



Universiteit
Leiden

The Netherlands

The impact of epidemiologic methods on findings in studies of causal effects and prediction modelling

Luijken, K.

Citation

Luijken, K. (2022, May 19). *The impact of epidemiologic methods on findings in studies of causal effects and prediction modelling*. Retrieved from <https://hdl.handle.net/1887/3304345>

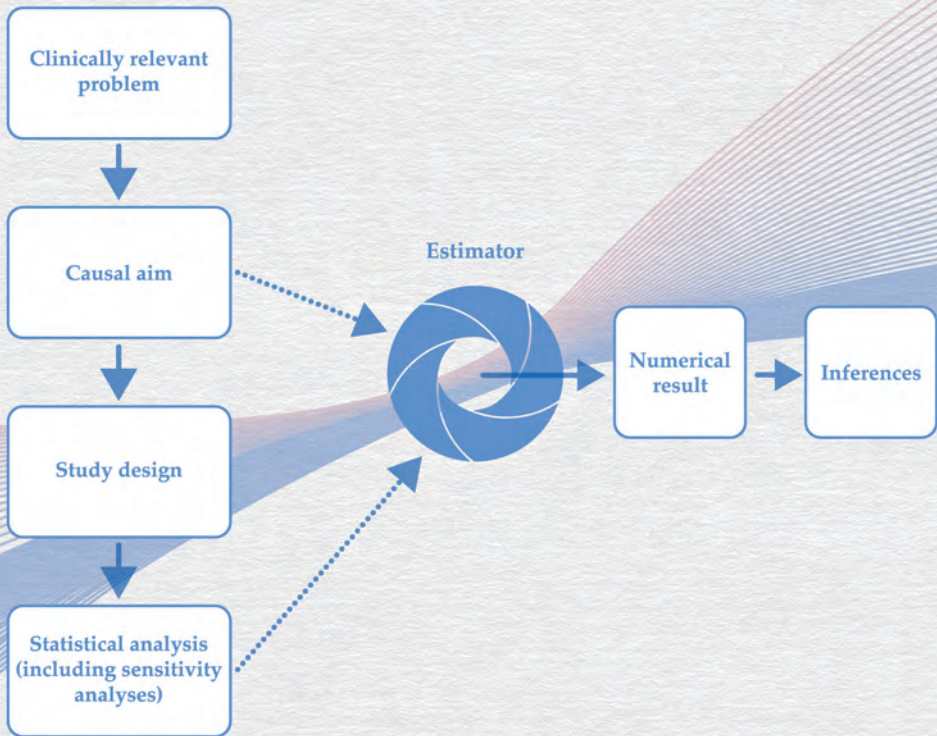
Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3304345>

Note: To cite this publication please use the final published version (if applicable).

4



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

A common view in epidemiology is that automated confounder selection methods, such as backward elimination, should be avoided as they can lead to biased and overly precise effect estimates. Nevertheless, backward elimination remains regularly applied. We investigated if and under which conditions causal effect estimation in observational studies can improve by using backward elimination on a prespecified set of potential confounders. An expression was derived that quantifies how variable omission relates to bias and variance of effect estimators. Additionally, 3,960 scenarios were defined and investigated by simulations comparing bias and mean squared error (MSE) of the conditional log odds ratio, $\log(\text{cOR})$, and the marginal log risk ratio, $\log(\text{mRR})$, between full models including all prespecified covariates and backward elimination of these covariates. Applying backward elimination resulted in a mean bias of 0.03 for $\log(\text{cOR})$ and 0.02 for $\log(\text{mRR})$, compared to 0.56 and 0.52 for $\log(\text{cOR})$ and $\log(\text{mRR})$, respectively, for a model without any covariate adjustment, and no bias for the full model. In less than 3% of the scenarios considered, the MSE of the $\log(\text{cOR})$ or $\log(\text{mRR})$ was slightly lower (max 3%) when backward elimination was used compared to the full model. When an initial set of potential confounders can be specified based on background knowledge, there is minimal added value of backward elimination. We advise not to use it and otherwise to provide ample arguments supporting its use.

This chapter was based on: Luijken K, Groenwold RHH, van Smeden M, Strohmaier S, Heinze G. 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. *Biometrical Journal* (in press).

1 | Background

Identification of causal effects from observational data relies on proper control for confounding. It is generally advised that confounders are determined based on the causal structure of the data, about which one may possess background knowledge, or one could at least make defensible assumptions¹⁻⁴, and that automated covariate selection methods, such as stepwise selection and backward elimination, should be avoided as they can lead to seriously biased estimated effect sizes and underestimation of statistical uncertainty by model-based confidence intervals (CIs)^{5,6}.

Despite these warnings, automated selection procedures for selection of confounders remain widely applied⁷⁻¹³. One reason for the popularity may be that, at least in theory, using automated selection as an add-on to selection of potential confounders based on background knowledge may lead to improved efficiency^{2,14,15}. For instance, backward elimination has occasionally been reported to improve estimation in terms of mean squared error (MSE) of the effect estimator¹⁶. Limited guidance exists about when backward elimination could be beneficial in observational studies in which confounding adjustment is needed¹⁷⁻¹⁹.

The aim of the current study was to extend recommendations for practicing statisticians on the use or avoidance of automated variable selection for descriptive models provided by Heinze and colleagues⁵ to a causal inference context. Specifically, we compare the efficiency of causal effect estimation by multivariable modelling in observational studies when fitting a model with all potential confounders (full model) compared to using backward elimination (see Box 1). Out of the many available methods for variable (or confounder) selection, we focus on backward elimination, since it is widely implemented in statistical software packages and is often considered superior to alternatives such as univariable screening, forward and stepwise selection²⁰. We focus on outcome-oriented selection of confounders, meaning that exposure-oriented selection procedures, for instance as part of propensity score methods, are beyond the scope of this article. Furthermore, we assume that sufficient clinical expertise is available to specify an outcome model with covariates presumably related to the exposure and/or outcome. This model is assumed to include at least all such covariates and to correctly specify all nonlinear covariate-outcome relations but may include covariates only related to the exposure (instruments) and/or true confounders, or irrelevant covariates.

In Section 2, we present comparative analyses of a motivating example. In Section 3, we discuss arguments in favor of and against the use of backward elimination as a means of automated selection among potential confounders. In Section 4, we perform simulation studies to investigate whether there is a benefit of using a backward-elimination estimator compared to a full-model estimator to estimate the target causal effect. We end with a discussion of the implications for clinical research.

Box 1. Motivation to compare backward elimination of potential confounders neutrally with a full model

- After identifying a set of potential confounders, uncertainty about the causal role of some covariates may remain. Although backward elimination can reduce the adjustment set, it assumes that the initial set of potential confounders is a sufficient adjustment set.
- The disjunctive cause criterion by VanderWeele and Shpitser can guide confounder selection¹⁵. This criterion states to control for covariates that are either a cause of the exposure or a cause of the outcome, which may lead to adjustment for instrumental variables. Therefore, they recommended implementing backward elimination or forward selection to eliminate such variables. On the other hand, Vansteelandt and colleagues argue that instrumental variables should not necessarily be eliminated from the adjustment set, because the uncertainty they introduce on the estimated exposure effect may reflect lack of information about the effect of interest¹⁸.
- Greenland and colleagues proposed to compare a model adjusted for a sufficient set of confounders where one confounder is deleted by hand to a full model by estimating the change in mean squared error (MSE) which was illustrated in an empirical data set¹⁴. Since similar bias and variance considerations apply to backward elimination, it is worthwhile to compare a full model and use of backward elimination in more settings.
- Backward elimination has been reported to improve estimation in terms of MSE of the effect estimator¹⁶.

2 | Motivating example: coronary artery bypass grafting study

We illustrate confounder selection using a study that investigated the causal effect of a computer tomography angiography (CTA) examination of the main coronary artery prior to coronary artery bypass grafting (CABG) surgery on the postoperative stroke risk of a patient²¹. We used a simulated dataset based on the empirical data (details in²¹) that was previously used for methodological work²². In the simulated dataset, the sample size and relationships between the variables were preserved and similar to the original dataset. In Online Supplementary File 1, we provide Rcode to allow replication of this example.

2.1 | Defining causal estimands

We defined two research questions and the corresponding estimands²³. The first research question compared the risk of post-operative stroke for patients with known characteristics when refraining from screening for aortic disease using CTA prior to CABG surgery versus the risk when patients were screened using CTA. The causal contrast, no CTA screening versus CTA screening given a set of characteristics, can, for instance, be expressed as a conditional risk difference, a conditional risk ratio or a conditional odds ratio (cOR). We defined the estimand as the cOR.

The second question of interest concerned the effect of not exposing an entire target population to CTA screening versus exposing everyone to CTA screening. The causal contrast could be expressed as a marginal risk difference, a marginal risk ratio (mRR) or a marginal odds ratio. We defined the estimand as the mRR.

2.2 | Linking the observed data to the estimand

Whether a causal effect can be identified from observational data depends, among other things, on the extent to which confounding can be adjusted for. Heinze and colleagues recommended to generate an initial working set of covariates based on clinical expertise and background knowledge, without yet using the dataset at hand^{5,24}. In studies of causal inference, it is often helpful to visualize assumed causal dependencies between covariates, where the level of formalization of those dependencies may sometimes reach that of a directed acyclic graph (DAG) (we refer to²⁵ for recommendations on implementation). In doing so, a researcher explicates knowledge about variables that are *irrelevant* to the study question, as leaving out variables is a stronger assumption

than including them. Accordingly, for covariates that are included in an initial working set, many decisions are still to be made regarding their causal role and relevance.

In the original study²¹, the initial working set contained 23 measured covariates that described the health state of a patient just before the decision to perform CTA or not. Detailed causal assumptions that could be represented in a DAG were not supported by the cross-sectional assessment of these covariates, but we could exclude collider stratification bias or presence of mediators when using these covariates as a confounding adjustment set.

2.3 | Defining causal estimands

We estimated the cOR by the exponentiated regression coefficient of no CTA screening in a multivariable logistic regression model with Firth's correction²⁶⁻²⁸ (CIs based on profile penalized likelihood) including the 23 covariates specified in the initial working set. We estimated the mRR based on predictions of potential outcomes from that multivariable logistic regression model²⁹⁻³¹ (CIs based on 500 bootstrap samples using the percentile method). Additionally, we applied data-driven selection of the 23 prespecified covariates by means of backward elimination at a significance level of 0.157 approximating selection by the Akaike information criterion⁵. For the backward-elimination estimator, we contrasted 'selected-model' CIs, which condition on the finally selected covariates, to 'global' bootstrap CI, where the selection process was repeated in each bootstrap resample. The selected-model CIs were based on profile penalized likelihood for the cOR and computed from fitting the finally selected model in 500 bootstrap samples using the percentile method for the mRR.

In this example, backward elimination reduced the adjustment set by eight potential confounders. While for both cOR and mRR the full-model CIs were wider than the (invalid) selected-model CIs, the global bootstrap CIs were the widest (Table 1). Clearly, additional variability arises from the uncertainty in the selection which must be captured by repeating the selection process in each bootstrap resample. Heinze and colleagues proposed to evaluate bias and added uncertainty by two bootstrap-based measures, relative conditional bias (RCB) and root mean squared difference ratio (RMSDR)^{5,32}. In our example, the RCB for the log cOR was -1.3% and RMSDR was 1.06. The RCB for the log mRR was -2.6% and RMSDR was 1.07. These measures also indicated a possible variance inflation by using backward elimination.

Table 1 Results for the CABG study.

Estimand	Model	Estimate	Confidence interval estimation approach	95% Confidence interval	Confidence interval width (upper/lower)
cOR	Full	4.48	PPL	[2.10, 10.01]	4.77
	Selected	3.83	Invalid: selected-model PPL	[1.99, 7.77]	3.90
			Global bootstrap	[2.10, 11.25]	5.36
mRR	Full	3.68	Bootstrap	[2.02, 7.09]	3.51
	Selected	3.24	Invalid: selected-model bootstrap	[1.87, 6.19]	3.31
			Global bootstrap	[1.93, 7.17]	3.72

Abbreviations: cOR = conditional odds ratio, mRR = marginal risk ratio, PPL = profile penalized likelihood

3 | Use of automated covariate selection

3.1 | Arguments in favor of automated selection of confounders

Bias and variance of an effect estimator can be combined in a single measure; the MSE. The MSE can be interpreted as the expected value of the squared distance of an estimate to the true value, which can be alternatively expressed as $MSE = bias^2 + variance$. For a linear regression model, the value of omitting a covariate in terms of reducing the MSE of an effect estimator can be quantified directly (see Appendix). We provide a simplified representation of this principle here that extends to settings with binary outcomes.

Consider a setting with an outcome, an exposure and one covariate. The effect of the exposure on the outcome is evaluated under two estimation strategies: ‘always include the covariate’ (full) versus ‘always omit the covariate’ (omit). Assuming that the bias in the exposure effect estimator of the ‘full’ strategy is 0, in terms of MSE we find a benefit in omitting the covariate if, for the effect of the exposure on the outcome, the following inequality holds:

$$Bias_{omit}^2 < Variance_{full} - Variance_{omit} \quad (1)$$

If (1) holds, the reduced variance of the ‘omit’ strategy outweighs the increase in squared bias, and thus there is a benefit of omitting the covariate in terms of MSE, and hence produces a more efficient estimate. If we ignore a possible small sample

bias^{33,34}, only the righthand side of (1) is inversely proportional to sample size. Thus, there should be a threshold sample size n , such that (1) holds for all values smaller than that n . Figure 1 illustrates this phenomenon. Figure A3 and A4 (in Appendix) illustrate that n increases with a stronger association between the exposure and the covariate, with a weaker association between the outcome and covariate, and with a lower variance of the exposure variable.

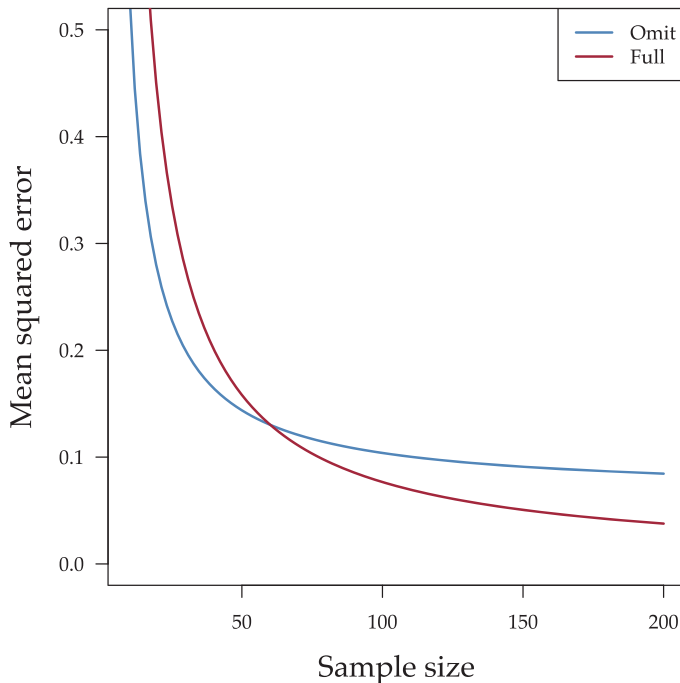


Figure 1. Illustration of the bias-variance trade-off for the ordinary least squares estimator of the exposure effect when including (Full) or omitting (Omit) covariate L . The blue and red line are computed using expressions for the mean squared error under the ‘Omit’ and ‘Full’ strategy, respectively (Appendix). The value of n for which the reduced variance by omitting L outweighs the increase in squared bias is around 60. This illustrates inequality (1) in the main text for a linear model and ordinary least squares estimation. For sample sizes < 60 , omission of the covariate resulted in a lower mean squared error of the exposure effect estimator in a linear setting.

3.2 | Arguments against automated selection of confounders

Selection of variables by statistical procedures is sometimes incorrectly thought to be a prerequisite for model building³⁵. However, a ‘statistically significant’ result neither confirms whether a covariate is indeed a confounder, nor does insignificance prove that it is not. A well-known counterargument against use of data-driven selection of confounders is that the causal structure of the data cannot be derived from observed

associations only. For example, a covariate has a different causal status being a *confounder* compared to being a *mediator*, but in both cases, it may be statistically associated with the exposure and/or outcome. Automated covariate selection procedures based on statistical associations only could result in inappropriate adjustment, selection bias or reduction of precision of the exposure effect estimate^{35,36}.

It has been claimed that post-selection inference cannot be valid at all⁶. Since research on this issue is ongoing^{37,38}, neutral comparison studies and user-friendly implementations are still lacking³⁹ and hence its advances are hardly accessible to epidemiologists. Frequentist statistical theory assumes that the parameters to be estimated in a model are fixed before observing the data, while variable selection involves the data in the selection process, meaning the model is not fixed a priori. Consequently, CIs based on the selected model are no longer valid and often underestimate uncertainty in the effect estimator^{5,20,37}.

Finally, there is no one-size-fits-all implementation of automated covariate selection^{5,20} and recommendations on covariate selection may not be applicable to a particular study. Choices regarding covariate selection should strongly depend on the aim of a study, which could be causal inference, prediction, or description^{40,41}. Statistical texts that explain variable selection do not always relate implementation of the procedure to those distinct research aims⁴⁰.

4 | Simulations

4.1 | Simulation design

Aim: We examined the effect of backward elimination vs. full model specification on the efficiency of causal effect evaluation in simulation studies. First, we performed a proof-of-concept simulation (Experiment 1) to confirm inequality (1). Additionally, we studied the value of backward elimination in efficiency of causal effect estimation in more complex and realistic settings (Experiment 2).

Data-generating mechanisms: The generated data consisted of a binary outcome, Y , a binary exposure, A , and a set of continuous covariates, L . The set of covariates was free of mediators and colliders and was the starting point for all backward elimination procedures. In Experiment 1, the generated data contained a single continuous covariate next to the exposure and outcome. The exposure effect was null, the sample size was set to 60 or 120 and the event fraction (i.e., $Pr(Y = 1)$) was set to 0.5 or 0.2. The conditional

associations $A-L$ and $Y-L$ varied between 0 and 0.5 on a log-odds scale. A total of 144 scenarios were evaluated. In Experiment 2, the $\log(\text{cOR})$ of the exposure was either $\log(1)$ or $\log(1.5)$. L consisted of 24 continuous covariates from a multivariate normal distribution with mean 0 and a variance-covariance matrix with 1s on the diagonal and 0.3 on all off-diagonal elements. The set consisted of a mix of 12 - 24 true confounders, 0 - 12 (near) instrumental variables, 0 - 12 (near) predictors of the outcome, and 0 - 12 noise variables, where the number of each covariate type was varied across simulation scenarios (see Table 2). The expected number of events was set to 50 or 200 and the expected event fraction was set to 0.2 or 0.03, resulting in samples with 250; 1,667; 1,000; or 6,667 observations. Table 2 presents the values of other simulation parameters. A total of 3,960 scenarios were evaluated.

Target estimand: The estimands were the cOR and the mRR of the association between A and Y , controlled for confounding.

Methods: The cOR was obtained from logistic regression models estimated using Firth's Logistic regression with Intercept Correction (FLIC) to avoid introduction of finite sample bias²⁶⁻²⁸ and issues with separation in the simulation⁴². The mRR was estimated using FLIC models that estimated potential outcomes²⁹⁻³¹. Estimates were evaluated on a logarithmic scale because of the asymmetrical nature of ORs and RRs.

Table 2 Simulation parameters of experiment 2.

Parameter	Value
Conditional exposure-outcome effect	0, $\log(1.5)$
Fixed confounders: conditional log odds ratio confounder-exposure association*	$\log(1.05)$
Fixed confounders: conditional log odds ratio confounder-outcome association*	$\log(1.05)$
Mixture of covariates: conditional log odds ratio covariate-exposure association (4 sets of 3 covariates)*	0, $\log(1.05)$, $\log(1.2)$
Mixture of covariates: conditional log odds ratio covariate-outcome association (4 sets of 3 covariates)*	0, $\log(1.05)$, $\log(1.2)$
Covariate correlation across all 24 covariates	0.3
Number of events	50, 200
Expected event fraction	0.2, 0.03

* Of the 24 continuous covariates, 12 were assumed to be fixed confounders, and 12 represented a mixture of true confounders ($\log(1.2)$), (near-) instrumental variables ($\log(1.05)$), (near-)predictors of the outcome ($\log(1.05)$) and noise variables (0). In each data set, the number of respective covariate types was determined by the combination of conditional covariate-exposure/outcome parameters.

Simulations were performed using R statistical software version 3.6.2.⁴³ using the package `logistf`⁴⁴ to implement Firth's correction. In Experiment 1, the MSE of the $\log(\widehat{cOR})$ and $\log(\widehat{mRR})$ was evaluated under two estimation strategies: 'always include covariate L' versus 'always omit covariate L'. In Experiment 2, we evaluated the MSE of the $\log(\widehat{cOR})$ and $\log(\widehat{mRR})$ obtained using a full model versus using backward elimination with cut-off value $p = 0.157$ (corresponding with using the Akaike information criterion)²⁴. We obtained the true mRR for each scenario by a large sample approximation ($N = 1,000,000$).

Performance measures: The MSE was defined as the average squared difference between the estimated $\log(\widehat{cOR})$ and true $\log(cOR)$ or the estimated $\log(\widehat{mRR})$ and true $\log(mRR)$ averaged per scenario over the simulation runs (10,000 for Experiment 1; 1,000 for Experiment 2). We compared the full and selected model in terms of relative efficiency of the $\log(\widehat{cOR})$ and $\log(\widehat{mRR})$, which was computed as a ratio of the MSE obtained from the backward elimination procedure divided by the MSE obtained from the full model.

This simulation design was reported following previous recommendations⁴⁵. All R code for simulations is available at https://github.com/Kluijken/CI_CovSel.

4.2 | Simulation results

Experiment 1: Inequality (1) held for most (90%) of the simulated scenarios in Experiment 1 for the $\log(\widehat{cOR})$ and for 20% of the scenarios regarding the $\log(\widehat{mRR})$. Hence, regarding the cOR, omitting the covariate was often more beneficial in terms of MSE than including it (see Online Supplementary File 2). Regarding the mRR, including the covariate was often more beneficial in terms of MSE than omitting it (Figure 2). Omitting the covariate was beneficial in terms of MSE only when the covariate was an instrument or a near-instrument. The benefit of omitting was larger when the event fraction was lower, 0.2 instead of 0.5, and, as expected, when sample size was lower, 60 compared to 120.

Experiment 2: In Experiment 2, the median relative efficiency of the $\log(\widehat{cOR})$ across all scenarios was 1.04, indicating the MSE was on average lower for the full model compared to the selected model. The median relative efficiency of the $\log(\widehat{mRR})$ across all scenarios was 1.05. Across all 3,960 scenarios, the bias of the full model was zero for both the $\log(\widehat{cOR})$ and $\log(\widehat{mRR})$, whereas the average bias across all scenarios of the backward

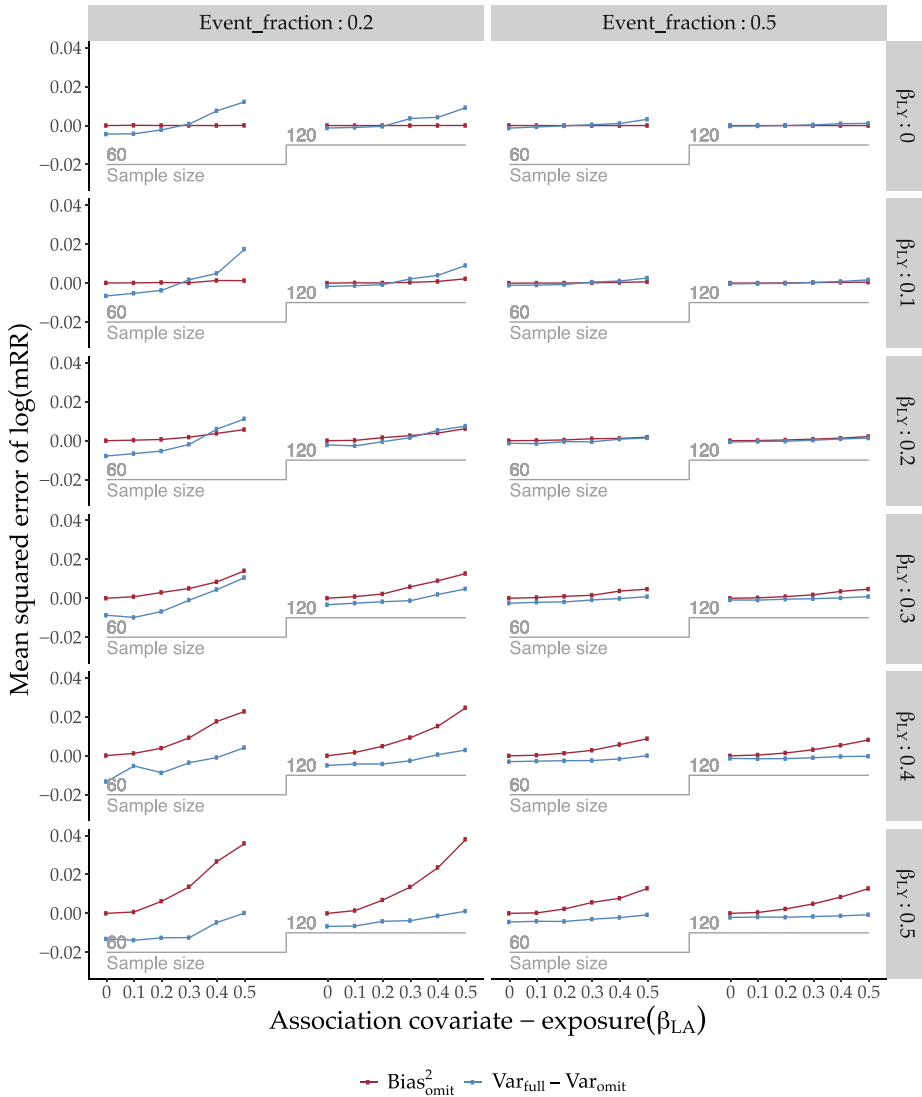


Figure 2. Results of simulation Experiment 1 for the marginal risk ratio (mRR). A single covariate L acts as a confounder, (near-)instrumental variable, (near-)predictor of the outcome, or noise variable in a setting where a binary exposure has a true null effect on a binary outcome. The squared bias and difference in variance is compared when L is always included or always omitted, illustrating principle (1) in the main text. β_{LA} and β_{LY} refer to the conditional log odds ratio of the covariate-exposure and covariate-outcome association, respectively. This figure was created using the `looplot` package⁴⁶.

eliminated model was 0.03 for the $\log(\widehat{cOR})$ and 0.02 for the $\log(\widehat{mRR})$, compared to 0.56 and 0.52 for $\log(\widehat{cOR})$ and $\log(\widehat{mRR})$, respectively, for a model without any covariate adjustment. We found 112 scenarios (2.8%) for cOR and 47 scenarios (1.1%) for mRR in which the MSE was lower for the selected than the full models.

Closer examination of the 112 scenarios in which the \widehat{cOR} estimated using backward elimination showed lower MSE than the full model revealed that 100 scenarios included at least three full instrumental variables and 37 scenarios included at least three noise variables (see Table 3). In these scenarios, the increased efficiency remained small, with a minimal relative efficiency of 0.97, meaning that the MSE when backward elimination was applied was only 3% lower than the MSE of the full model in the most beneficial setting. For the $\log(\widehat{mRR})$, we found that 42 scenarios included at least three full instrumental variables and 18 scenarios included at least three noise variables (see Table 4). Again, the increased efficiency remained small, with a minimal relative efficiency of 0.97. Full results of the simulations are presented in Online Supplementary File 3.

5 | Discussion

Our simulation results show that, compared to estimating a model with all prespecified confounders, application of backward elimination was unlikely to reduce the MSE of the exposure effect estimator (defined as the cOR and the mRR), while introducing a bias. We identified some settings in which the MSE of the effect estimators was lower with backward elimination than without, yet the reduction in MSE was small. The results are driven by two antagonist effects; an MSE-reducing effect of omitting weak confounders, and an MSE-increasing effect caused by additional uncertainties incurred by applying automated selection as explained by Heinze et al (2018).

Our findings support and extend previous recommendations on automated covariate selection. VanderWeele and Shpitser proposed to use the disjunctive cause criterion for confounder selection¹⁵. This criterion states to control for covariates that are either a cause of the exposure or a cause of the outcome, which may lead to adjustment for instrumental variables. Therefore, they recommended to implement backward elimination or forward selection to eliminate such variables. Our findings provide weak support for the use of variable selection in this case. In an overview and classification of covariate selection strategies, Witte and Didelez found that backward elimination performed well in terms of bias in the effect estimator in settings that

Table 3 Summary of simulation experiment 2; results for the conditional odds ratio (cOR). Each row represents 495 scenarios with varying associations between the covariates and the exposure and/or outcome. Mean bias indicates the average bias of the log(cOR) for the full and backward eliminated model, respectively. Relative efficiency of the mean squared error (MSE) of the cOR is computed as a ratio of the backward elimination MSE divided by the full model MSE.

Conditional exposure effect	Event fraction	Number of events	Mean bias full	Mean bias BE	Median relative efficiency	Minimum relative efficiency	Maximum relative efficiency	Number of scenarios BE < full	Number of scenarios MSE(cOR) 3 IVs in DGM	At least 3 variables in DGM	No IVs or noise in DGM
0	0.20	50	0.00	0.04	1.08	1.00	1.20	0	0	0	0
0	0.20	200	0.00	0.02	1.03	0.98	1.13	10	9	3	0
0	0.03	50	0.00	0.03	1.03	0.98	1.10	6	6	1	0
0	0.03	200	0.00	0.02	1.02	0.97	1.08	42	40	14	0
log(1.5)	0.20	50	0.00	0.06	1.09	1.02	1.27	0	0	0	0
log(1.5)	0.20	200	0.00	0.03	1.04	0.98	1.13	8	8	3	0
log(1.5)	0.03	50	0.00	0.04	1.04	0.99	1.10	7	5	3	1
log(1.5)	0.03	200	0.00	0.02	1.02	0.98	1.09	39	32	13	3
Overall results			0.00	0.03	1.04	0.97	1.27	112	100	37	4

Abbreviations: BE = backward elimination, cOR = conditional odds ratio, DGM = data-generating mechanism, IV = instrumental variable, MSE = mean squared error.

Table 4 Summary of simulation Experiment 2; results for the marginal risk ratio (mRR). Each row represents 495 scenarios with varying associations between the covariates and the exposure and/or outcome. Mean bias indicates the average bias of the log(mRR) for the full and backward eliminated model, respectively. Relative efficiency of the mean squared error (MSE) of the mRR is computed as a ratio of the backward elimination MSE divided by the full model MSE.

Conditional exposure effect	Event fraction	Number of events	Mean bias full	Mean bias BE	Median relative efficiency	Minimum relative efficiency	Maximum relative efficiency	Number of scenarios BE < full	At least 3 IVs in DGM	At least 3 noise variables in DGM	No IVs or noise in DGM
0	0.20	50	0.00	0.03	1.12	1.06	1.25	0	0	0	0
0	0.20	200	0.00	0.02	1.05	1.00	1.17	0	0	0	0
0	0.03	50	0.00	0.03	1.05	1.00	1.12	0	0	0	0
0	0.03	200	0.00	0.02	1.02	0.97	1.09	21	19	8	0
log(1.5)	0.20	50	-0.02	0.03	1.12	1.05	1.31	0	0	0	0
log(1.5)	0.20	200	0.00	0.02	1.05	1.00	1.15	1	1	0	0
log(1.5)	0.03	50	-0.01	0.03	1.05	1.01	1.12	0	0	0	0
log(1.5)	0.03	200	0.00	0.02	1.02	0.98	1.10	25	22	10	0
Overall results			0.00	0.02	1.05	0.97	1.31	47	42	18	0

Abbreviations: BE = backward elimination, DGM = data-generating mechanism, IV = instrumental variable, mRR = marginal risk ratio, MSE = mean squared error.

contained strong confounders and instrumental variables and did not perform well when applied to a sufficient adjustment set in which each confounder was responsible for a small degree of confounding¹⁷. We found similar patterns in terms of the MSE of the effect estimator, irrespective of conditional or marginal effects are of interest. On the other hand, Vansteelandt and colleagues recommended against the use of automated covariate selection even when there is a potential efficiency gain by excluding an instrumental variable, because this would prevent overstating the precision with which a causal effect is known¹⁸. Summarizing, the true number of irrelevant covariates and instruments included in the prespecified set of adjustment variables, and the strength of association of true confounders with the outcome greatly affect the relative performance of applying backward elimination. In practice, these conditions are usually unknown, but the more domain expertise is available to define the set, the less a researcher has to rely on data-driven selection.

Our motivating example was typical for clinical observational studies where a set of covariates is available that accurately describes the health state of a subject just before the decision to perform an intervention or not, but where dependencies among these covariates are difficult to assess. Therefore, we only assumed that the set of covariates was free of mediators and that there was no unmeasured confounding. These assumptions were based on clinical expertise and allowed specification of an initial working set without explicitly specifying a full DAG. Under these conditions, backward elimination was applied to potentially increase the efficiency of the effect estimate by setting weak covariate effects to zero, but not to change the underlying assumptions.

A limitation of our study is that we did not consider scenarios in which clinical expertise is not available. In many clinical settings, it is questionable whether the assumption of no residual confounding really holds. Furthermore, it is difficult to judge to what extent preselection can be reliably done. This depends on the novelty of a research field and often one will rely on previous research to derive assumptions. Doing so, researchers should be aware of inappropriate methodology, such as questionable conclusions stemming from observed bivariate associations, which typically do not reflect multivariable relations represented in a causal network³⁶. It is up to the researcher to explain to what extent pre-processing based on background knowledge is possible and hence whether data-driven selection could be of added value.

Additionally, as our paper was intended to evaluate a common practice, we did not consider more sophisticated approaches for data-driven confounder selection. Whereas backward elimination is an outcome-oriented selection procedure, other approaches,

such as Lasso penalized regression approaches^{47,48}, take into account both covariate-outcome and covariate-exposure relations. Such approaches might lead to more robust and efficient effect estimation compared to backward elimination; however, they are hardly ever used in epidemiological studies. We also excluded augmented backward elimination¹⁶ and other novel approaches as we were either involved in developing these methods or lack the necessary expertise to apply them routinely.

We conclude that backward elimination for confounder selection is unlikely to have added value when an initial set of covariates related to the exposure and/or outcome can be specified based on background knowledge. If researchers choose to perform backward elimination of potential confounders, selection should be justified, e.g., because a large number of potential confounders are anticipated to function as (near-)instruments, and the approach should be prespecified in a statistical analysis plan. Covariate selection based solely on statistical criteria should be avoided due to the possible selection of mediators and colliders. Irrespective of whether or not covariate selection strategies are being applied, we recommend to always provide information about the assumed causal structure, ideally by a depiction of assumed causal dependencies, but at least by excluding mediators and the possibility of unmeasured confounding.

Appendix

Mean squared error of ordinary least squares exposure effect estimator of full and reduced model

Notation and set-up

Consider the model depicted in Figure A1. Let A denote the exposure, L a covariate and Y the outcome. Each variable is a linear combination of the variables affecting it (indicated by the directed arrows in Figure 1) plus an error term. The coefficients of the model are denoted α for the relation between A and L , γ for the relation between Y and L conditional on A and β for the relation between Y and A conditional on L . All variables are normally distributed, where $L \sim \mathcal{N}(\mu_L, \sigma_L^2)$, $A \sim \mathcal{N}(\mu_A, \sigma_A^2)$, and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$.

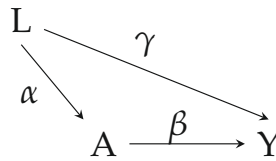


Figure A1: Directed acyclic graph.

To explore when elimination of a covariate improves efficiency of an ordinary least squares estimator of the association between exposure and the outcome (in short exposure effect estimator), we derive expressions for the MSE of the regression coefficient of Y on A when L is included in the model (full model), $\text{MSE}(\hat{\beta})$, and for the model where L is omitted (reduced model), $\text{MSE}(\hat{\beta}_{omit})$. The expression for bias of the exposure effect estimator is derived using path-tracing rules, as described by Wright⁴⁹ and Pearl⁵⁰. The variance of the exposure effect estimator is specified for an ordinary least squares estimator in a finite sample with n observations, similar to the expressions for the extended omitted variable framework by Cinelli and Hazlett⁵¹.

Full model

Consider an ordinary least squares regression of the model in Figure A1,

$$Y = \hat{\beta}A + \hat{\gamma}L + \hat{\epsilon},$$

where Y is an $n \times 1$ vector containing the outcome of interest for each of the n observations, A is an $n \times 1$ vector of the continuous exposure variable, L is an $n \times 1$ vector of the continuous covariate, $\hat{\beta}$ and $\hat{\gamma}$ are ordinary least squares coefficients of the association between Y and A and between Y and L , respectively, and $\epsilon \sim \mathcal{N}(0, \sigma_{Y^{\perp A, L}}^2)$. Let $Y^{\perp A, L}$ denote the variable Y after removing the components linearly explained by A and L , $A^{\perp L}$ denote the variable A after removing the components linearly explained by L , $\hat{\alpha}$ denote the ordinary least squares estimator of the association between A and L and $\text{var}(\cdot)$ and $\text{cov}(\cdot)$ denote the sample variances and covariances, respectively⁵¹. Then, using path-tracing rules^{49,50}

$$\begin{aligned} \text{var}(L) &:= \hat{\sigma}_L^2 \\ \text{var}(A) &:= \hat{\sigma}_A^2 \\ \text{cov}(A, L) &:= \hat{\sigma}_L^2 \hat{\alpha} \\ \text{var}(A^{\perp L}) &:= \hat{\sigma}_A^2 - \hat{\sigma}_L^2 \hat{\alpha}^2 \\ \text{var}(Y) &:= \hat{\sigma}_Y^2 \\ \text{cov}(Y^{\perp L}, A^{\perp L}) &:= (\hat{\sigma}_A^2 - \hat{\sigma}_L^2 \hat{\alpha}^2) \hat{\beta} \\ \text{var}(Y^{\perp A, L}) = \hat{\sigma}_{Y^{\perp A, L}}^2 &:= \hat{\sigma}_Y^2 - \hat{\sigma}_A^2 \hat{\beta}^2 - \hat{\sigma}_L^2 \hat{\gamma}^2 - 2\hat{\sigma}_L^2 \hat{\alpha} \hat{\beta} \hat{\gamma}. \end{aligned}$$

As described in^{49,50}, partial regression coefficients can be readily read from a path diagram such as Figure A1. The expected value for the partial regression coefficient regressing Y on A given L , $\hat{\beta}$, can be expressed as

$$\begin{aligned} \hat{\beta} &= \frac{\text{cov}(Y^{\perp L}, A^{\perp L})}{\text{var}(A^{\perp L})}, \text{ and} \\ \mathbb{E}(\hat{\beta}) &= \frac{(\sigma_A^2 - \sigma_L^2 \alpha^2) \beta}{\sigma_A^2 - \sigma_L^2 \alpha^2} \\ &= \beta. \end{aligned} \tag{1}$$

Hence, the bias of the effect estimator in the full model is $\mathbb{E}(\hat{\beta}) - \beta = 0$.

Let df denote the ordinary least squares regression's degrees of freedom. An expression for the variance of $\hat{\beta}$ is⁵¹

$$\begin{aligned}\text{var}(\hat{\beta}) &= \frac{\text{var}(Y^{\perp A, L})}{\text{var}(A^{\perp L})} \frac{1}{\text{df} - 1} \\ &= \frac{\hat{\sigma}_Y^2 - \hat{\sigma}_A^2 \hat{\beta}^2 - \hat{\sigma}_L^2 \hat{\gamma}^2 - 2\hat{\sigma}_L^2 \hat{\alpha} \hat{\beta} \hat{\gamma}}{\hat{\sigma}_A^2 - \hat{\sigma}_L^2 \hat{\alpha}^2} \frac{1}{n - 3}.\end{aligned}$$

To derive the expected variance of $\hat{\beta}$, we use the property that $\hat{\sigma}^2 \approx \hat{\sigma}^2 \frac{n-1}{n} \approx \sigma^2$ for $\hat{\sigma}_A^2$, $\hat{\sigma}_L^2$ and $\hat{\sigma}_Y^2$. Furthermore, we assume that the residuals $\sigma_{Y^{\perp A, L}}^2$ are independent from A and L and apply a first-order approximation⁵². Then,

$$\mathbb{E}[\text{var}(\hat{\beta})] \approx \frac{\sigma_Y^2 - \sigma_A^2 \beta^2 - \sigma_L^2 \gamma^2 - 2\sigma_L^2 \alpha \beta \gamma}{\sigma_A^2 - \sigma_L^2 \alpha^2} \frac{1}{n - 3}. \quad (2)$$

Since $MSE = bias^2 + variance$,

$$MSE(\hat{\beta}) \approx \frac{\sigma_Y^2 - \sigma_A^2 \beta^2 - \sigma_L^2 \gamma^2 - 2\sigma_L^2 \alpha \beta \gamma}{\sigma_A^2 - \sigma_L^2 \alpha^2} \frac{1}{n - 3}. \quad (3)$$

Reduced model

Consider the ordinary least squares regression of the model in Figure 1 where variable L is omitted

$$Y = \hat{\beta}_{omit} A + \hat{\epsilon}_{omit},$$

where Y is an $n \times 1$ vector containing the outcome of interest for each of the n observations, A is an $n \times 1$ continuous exposure variable, $\hat{\beta}_{omit}$ is an ordinary least squares coefficient estimator and $\epsilon_{omit} \sim \mathcal{N}(0, \sigma_{Y^{\perp A}}^2)$. Let $Y^{\perp A}$ denote the variable Y after removing the components linearly explained by A . Then,

$$\begin{aligned}\text{cov}(Y, A) &:= \hat{\sigma}_A^2 \hat{\beta} + \hat{\sigma}_L^2 \hat{\alpha} \hat{\gamma} \\ \text{var}(Y^{\perp A}) &= \hat{\sigma}_{Y^{\perp A}}^2 := \hat{\sigma}_Y^2 - \hat{\sigma}_A^2 \hat{\beta}^2 - 2\hat{\sigma}_L^2 \hat{\alpha} \hat{\beta} \hat{\gamma}.\end{aligned}$$

Again, reading partial regressions from Figure A1 as described in ^{49,50}, we find that the expected value for the marginal association between A and Y can be expressed as

$$\begin{aligned}\hat{\beta}_{omit} &= \frac{\text{cov}(Y, A)}{\text{var}(A)}, \text{ and} \\ \mathbb{E}(\hat{\beta}_{omit}) &= \frac{\sigma_A^2 \beta + \sigma_L^2 \alpha \gamma}{\sigma_A^2} \\ &= \beta + \frac{\sigma_L^2 \alpha \gamma}{\sigma_A^2}.\end{aligned} \quad (4)$$

Hence, the expected bias in the reduced model is $\mathbb{E}(\hat{\beta}_{omit}) - \beta = \frac{\sigma_L^2 \alpha \gamma}{\sigma_A^2}$. We obtain the expression for the variance of $\hat{\beta}_{omit}$ ⁵¹

$$\begin{aligned} \text{var}(\hat{\beta}_{omit}) &= \frac{\text{var}(Y^{\perp A})}{\text{var}(A)} \frac{1}{df} \\ &= \frac{\hat{\sigma}_Y^2 - \hat{\sigma}_A^2 \hat{\beta}^2 - 2\hat{\sigma}_L^2 \hat{\alpha} \hat{\beta} \hat{\gamma}}{\hat{\sigma}_A^2} \frac{1}{n-2}. \end{aligned}$$

To derive the expected variance of $\hat{\beta}_{omit}$, we use the property that $\hat{\sigma}^2 \approx \hat{\sigma}^2 \frac{n-1}{n} \approx \sigma^2$ for $\hat{\sigma}_A^2$ and $\hat{\sigma}_Y^2$. Furthermore, we assume that the residuals $\sigma_{Y^{\perp A}}^2$ are independent from A and apply a first-order approximation⁵². Then,

$$\mathbb{E}[\text{var}(\hat{\beta}_{omit})] \approx \frac{\sigma_Y^2 - \sigma_A^2 \beta^2 - 2\sigma_L^2 \alpha \beta \gamma}{\sigma_A^2} \frac{1}{n-2}, \quad (5)$$

where the final step is performed using a first-order approximation⁵² and under the assumption that the residuals $\sigma_{Y^{\perp A}}^2$ are independent from A . Since $MSE = bias^2 + variance$,

$$\begin{aligned} MSE(\hat{\beta}_{omit}) &\approx \left(\frac{\sigma_L^2 \alpha \gamma}{\sigma_A^2} \right)^2 + \frac{\sigma_Y^2 - \sigma_A^2 \beta^2 - 2\sigma_L^2 \alpha \beta \gamma}{\sigma_A^2} \frac{1}{n-2} \\ &\approx \frac{\sigma_L^4 \alpha^2 \gamma^2}{\sigma_A^4} + \frac{\sigma_Y^2 - \sigma_A^2 \beta^2 - 2\sigma_L^2 \alpha \beta \gamma}{\sigma_A^2} \frac{1}{n-2}. \end{aligned} \quad (6)$$

Comparison full and reduced model

Plotting equation (3) and equation (6) for the arbitrarily chosen values $\alpha = 0.4$, $\beta = 0.3$, $\gamma = 0.2$, $\sigma_A^2 = 2.5$, $\sigma_L^2 = 8$, and $\sigma_Y^2 = 10$ yields the following result.

The lines in Figure A2 show that $MSE(\hat{\beta}_{omit}) < MSE(\hat{\beta})$ for smaller sample size n , and that the full model has a lower MSE for $\hat{\beta}$ with larger sample sizes. In other words, if the inequality $Bias_{omit}^2 < Variance_{full} - Variance_{omit}$ holds, i.e., when

$$\frac{\sigma_L^4 \alpha^2 \gamma^2}{\sigma_A^4} < \frac{\sigma_Y^2 - \sigma_A^2 \beta^2 - \sigma_L^2 \gamma^2 - 2\sigma_L^2 \alpha \beta \gamma}{\sigma_A^2 - \sigma_L^2 \alpha^2} \frac{1}{n-3} - \frac{\sigma_Y^2 - \sigma_A^2 \beta^2 - 2\sigma_L^2 \alpha \beta \gamma}{\sigma_A^2} \frac{1}{n-2}$$

holds, then the reduced variance by omitting L outweighs the increase in squared bias. A simple approximation of the value of n for which this is the case, denoted n' , can be

found by assuming $n - 3 \approx n - 2 \approx n$ (which is reasonable to assume for sufficiently large sample sizes). This yields the following expression

$$n' < \frac{\sigma_A^4 \sigma_Y^2 - \sigma_A^6 \beta^2 - \sigma_A^4 \sigma_L^2 \gamma^2 - 2\sigma_A^4 \sigma_L^2 \alpha \beta \gamma}{(\sigma_A^2 - \sigma_L^2 \alpha^2) \sigma_L^4 \alpha^2 \gamma^2} - \frac{\sigma_A^2 \sigma_Y^2 - \sigma_A^4 \beta^2 - 2\sigma_A^2 \sigma_L^2 \alpha \beta \gamma}{\sigma_L^4 \alpha^2 \gamma^2}. \quad (7)$$

Similar to the plot above, Equation (7) indicates that for sample sizes smaller than the critical n' , omitting L results in a lower mean squared error of the exposure estimate than including it. The impact of many of the parameters in the equation depends on the values of other parameters. Their joint effect on the critical sample size could be assessed by means of plotting the result of equation for various parameters values, as is presented in Figures A3 and A4.

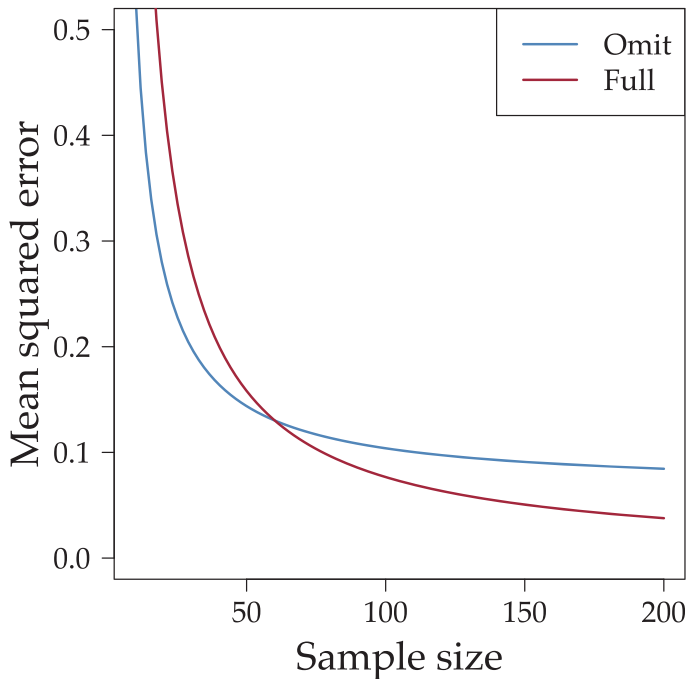


Figure A2. Illustration of the bias-variance trade-off for the ordinary least squares estimator of the exposure effect when including (Full) or omitting (Omit) covariate L . The red and blue line are computed using expression (3) and (6), respectively, for sample size n ranging from 0 to 200 and the values $\alpha = 0.4$, $\beta = 0.3$, $\gamma = 0.2$, $\sigma_A^2 = 2.5$, $\sigma_L^2 = 8$, and $\sigma_Y^2 = 10$. The value of n for which the reduced variance by omitting L outweighs the increase in squared bias is around 60.

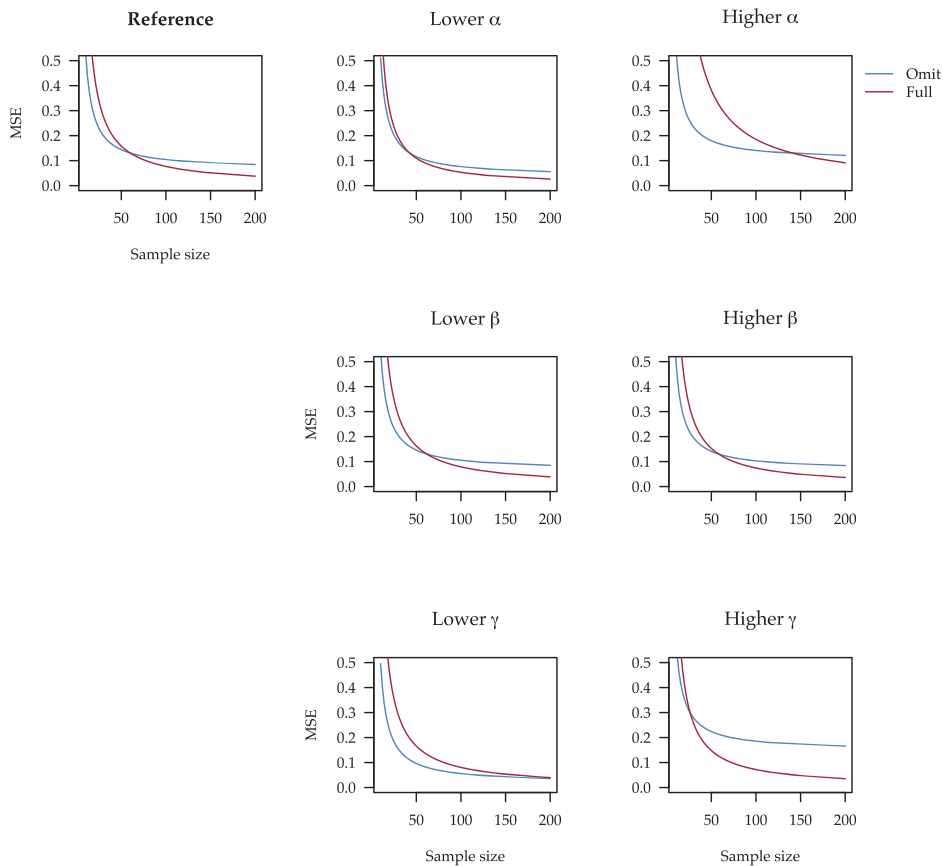


Figure A3. Illustration of the bias-variance trade-off for the ordinary least squares estimator of the exposure effect when including (Full) or omitting (Omit) covariate L . The red and blue line are computed using expression (3) and (6), respectively, for sample size n ranging from 0 to 200. The parameters σ_A^2 , σ_L^2 , and σ_Y^2 are fixed to values 2.5, 8 and 10, respectively, while the values of parameters α , β , and γ are varied. The reference plot is created using the arbitrarily chosen values $\alpha = 0.4$, $\beta = 0.3$, $\gamma = 0.2$, $\sigma_A^2 = 2.5$, $\sigma_L^2 = 8$, and $\sigma_Y^2 = 10$. Lower α indicates $\alpha = 0.3$ and higher $\alpha = 0.5$. Lower β indicates $\beta = 0.2$ and higher $\beta = 0.4$. Lower γ indicates $\gamma = 0.1$ and higher $\gamma = 0.3$.

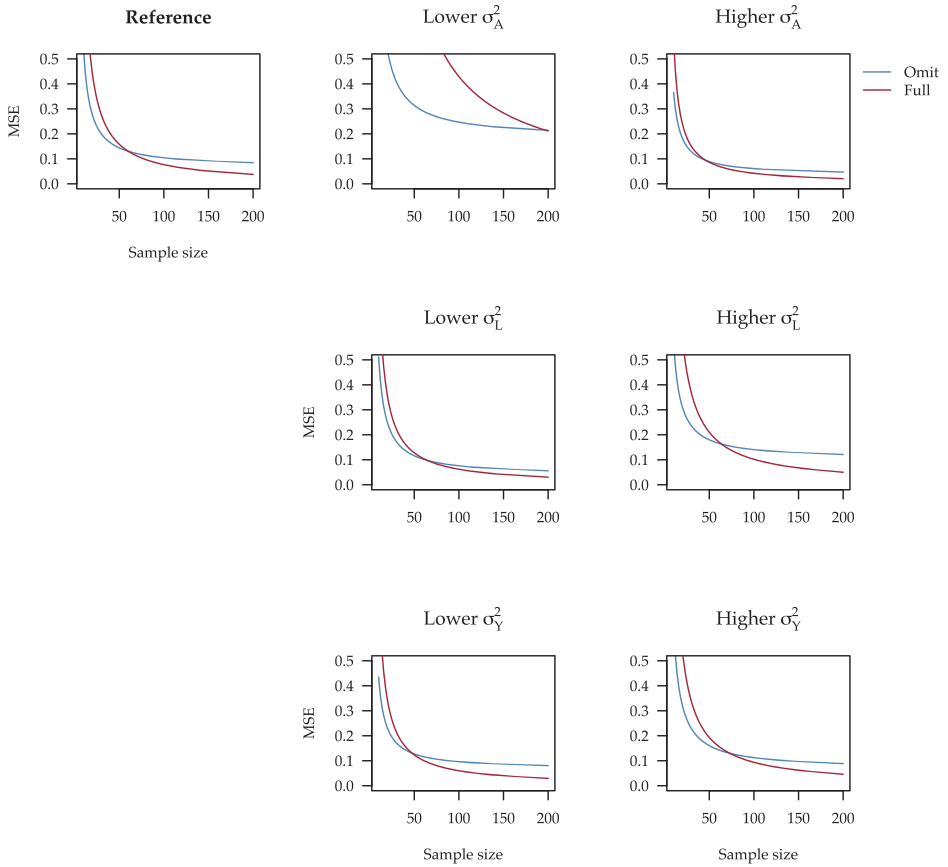


Figure A4. Illustration of the bias-variance trade-off for the ordinary least squares estimator of the exposure effect when including (Full) or omitting (Omit) covariate L . The red and blue line are computed using expression (3) and (6), respectively, for sample size n ranging from 0 to 200. The parameters α , β , and γ are fixed to values 0.4, 0.3 and 0.2, respectively, while the values of parameters σ_A^2 , σ_L^2 , and σ_γ^2 are varied. The reference plot is created using the arbitrarily chosen values $\alpha = 0.4$, $\beta = 0.3$, $\gamma = 0.2$, $\sigma_A^2 = 2.5$, $\sigma_L^2 = 8$, and $\sigma_\gamma^2 = 10$. Lower σ_A^2 indicates $\sigma_A^2 = 1.5$ and higher $\sigma_A^2 = 3.5$. Lower σ_L^2 indicates $\sigma_L^2 = 6$ and higher $\sigma_L^2 = 10$. Lower σ_γ^2 indicates $\sigma_\gamma^2 = 8$ and higher $\sigma_\gamma^2 = 12$.

Online Supplementary Files

The supplementary files referred to in this Chapter are available online at https://github.com/KLuijken/Dissertation_Online_Supplements/tree/main/Chapter_4

References

1. Hernán MA, Robins JM. *Causal inference: what if*. In: Boca Raton: Chapman & Hall/CRC; 2020.
2. VanderWeele TJ. Principles of confounder selection. *European Journal of Epidemiology*. 2019;34(3):211-219.
3. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003;14(3):300-306.
4. Ding P, Miratrix LW. To adjust or not to adjust? Sensitivity analysis of M-bias and butterfly-bias. *Journal of Causal Inference*. 2015;3(1):41-57.
5. Heinze G, Wallisch C, Dunkler D. Variable selection—a review and recommendations for the practicing statistician. *Biometrical Journal*. 2018;60(3):431-449.
6. Leeb H, Pötscher BM. Model selection and inference: Facts and fiction. *Econometric Theory*. 2005:21-59.
7. Talbot D, Massamba VK. A descriptive review of variable selection methods in four epidemiologic journals: there is still room for improvement. *European Journal of Epidemiology*. 2019;34(8):725-730.
8. Groenwold RH, Van Deursen AM, Hoes AW, Hak E. Poor quality of reporting confounding bias in observational intervention studies: a systematic review. *Annals of Epidemiology*. 2008;18(10):746-751.
9. Ali MS, Groenwold RH, Belitser SV, et al. Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. *Journal of Clinical Epidemiology*. 2015;68(2):122-131.
10. Walter S, Tiemeier H. Variable selection: current practice in epidemiological studies. *European Journal of Epidemiology*. 2009;24(12):733-736.
11. Pouwels KB, Widyakusuma NN, Groenwold RH, Hak E. Quality of reporting of confounding remained suboptimal after the STROBE guideline. *Journal of Clinical Epidemiology*. 2016;69:217-224.
12. Klein-Geltink J, Rochon P, Dyer S, Laxer M, Anderson G. Readers should systematically assess methods used to identify, measure and analyze confounding in observational cohort studies. *Journal of Clinical Epidemiology*. 2007;60(8):766. e1-11.
13. Hemkens LG, Ewald H, Naudet F, et al. Interpretation of epidemiologic studies very often lacked adequate consideration of confounding. *Journal of Clinical Epidemiology*. 2018;93:94-102.
14. Greenland S, Daniel R, Pearce N. Outcome modelling strategies in epidemiology: traditional methods and basic alternatives. *International Journal of Epidemiology*. 2016;45(2):565-575.
15. VanderWeele TJ, Shpitser I. A new criterion for confounder selection. *Biometrics*. 2011;67(4):1406-1413.
16. Dunkler D, Plischke M, Leffondré K, Heinze G. Augmented backward elimination: a pragmatic and purposeful way to develop statistical models. *PloS one*. 2014;9(11):e113677.
17. Witte J, Didelez V. Covariate selection strategies for causal inference: Classification and comparison. *Biometrical Journal*. 2019;61(5):1270-1289.
18. Vansteelandt S, Bekaert M, Claeskens G. On model selection and model misspecification in causal inference. *Statistical Methods in Medical Research*. 2012;21(1):7-30.
19. Greenland S. Invited commentary: variable selection versus shrinkage in the control of multiple confounders. *American Journal of Epidemiology*. 2008;167(5):523-529.
20. Sauerbrei W, Perperoglou A, Schmid M, et al. State of the art in selection of variables and functional forms in multivariable analysis—outstanding issues. *Diagnostic and Prognostic Research*. 2020;4:1-18.
21. Sandner SE, Nolz R, Loewe C, et al. Routine preoperative aortic computed tomography angiography is associated with reduced risk of stroke in coronary artery bypass grafting: a propensity-matched analysis. *European Journal of Cardio-Thoracic Surgery*. 2020;57(4):684-690.
22. Gregorich MG. *A comparison of methods for causal inference with a rare binary outcome* [Master thesis (unpublished)], Wien; 2018.
23. Goetghebeur E, le Cessie S, De Stavola B, Moodie EE, Waernbaum I. Topic group Causal Inference (TG7) of the STRATOS initiative. Formulating causal questions and principled statistical answers. *Statistics in medicine*. 2020;39(30):4922-4948.

24. Harrell Jr FE. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer; 2015.
25. Tennant P, Murray E, Arnold K, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *International Journal of Epidemiology*. 2020;dyaa213.
26. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. *Statistics in medicine*. 2002;21(16):2409-2419.
27. Firth D. Bias reduction of maximum likelihood estimates. *Biometrika*. 1993;80(1):27-38.
28. Puhr R, Heinze G, Nold M, Lusa L, Geroldinger A. Firth's logistic regression with rare events: accurate effect estimates and predictions? *Statistics in Medicine*. 2017;36(14):2302-2317.
29. Austin PC. Absolute risk reductions, relative risks, relative risk reductions, and numbers needed to treat can be obtained from a logistic regression model. *Journal of Clinical Epidemiology*. 2010;63(1):2-6.
30. Greenland S. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. *American Journal of Epidemiology*. 2004;160(4):301-305.
31. Localio AR, Margolis DJ, Berlin JA. Relative risks and confidence intervals were easily computed indirectly from multivariable logistic regression. *Journal of Clinical Epidemiology*. 2007;60(9):874-882.
32. Wallisch C, Dunkler D, Rauch G, De Bin R, Heinze G. Selection of variables for multivariable models: Opportunities and limitations in quantifying model stability by resampling. *Statistics in Medicine*. 2021;40(2):369-381.
33. Schaefer RL. Bias correction in maximum likelihood logistic regression. *Statistics in Medicine*. 1983;2(1):71-78.
34. Cordeiro GM, McCullagh P. Bias correction in generalized linear models. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1991;53(3):629-643.
35. Heinze G, Dunkler D. Five myths about variable selection. *Transplant International*. 2017;30(1):6-10.
36. Sun G-W, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *Journal of Clinical Epidemiology*. 1996;49(8):907-916.
37. Berk R, Brown L, Buja A, Zhang K, Zhao L. Valid post-selection inference. *The Annals of Statistics*. 2013;41(2):802-837.
38. Belloni A, Chernozhukov V, Wei Y. Post-selection inference for generalized linear models with many controls. *Journal of Business & Economic Statistics*. 2016;34(4):606-619.
39. Kammer M, Dunkler D, Michiels S, Heinze G. Evaluating methods for Lasso selective inference in biomedical research by a comparative simulation study. In. *arXiv preprint arXiv:2005.07484*2020.
40. Shmueli G. To explain or to predict? *Statistical Science*. 2010;25(3):289-310.
41. Hernán MA, Hsu J, Healy B. A second chance to get causal inference right: a classification of data science tasks. *Chance*. 2019;32(1):42-49.
42. van Smeden M, de Groot JA, Moons KG, et al. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Medical Research Methodology*. 2016;16(1):1-12.
43. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2020.
44. Heinze G, Ploner M, Jiricka L. *logistf: Firth's Bias-Reduced Logistic Regression*. R package version 1.24. 2020 <https://cemsii.meduniwien.ac.at/en/kb/science-research/software/statistical-software/fllogistf/>
45. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Statistics in Medicine*. 2019;38(11):2074-2102.
46. Kammer M. *looplot: Create nested loop plots*. R package version 0.5.0.9001. 2020 <https://github.com/matherealize/looplot/>
47. Ertefaie A, Asgharian M, Stephens DA. Variable selection in causal inference using a simultaneous penalization method. *Journal of Causal Inference*. 2018;6(1).

48. Wilson A, Reich BJ. Confounder selection via penalized credible regions. *Biometrics*. 2014;70(4):852-861.
49. Wright S. Correlation and causation. *Journal of Agricultural Research*. 1921;20:557-85.
50. Pearl J. Linear models: A useful “microscope” for causal analysis. *Journal of Causal Inference*. 2013;1(1):155-170.
51. Cinelli C, Hazlett C. Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2020;82(1):39-67.
52. Henckel L, Perković E, Maathuis MH. Graphical criteria for efficient total effect estimation via adjustment in causal linear models. *arXiv preprint arXiv:190702435*. 2019.

