Supplementary File 4

Simulation experiment 2, results

This file provides the results of simulation experiment 2, referred to in section 3 and 4 in the main text of "A comparison of full model specification and backward elimination of potential confounders when estimating marginal and conditional causal effects on binary outcomes from observational data", by Kim Luijken, Susanne Strohmaier, Maarten van Smeden, Rolf H.H. Groenwold, and Georg Heinze. To facilitate replicability of simulation results, we present some additional descriptives of the simulation ouput.

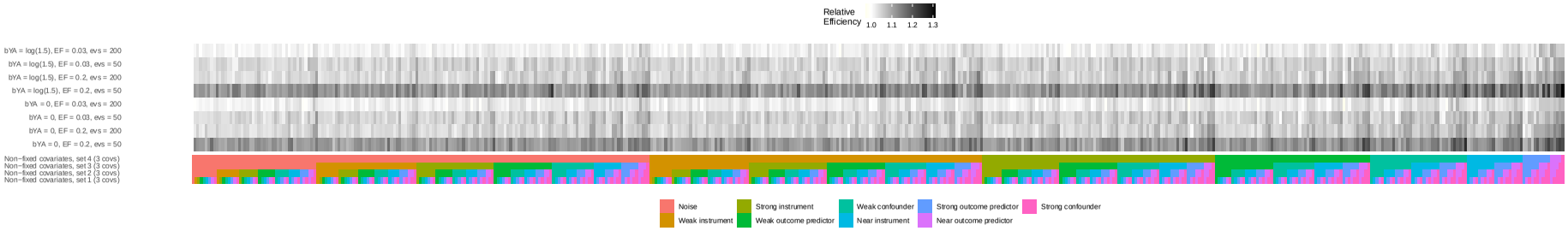


Figure 1: The relative efficiency of the marginal risk ratio estimated using Firth's corrected Logistic regression with intercept-correction between the full and selected model in each of the 3960 scenarios. The relative efficiency is computed as $\frac{MSE_{selected}}{MSE_{full}}$

Marginal Risk Ratio, Firth's corrected Logistic regression with intercept-correction

Values

			Mean			Min			Max			Warnings		
bYA	event fraction	nevents	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.
0.00	0.20	50.00	0.09	0.12	0.59	-0.02	-0.00	0.09	0.31	0.37	1.46	0	0	0
0.00	0.20	200.00	0.10	0.12	0.61	-0.02	-0.01	0.16	0.32	0.34	1.66	0	0	0
0.00	0.03	50.00	0.13	0.15	0.67	-0.02	-0.00	0.11	0.41	0.47	2.02	0	0	0
0.00	0.03	200.00	0.13	0.15	0.65	-0.03	-0.01	0.16	0.41	0.44	1.60	0	0	0
log(1.5)	0.20	50.00	0.22	0.25	0.70	-0.01	0.00	0.15	0.41	0.45	1.57	0	0	0
log(1.5)	0.20	200.00	0.22	0.24	0.72	-0.01	-0.00	0.09	0.41	0.45	1.64	0	0	0
log(1.5)	0.03	50.00	0.22	0.25	0.78	-0.03	0.00	0.19	0.40	0.43	1.65	0	0	0
log(1.5)	0.03	200.00	0.22	0.24	0.77	-0.01	0.00	0.11	0.40	0.44	2.04	0	0	0

Table 1: The values of the log(mRR) estimated using FLIC. Each row represents 495 scenarios with varying associations between the covariates and the exposure and/or outcome.

Bias

			Mean			Min			Max			Warnings		
bYA	event fraction	nevents	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.
0.00	0.20	50	-0.00	0.03	0.49	-0.04	-0.00	0.09	0.02	0.10	1.27	0	0	0
0.00	0.20	200	-0.00	0.02	0.51	-0.02	-0.01	0.10	0.03	0.07	1.66	0	0	0
0.00	0.03	50	-0.00	0.03	0.54	-0.04	-0.00	0.11	0.03	0.09	1.67	0	0	0
0.00	0.03	200	0.00	0.02	0.52	-0.03	-0.01	0.12	0.02	0.08	1.28	0	0	0
log(1.5)	0.20	50	-0.00	0.03	0.48	-0.04	-0.01	0.12	0.03	0.09	1.20	0	0	0
log(1.5)	0.20	200	-0.00	0.02	0.49	-0.04	-0.01	0.09	0.02	0.08	1.28	0	0	0
log(1.5)	0.03	50	-0.01	0.03	0.56	-0.04	-0.01	0.09	0.03	0.09	1.64	0	0	0
log(1.5)	0.03	200	-0.01	0.02	0.54	-0.04	-0.01	0.11	0.01	0.07	1.69	0	0	0

Table 2: The bias of the log(mRR) estimated using FLIC. Each row represents 495 scenarios with varying associations between the covariates and the exposure and/or outcome.

Mean Squared Error

			Mean			Min			Max			Warnings		
bYA	event fraction	nevents	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.
0.00	0.20	50	0.06	0.06	0.35	0.02	0.02	0.06	0.10	0.12	1.65	0	0	0
0.00	0.20	200	0.04	0.04	0.36	0.01	0.01	0.03	0.08	0.09	2.78	0	0	0
0.00	0.03	50	0.08	0.08	0.43	0.01	0.01	0.04	0.20	0.21	3.01	0	0	0
0.00	0.03	200	0.04	0.05	0.36	0.01	0.01	0.03	0.11	0.13	1.74	0	0	0
log(1.5)	0.20	50	0.06	0.07	0.33	0.01	0.01	0.04	0.15	0.16	1.59	0	0	0
log(1.5)	0.20	200	0.05	0.05	0.34	0.01	0.01	0.02	0.15	0.16	1.78	0	0	0
log(1.5)	0.03	50	0.07	0.07	0.45	0.02	0.02	0.07	0.14	0.16	2.85	0	0	0
log(1.5)	0.03	200	0.04	0.04	0.41	0.02	0.02	0.03	0.10	0.11	2.89	0	0	0

Table 3: The mean squared error (MSE) of the log(mRR) estimated using FLIC. Each row represents 495 scenarios with varying associations between the covariates and the exposure and/or outcome.

Conditional Odds Ratio, Firth's corrected Logistic regression with intercept-correction

Values

			Mean			Min			Max			Warnings		
bYA	event fraction	nevents	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.
0.00	0.20	50.00	0.14	0.18	0.69	-0.03	-0.01	0.12	0.44	0.58	1.77	0	0	0
0.00	0.20	200.00	0.13	0.16	0.70	-0.03	-0.01	0.20	0.42	0.45	1.70	0	0	0
0.00	0.03	50.00	0.13	0.17	0.72	-0.02	-0.01	0.12	0.43	0.51	2.07	0	0	0
0.00	0.03	200.00	0.14	0.16	0.71	-0.03	-0.01	0.20	0.42	0.45	1.64	0	0	0
log(1.5)	0.20	50.00	0.27	0.31	0.81	-0.03	-0.00	0.19	0.44	0.48	1.61	0	0	0
log(1.5)	0.20	200.00	0.27	0.30	0.82	-0.02	-0.00	0.11	0.44	0.50	1.76	0	0	0
log(1.5)	0.03	50.00	0.27	0.31	0.85	-0.03	-0.00	0.20	0.44	0.53	1.69	0	0	0
log(1.5)	0.03	200.00	0.27	0.30	0.84	-0.01	0.00	0.11	0.44	0.51	2.08	0	0	0

Table 4: The values of the log(cOR) estimated using FLIC. Each row represents 495 scenarios with varying associations between the covariates and the exposure and/or outcome.

Bias

			Mean			Min			Max			Warnings		
bYA	event fraction	nevents	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.
0.20	50	0.00	0.04	0.55	-0.03	-0.01	0.12	0.03	0.18	1.36	0	0	0	0.00
0.00	0.20	200	-0.00	0.03	0.56	-0.03	-0.01	0.12	0.04	0.09	1.70	0	0	0
0.00	0.03	50	-0.00	0.03	0.58	-0.04	-0.01	0.12	0.03	0.10	1.66	0	0	0
0.00	0.03	200	0.00	0.03	0.58	-0.03	-0.01	0.12	0.02	0.09	1.43	0	0	0
log(1.5)	0.20	50	-0.00	0.04	0.54	-0.04	-0.01	0.12	0.03	0.13	1.40	0	0	0
log(1.5)	0.20	200	-0.00	0.03	0.55	-0.03	-0.00	0.11	0.03	0.09	1.35	0	0	0
log(1.5)	0.03	50	0.00	0.04	0.58	-0.03	-0.00	0.10	0.03	0.12	1.69	0	0	0
log(1.5)	0.03	200	0.00	0.03	0.57	-0.03	0.00	0.11	0.03	0.11	1.68	0	0	0

Table 5: The bias of the log(cOR) estimated using FLIC. Each row represents 495 scenarios with varying associations between the covariates and the exposure and/or outcome.

Mean Squared Error

			Mean			Min			Max			Warnings		
bYA	event fraction	nevents	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.
0.00	0.20	50	0.12	0.13	0.44	0.02	0.02	0.07	0.24	0.31	2.01	0	0	0
0.00	0.20	200	0.07	0.08	0.44	0.02	0.02	0.04	0.22	0.25	2.92	0	0	0
0.00	0.03	50	0.09	0.10	0.48	0.03	0.03	0.07	0.26	0.28	2.98	0	0	0
0.00	0.03	200	0.06	0.06	0.44	0.02	0.02	0.04	0.15	0.17	2.07	0	0	0
log(1.5)	0.20	50	0.11	0.11	0.43	0.03	0.03	0.06	0.25	0.29	2.11	0	0	0
log(1.5)	0.20	200	0.06	0.07	0.41	0.03	0.03	0.04	0.18	0.20	1.87	0	0	0
log(1.5)	0.03	50	0.10	0.11	0.49	0.02	0.02	0.07	0.21	0.23	2.99	0	0	0
log(1.5)	0.03	200	0.07	0.08	0.44	0.02	0.02	0.03	0.23	0.27	2.86	0	0	0

Table 6: The mean squared error of the log(cOR) estimated using FLIC. Each row represents 495 scenarios with varying associations between the covariates and the exposure and/or outcome.

Marginal Risk Ratio estimated using Maximum Likelihood

Values

			Mean			Min			Max			Warnings		
bYA	event fraction	nevents	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.
0.00	0.20	50.00	0.10	0.13	0.60	-0.02	-0.01	0.09	0.35	0.38	1.50	0	0	0
0.00	0.20	200.00	0.10	0.12	0.61	-0.02	-0.01	0.16	0.33	0.34	1.67	0	0	0
0.00	0.03	50.00	0.14	0.17	0.68	-0.02	-0.00	0.11	0.44	0.50	2.10	9	9	9
0.00	0.03	200.00	0.13	0.15	0.66	-0.02	-0.00	0.16	0.41	0.44	1.61	0	0	0
log(1.5)	0.20	50.00	0.23	0.26	0.72	-0.02	0.00	0.15	0.44	0.47	1.62	0	0	0
log(1.5)	0.20	200.00	0.23	0.25	0.73	-0.01	-0.00	0.09	0.44	0.48	1.70	0	0	0
log(1.5)	0.03	50.00	0.24	0.26	0.79	-0.02	0.01	0.20	0.41	0.44	1.70	0	0	0
log(1.5)	0.03	200.00	0.23	0.25	0.77	-0.01	0.00	0.11	0.40	0.45	2.05	1	1	1

Table 7: The values of the log(mRR) estimated using ML. Each row represents 495 scenarios with varying associations between the covariates and the exposure and/or outcome. In 10 datasets, the following warning occured when the unadjusted/full/selected model was fitted: "simpleWarning: glm.fit: fitted probabilities numerically 0 or 1 occurred". We discarded these datasets, which implies that 10 mRR estimates are missing.

Bias

			Mean			Min			Max			Warnings		
bYA	event fraction	nevents	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.
0.00	0.20	50.00	0.01	0.03	0.50	-0.02	-0.01	0.09	0.04	0.11	1.28	0	0	0
0.00	0.20	200.00	0.00	0.02	0.51	-0.02	-0.01	0.10	0.04	0.07	1.67	0	0	0
0.00	0.03	50.00	0.01	0.04	0.56	-0.02	-0.00	0.11	0.06	0.14	1.75	9	9	9
0.00	0.03	200.00	0.01	0.03	0.53	-0.02	-0.00	0.12	0.03	0.10	1.31	0	0	0
log(1.5)	0.20	50.00	0.01	0.03	0.49	-0.02	0.00	0.13	0.05	0.09	1.25	0	0	0
log(1.5)	0.20	200.00	0.01	0.03	0.50	-0.01	-0.00	0.09	0.05	0.12	1.34	0	0	0
log(1.5)	0.03	50.00	0.01	0.04	0.57	-0.02	-0.00	0.09	0.04	0.12	1.70	0	0	0
log(1.5)	0.03	200.00	0.01	0.03	0.55	-0.01	0.00	0.11	0.04	0.07	1.70	1	1	1

Table 8: The bias of the log(mRR) estimated using ML. Each row represents 495 scenarios with varying associations between the covariates and the exposure and/or outcome. In 10 datasets, the following warning occured when the unadjusted/full/selected model was fitted: "simpleWarning: glm.fit: fitted probabilities numerically 0 or 1 occurred". We discarded these datasets, which implies that 10 mRR estimates are missing.

Mean Squared Error

			Mean			Min			Max			Warnings		
bYA	event fraction	nevents	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.
0.00	0.20	50	0.07	0.07	0.36	0.02	0.02	0.07	0.13	0.13	1.71	0	0	0
0.00	0.20	200	0.04	0.04	0.37	0.01	0.01	0.03	0.11	0.10	2.82	0	0	0
0.00	0.03	50	0.08	0.09	0.46	0.01	0.01	0.04	0.23	0.25	3.30	9	9	9
0.00	0.03	200	0.05	0.05	0.37	0.01	0.01	0.03	0.13	0.14	1.85	0	0	0
log(1.5)	0.20	50	0.07	0.07	0.35	0.01	0.01	0.05	0.17	0.19	1.74	0	0	0
log(1.5)	0.20	200	0.05	0.05	0.35	0.01	0.01	0.02	0.17	0.19	1.96	0	0	0
log(1.5)	0.03	50	0.08	0.08	0.47	0.02	0.02	0.07	0.17	0.19	3.06	0	0	0
log(1.5)	0.03	200	0.05	0.05	0.42	0.02	0.02	0.03	0.12	0.12	2.96	1	1	1

Table 9: The MSE of the log(mRR) estimated using ML. Each row represents 495 scenarios with varying associations between the covariates and the exposure and/or outcome. In 10 datasets, the following warning occured when the unadjusted/full/selected model was fitted: "simpleWarning: glm.fit: fitted probabilities numerically 0 or 1 occurred". We discarded these datasets, which implies that 10 mRR estimates are missing.

Conditional Odds Ratio estimated using Maximum Likelihood

Values

			Mean			Min			Max			Warnings		
bYA	event fraction	nevents	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.
0.00	0.20	50.00	0.16	0.18	0.70	-0.03	-0.01	0.12	0.53	0.62	1.82	0	0	0
0.00	0.20	200.00	0.14	0.16	0.70	-0.04	-0.02	0.20	0.43	0.46	1.71	0	0	0
0.00	0.03	50.00	0.15	0.18	0.73	-0.02	-0.01	0.12	0.58	0.68	2.27	9	9	9
0.00	0.03	200.00	0.14	0.17	0.72	-0.03	-0.00	0.21	0.43	0.46	1.65	0	0	0
log(1.5)	0.20	50.00	0.28	0.31	0.83	-0.03	-0.00	0.19	0.47	0.51	1.66	0	0	0
log(1.5)	0.20	200.00	0.29	0.31	0.83	-0.02	-0.00	0.11	0.48	0.54	1.77	0	0	0
log(1.5)	0.03	50.00	0.30	0.32	0.86	-0.03	0.01	0.20	0.51	0.56	1.74	0	0	0
log(1.5)	0.03	200.00	0.30	0.31	0.85	-0.01	0.00	0.11	0.53	0.55	2.10	1	1	1

Table 10: The values of the log(cOR) estimated using ML. Each row represents 495 scenarios with varying associations between the covariates and the exposure and/or outcome. In 10 datasets, the following warning occured when the unadjusted/full/selected model was fitted: "simpleWarning: glm.fit: fitted probabilities numerically 0 or 1 occurred". We discarded these datasets, which implies that 10 mRR estimates are missing.

Bias

			Mean			Min			Max			Warnings		
bYA	event fraction	nevents	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.
0.00	0.20	50	0.03	0.05	0.57	-0.03	-0.01	0.12	0.13	0.21	1.41	0	0	0
0.00	0.20	200	0.00	0.03	0.57	-0.04	-0.02	0.12	0.05	0.09	1.71	0	0	0
0.00	0.03	50	0.01	0.04	0.60	-0.02	-0.01	0.12	0.17	0.27	1.87	9	9	9
0.00	0.03	200	0.00	0.03	0.58	-0.03	-0.00	0.12	0.03	0.11	1.44	0	0	0
log(1.5)	0.20	50	0.01	0.04	0.56	-0.03	-0.00	0.13	0.06	0.13	1.44	0	0	0
log(1.5)	0.20	200	0.02	0.04	0.56	-0.02	-0.00	0.11	0.07	0.13	1.36	0	0	0
log(1.5)	0.03	50	0.03	0.05	0.59	-0.03	0.00	0.11	0.10	0.15	1.74	0	0	0
log(1.5)	0.03	200	0.03	0.04	0.58	-0.01	0.00	0.11	0.12	0.14	1.70	1	1	1

Table 11: The bias of the log(cOR) estimated using ML. Each row represents 495 scenarios with varying associations between the covariates and the exposure and/or outcome. In 10 datasets, the following warning occured when the unadjusted/full/selected model was fitted: "simpleWarning: glm.fit: fitted probabilities numerically 0 or 1 occurred". We discarded these datasets, which implies that 10 mRR estimates are missing.

Mean Squared Error

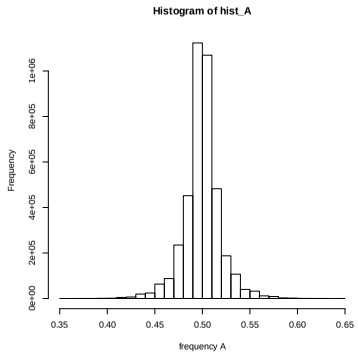
			Mean			Min			Max			Warnings		
bYA	event fraction	nevents	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.	Full	Selected	Unadj.
0.00	0.20	50	0.16	0.14	0.46	0.02	0.02	0.07	0.40	0.38	2.16	0	0	0
0.00	0.20	200	0.10	0.09	0.45	0.02	0.02	0.04	0.33	0.29	2.97	0	0	0
0.00	0.03	50	0.12	0.12	0.53	0.03	0.03	0.07	4.02	4.08	7.51	9	9	9
0.00	0.03	200	0.06	0.06	0.45	0.02	0.02	0.04	0.18	0.19	2.10	0	0	0
log(1.5)	0.20	50	0.14	0.13	0.45	0.03	0.03	0.06	0.39	0.34	2.22	0	0	0
log(1.5)	0.20	200	0.07	0.07	0.42	0.03	0.03	0.04	0.22	0.23	1.96	0	0	0
log(1.5)	0.03	50	0.13	0.12	0.51	0.02	0.02	0.07	0.32	0.28	3.21	0	0	0
log(1.5)	0.03	200	0.10	0.09	0.46	0.02	0.02	0.03	1.37	1.33	2.93	1	1	1

Table 12: The MSE of the log(cOR) estimated using ML. Each row represents 495 scenarios with varying associations between the covariates and the exposure and/or outcome. In 10 datasets, the following warning occured when the unadjusted/full/selected model was fitted: "simpleWarning: glm.fit: fitted probabilities numerically 0 or 1 occurred". We discarded these datasets, which implies that 10 mRR estimates are missing.

Descriptives data

Frequency of exposure A

bYA	event fraction	nevents	Mean (sd)	Min	Max
0.00	0.20	50	0.5(0.032)	0.35	0.64
0.00	0.20	200	0.5(0.016)	0.43	0.57
0.00	0.03	50	0.5(0.012)	0.45	0.56
0.00	0.03	200	0.5(0.006)	0.47	0.53
log(1.5)	0.20	50	0.5(0.032)	0.35	0.64
log(1.5)	0.20	200	0.5(0.016)	0.43	0.58
log(1.5)	0.03	50	0.5(0.012)	0.44	0.56
log(1.5)	0.03	200	0.5(0.006)	0.47	0.53



Frequency of outcome Y

bYA	event fraction	nevents	Mean (sd)	Min	Max
0.00	0.20	50	0.2(0.025)	0.09	0.32
0.00	0.20	200	0.2(0.013)	0.14	0.26
0.00	0.03	50	0.03(0.004)	0.01	0.05
0.00	0.03	200	0.03(0.002)	0.02	0.04
log(1.5)	0.20	50	0.2(0.025)	0.09	0.32
log(1.5)	0.20	200	0.2(0.013)	0.15	0.26
log(1.5)	0.03	50	0.03(0.004)	0.01	0.05
log(1.5)	0.03	200	0.03(0.002)	0.02	0.04