67
MSLOM(X3478)	1.610(0.324)	0.332	0.952	0.040	2.42
MSLOM(X5678)	1.609(0.296)	0.316	0.961	0.041	2.48
MSLOM(X123478)	1.605(0.384)	0.387	0.947	0.045	2.73
MSLOM(all)	1.604(0.375)	0.387	0.955	0.046	2.79
DR	2.683(0.504)	0.229	0.637	0.043	1.63

 $RERI_{OR}^{true} = 2.64$; Empirical: $RERI_{OR} = 1.650$, $RERI_{RR} = 0.511$

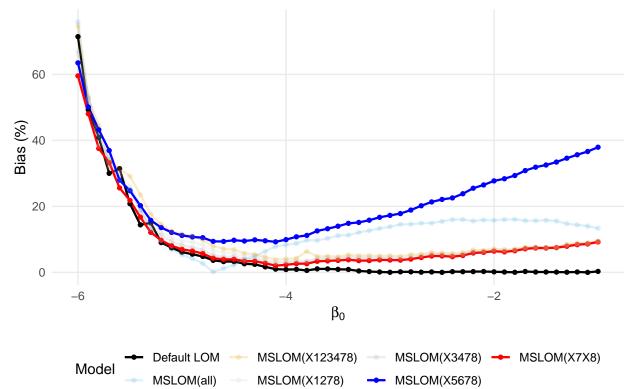
Graphs

Bias of $\mathsf{RERI}_\mathsf{OR}$ across models, with weights using control data only









Bias of $\mathsf{RERI}_\mathsf{OR}$ across MSLOM models, with weights using both data

