OPEN

Correspondence and
requests for materials
should be addressed to
P.H.L. (paul.h.lee@
polyu.edu.hk)

Should we adjust for a confounder if empirical and theoretical criteria yield contradictory results? A simulation study

Paul H. Lee

School of Nursing, Hong Kong Polytechnic University.

Confounders can be identified by one of two main strategies: empirical or theoretical. Although confounder identification strategies that combine empirical and theoretical strategies have been proposed, the need for adjustment remains unclear if the empirical and theoretical criteria yield contradictory results due to random error. We simulated several scenarios to mimic either the presence or the absence of a confounding effect and tested the accuracy of the exposure-outcome association estimates with and without adjustment. Various criteria (significance criterion, Change-in-estimate(CIE) criterion with a 10% cutoff and with a simulated cutoff) were imposed, and a range of sample sizes were trialed. In the presence of a true confounding effect, unbiased estimates were obtained only by using the CIE criterion with a simulated cutoff. In the absence of a confounding effect, all criteria performed well regardless of adjustment. When the confounding factor was affected by both exposure and outcome, all criteria yielded accurate estimates without adjustment, but the adjusted estimates were biased. To conclude, theoretical confounders should be adjusted for regardless of the empirical evidence found. The adjustment for factors that do not have a confounding effect minimally effects. Potential confounders affected by both exposure and outcome should not be adjusted for.

Most evidence-based medical studies of causal relationship between an exposure and an outcome are based on observational data. However, the true causal effect between exposure and outcome is affected by confounders[1], defined as variables that are associated with both exposure and outcome, but influenced by neither[2]. Without proper adjustment for confounders, the crude exposure-outcome association will be a biased estimate of the true association. Once identified, a confounder may be adjusted for by controlling it via appropriate study design techniques, such as restriction and matching of risk factors[3], or by analyzing the exposure-outcome association with appropriate statistical techniques, such as regression[4]. The issue of confounding is not only limited to observational studies, as some researchers have even suggested that confounders should be adjusted for in analysis of randomized controlled trials data[5,6]. Therefore, confounder identification is an important issue in medical research.

Confounders can be identified by either empirical or theoretical strategies. Empirical strategies select a confounder based on objective criteria in the current working dataset. Examples of this strategy include forward, backward, and stepwise regression[4], and change-in-estimate (CIE) criterion[7,8]. With forward, backward, and stepwise regression techniques, confounders are defined as those variables with a regression β on the outcome at a level of significance below a pre-specified. Popular choices for the pre-specified level of significance include 0.05, 0.10, and 0.20[1]. With the CIE criterion, confounders are defined as those variables for which the percent difference between the values of the regression β when the variable is adjusted for compare with when it is not adjusted for is larger than a pre-specified value, usually 10%[9]. Use of the CIE criterion with a fixed cutoff level has been found inappropriate[10], and some researchers recently proposed using a data-driven, simulated cutoff that yields a 5% of type I error rate[10]. Theoretical confounder identification strategies select the confounders from the results of previous studies or expert knowledge. This strategy is exemplified by directed acyclic graphs (DAGs)[11].

In confounder identification, some researchers have relied on empirical criteria only when theoretical evidence is not available[12], but other researchers have identified a need for combining empirical and theoretical criteria[9,13]. Simulation studies have shown that the CIE criterion combined with DAGs more accurately estimates exposure-outcome associations than do DAGs alone[14]. However, the decision of whether to adjust for remains unclear if the results from empirical and theoretical criteria are contradictory due to random error. These contradictions can be further divided into two cases, analogous to type I and type II errors. First, a variable without a confounding effect

may show empirical confounding evidence in a particular dataset due to random error. Second, a true confounder may show no empirical confounding evidence. Most researchers have recommended first listing all theoretically possible confounders identified by DAGs and then selecting those that should be adjusted for by empirical methods[13,14]. However, these studies have provided no guidelines for handling confounders that are identified by empirical criteria but excluded in the DAGs; that is, potential confounders without theoretical background. How to handle these variables, and, equally, confounders suggested by DAGs that cannot be empirically validated, is a natural question. This study aims to answer the following research question "Should we adjust for a confounder if empirical and theoretical criteria yield contradictory results?" We simulated several scenarios in which a potential confounder exerted no confounding effect, and other scenarios in which a true confounder exerted a confounding effect. Simulations were performed for different sample sizes and exposure-confounder correlations. By comparing the accuracy of exposure-outcome associations under different scenarios, we could determine when an adjustment for a potential confounder was required.

Results

Table 1 summarizes the results of Simulation 1, in which a true confounding effect was present. In a simulation of size $N=200$, the average probability of accurately detecting a confounding effect under the significance criterion ($p < 0.05$), significance criterion ($p < 0.20$), CIE criterion with a 10% cutoff, and CIE criterion with a simulated cutoff was 9.77%, 29.32%, 12.32%, and 59.14%, respectively. The probability decreased with increasing sample size under the CIE criterion with a 10% cutoff and increased with sample size under the significance criteria and the CIE criterion with a simulated cutoff. The significance criteria and the CIE criterion with a 10% cutoff underestimated the true OR, even when correctly identifying the confounding effect, although the performance of the significance criterion was better using the cutoff p-value of 0.20 compared with 0.05. Confounder adjustment led to underestimates as large as 16.63% ($N=500$, CIE criterion with 10% cutoff, exposure-confounder association=0.2), while no adjustment yielded overestimates as large as 6.64% ($N=200$, CIE criterion with 10% cutoff, exposure-confounder association=0.5). Under the significance criteria, however, these biases reduced as the sample size increased. While the significance criteria and the CIE criterion with a 10% cutoff performed poorly, the CIE criterion with a simulated cutoff provided accurate estimates. Under this criterion, the absolute percentage errors of the adjusted OR were all within 0.98% (the largest absolute percentage error was attained at $N=200$, exposure-confounder association=0.1). At a sample size of 1,000, all absolute percentage errors were within 0.2%. If no confounding effect was identified, the estimates were positively biased under all criteria. Under the significance criteria and the CIE criterion with a 10% cutoff, this bias was reduced by adjusting for the confounder, which reduced the percentage error by 4.28% at most ($N=1,000$, CIE criterion with a 10% cutoff, exposure-confounder association=0.5). To visualize the simulation results, the percentage error in the simulated OR for the case of $N=200$ is shown in Figure 1.

Table 2 shows the results of Simulation 2, in which the potential confounder exerted no causal effect on the outcome. In a simulation of size $N=200$, the average probability of accurately determining a null confounding effect under the significance criterion ($p < 0.05$), significance criterion ($p < 0.20$), CIE criterion with a 10% cutoff, and CIE criterion with a simulated cutoff was 95.24%, 80.02%, 92.15%, and 46.21%, respectively. This probability was independent of sample size under the significance criteria, decreased with increasing sample size under the CIE criterion with a 10% cutoff, and increased with sample size under the CIE criterion with a simulated cutoff. Under all criteria, if no confounding effect was identified, the

accuracy of the ORs was essentially unaltered by adjustment. The absolute differences in the percentage errors were all within 0.43% (the largest absolute difference in the percentage error was attained at $N=200$, CIE criterion with 10% cutoff, exposure-confounder association=0.5). However, if a confounding effect was identified, confounder adjustment increased the percentage error and RMSE. The percentage error in the simulated OR for the case of $N=200$ is shown in Figure 2. Not surprisingly, the accuracy of the estimation improved with increasing sample size.

Table 3 shows the results of Simulations 3 and 4, in which the exposure and the potential confounder were uncorrelated. The average probability of accurately determining a null confounding effect was high under all four criteria (significance criterion ($p < 0.05$), significance criterion ($p < 0.20$), CIE criterion with a 10% cutoff, and CIE criterion with a simulated cutoff). Provided that the potential confounder exerted no causal effect on the outcome (Simulation 3), these probabilities were independent of sample size; otherwise (Simulation 4), the probabilities reduced as sample size increased. As in the the results of Simulation 2, if no confounding effect was identified, the accuracy of the ORs was highly independent of imposed criterion and confounder adjustment. The percentage errors were all within 1.58% (the largest absolute percentage error was attained at $N=200$, significance criterion, no confounder-outcome effect, confounder adjusted). Similarly, if a confounding effect was identified, no difference was found whether the confounder was adjusted for or not. Again, the accuracy of the estimation improved with increasing sample size.

Table 4 shows the results of Simulation 5, in which the potential confounder was affected by both the exposure and the outcome. The probability of correctly determining a null confounding effect was low under all criteria. Despite their poor performance in identifying a non-confounding effect, under all criteria the estimated OR for the exposure-outcome association were of acceptable accuracy. Even with a small sample size ($N=200$), the absolute percentage error differences in the unadjusted ORs were all within 1.43% (the largest absolute difference was attained under all criteria, slope of regression line relating exposure to the potential confounder = 0.4). However, all adjusted ORs were positively biased and the biases increased with the slope of the regression line relating exposure to the potential confounder. The percent error in the simulated OR for the case of $N=200$ is shown in Figure 3. As in Simulations 2 through 4, accuracy improved with increasing sample size.

Discussion

Five different scenarios, based on various empirical and theoretical criteria, were simulated to evaluate the accuracy of exposure-outcome association estimates. Simulations containing a confounding effect and simulations without a confounding effect will be discussed in turn.

In the presence of a confounding effect, the OR estimates were biased under the significance criteria and the CIE criterion with a 10% cutoff. In simulations that indicated a confounding effect, the unadjusted OR was overestimated, but an adjustment for the confounding effect yielded underestimates of the true OR in subsequent logistic regression. These biases increased with increasing exposure-confounder correlation. The absolute percentage errors were largely unaffected by confounder adjustment. However, using the CIE criterion with a simulated cutoff yielded unbiased OR estimates. In simulations that indicated no confounding effect, adjustment only reduced part of the bias (the adjusted OR was less biased than the unadjusted OR). This result is expected under the significance criterion, which is subject to multi-collinearity effects. Under this criterion, adjusting for confounding is beneficial. However, we should note that the exposure-outcome association obtained from the dataset is biased regardless of confounder adjustment. The RMSEs yielded by all criteria were similar, and were reduced with increasing

Table 1 | Percentage error (root mean square error) in the simulated odds ratio (OR) with confounder (simulation size = 10,000)

	N=200					N=500					N=1,000				
	OR_x	$OR_{x,z}$	OR_x'	$OR_{x,z}'$	% indicating confounding effect	OR_x	$OR_{x,z}$	OR_x'	$OR_{x,z}'$	% indicating confounding effect	OR_x	$OR_{x,z}$	OR_x'	$OR_{x,z}'$	% indicating confounding effect
Significance criterion (p < 0.05)															
Corr (X, Z)															
0.1	2.50 (0.16)	−0.46 (0.16)	2.36 (0.17)	1.69 (0.17)	10.72	1.14 (0.10)	−1.01 (0.10)	1.35 (0.10)	0.68 (0.10)	19.51	1.17 (0.07)	−0.44 (0.07)	1.27 (0.07)	0.63 (0.07)	34.67
0.2	3.40 (0.17)	−3.15 (0.17)	3.10 (0.17)	1.61 (0.17)	10.21	2.82 (0.11)	−1.69 (0.10)	2.36 (0.11)	0.98 (0.10)	19.32	2.54 (0.08)	−0.82 (0.07)	2.22 (0.08)	0.93 (0.07)	33.18
0.3	4.44 (0.18)	−5.21 (0.17)	4.28 (0.17)	2.03 (0.17)	9.30	3.45 (0.11)	−3.43 (0.10)	3.50 (0.11)	1.43 (0.11)	18.58	3.22 (0.08)	−1.97 (0.07)	3.30 (0.08)	1.32 (0.08)	32.64
0.4	5.38 (0.18)	−8.10 (0.18)	5.59 (0.18)	2.74 (0.18)	9.46	4.61 (0.12)	−5.00 (0.11)	4.63 (0.12)	1.80 (0.11)	17.69	4.35 (0.09)	−2.85 (0.08)	4.38 (0.09)	1.75 (0.08)	30.18
0.5	6.34 (0.19)	−10.79 (0.20)	6.45 (0.19)	2.82 (0.19)	9.17	5.56 (0.12)	−6.93 (0.11)	5.61 (0.12)	2.09 (0.12)	16.17	5.28 (0.09)	−4.18 (0.08)	5.36 (0.10)	1.97 (0.08)	27.46
		[0.23]					[0.13]					[0.08]			
Significance criterion (p < 0.20)															
Corr (X, Z)															
0.1	2.58 (0.17)	−0.42 (0.17)	2.28 (0.17)	1.89 (0.17)	29.79	1.21 (0.10)	−0.48 (0.10)	1.39 (0.10)	0.99 (0.10)	43.67	1.25 (0.07)	−0.09 (0.07)	1.21 (0.07)	0.84 (0.07)	62.00
0.2	3.34 (0.17)	−1.25 (0.17)	3.03 (0.17)	2.15 (0.17)	30.31	2.69 (0.11)	−0.81 (0.10)	2.25 (0.11)	1.46 (0.10)	43.93	2.33 (0.08)	−0.45 (0.07)	2.32 (0.08)	1.56 (0.07)	60.23
0.3	4.37 (0.17)	−2.59 (0.17)	4.26 (0.18)	3.00 (0.18)	29.46	3.39 (0.11)	−1.99 (0.10)	3.57 (0.11)	2.35 (0.11)	42.07	3.26 (0.08)	−1.02 (0.07)	3.29 (0.08)	2.13 (0.08)	59.74
0.4	5.85 (0.18)	−3.44 (0.18)	5.46 (0.18)	3.80 (0.18)	28.75	4.49 (0.11)	−2.95 (0.11)	4.71 (0.12)	3.10 (0.11)	41.36	4.38 (0.09)	−1.53 (0.08)	4.37 (0.09)	2.82 (0.08)	56.55
0.5	6.50 (0.19)	−5.29 (0.19)	6.42 (0.19)	4.29 (0.19)	28.32	5.54 (0.12)	−4.08 (0.12)	5.64 (0.12)	3.54 (0.12)	38.16	5.19 (0.09)	−2.52 (0.08)	5.50 (0.10)	3.50 (0.09)	53.52
		[0.21]					[0.12]					[0.08]			
Change-in-estimate criterion (10%)															
Corr (X, Z)															
0.1	3.48 (0.11)	−5.43 (0.14)	2.37 (0.17)	1.47 (0.17)	0.15	N/A	N/A	1.31 (0.10)	0.35 (0.10)	0	N/A	N/A	1.24 (0.07)	0.26 (0.07)	0
0.2	0.44 (0.15)	−10.23 (0.18)	3.18 (0.17)	1.35 (0.17)	1.97	−8.01 (0.10)	−16.63 (0.19)	2.45 (0.11)	0.47 (0.10)	0.04	N/A	N/A	2.33 (0.08)	0.26 (0.07)	0
0.3	3.56 (0.18)	−7.08 (0.19)	4.36 (0.18)	2.09 (0.17)	8.02	2.38 (0.11)	−7.91 (0.13)	3.51 (0.11)	0.65 (0.11)	1.46	−2.98 (0.09)	−7.07 (0.11)	3.28 (0.08)	0.26 (0.07)	0.10
0.4	5.88 (0.19)	−5.21 (0.19)	5.50 (0.18)	3.39 (0.18)	19.45	4.54 (0.12)	−6.83 (0.12)	4.63 (0.12)	1.28 (0.11)	8.43	4.65 (0.10)	−5.89 (0.10)	4.37 (0.09)	0.53 (0.08)	2.58
0.5	6.64 (0.19)	−4.86 (0.21)	6.35 (0.19)	4.60 (0.19)	32.03	5.68 (0.12)	−6.54 (0.13)	5.58 (0.12)	2.32 (0.12)	19.04	5.04 (0.09)	−6.30 (0.10)	5.37 (0.10)	1.09 (0.08)	10.99

Table 1 | Continued

Change-in-estimate criterion (simulated cutoff)	N=200				% indicating confounding effect/simulated cutoff	N=500				% indicating confounding effect/simulated cutoff	N=1,000				% indicating confounding effect/simulated cutoff
Corr (X, Z)															
0.1	1.88 (0.16)	2.51 (0.17)	2.16 (0.17)	−0.98 (0.16)	22.46/2.24%	1.18 (0.10)	1.43 (0.10)	1.16 (0.10)	−0.52 (0.10)	48.39/0.90%	1.25 (0.07)	1.18 (0.07)	1.01 (0.07)	0.06 (0.07)	78.55/0.44%
0.2	3.41 (0.17)	2.85 (0.17)	2.42 (0.17)	−0.21 (0.17)	49.44/2.29%	2.56 (0.11)	2.09 (0.10)	1.90 (0.10)	0.02 (0.10)	76.56/0.92%	2.35 (0.08)	2.07 (0.08)	1.99 (0.08)	0.19 (0.07)	91.14/0.45%
0.3	4.38 (0.18)	4.13 (0.17)	3.88 (0.17)	0.13 (0.18)	67.45/2.20%	3.52 (0.11)	3.34 (0.11)	3.22 (0.11)	0.01 (0.11)	83.76/0.98%	3.30 (0.08)	2.82 (0.08)	2.78 (0.08)	0.11 (0.07)	94.58/0.44%
0.4	5.65 (0.18)	5.35 (0.18)	5.11 (0.18)	0.53 (0.19)	74.01/2.51%	4.57 (0.12)	5.02 (0.12)	4.96 (0.12)	0.03 (0.11)	88.44/0.93%	4.40 (0.09)	3.80 (0.08)	3.76 (0.08)	0.20 (0.08)	95.53/0.44%
0.5	6.72 (0.19)	5.17 (0.18)	5.10 (0.18)	0.82 (0.20)	82.34/2.25%	5.53 (0.12)	6.28 (0.13)	6.23 (0.13)	0.11 (0.12)	91.36/0.91%	5.32 (0.09)	5.68 (0.10)	5.66 (0.10)	0.06 (0.08)	96.08/0.44%

Data are simulated from the model $Prob(Y=1)/Prob(Y=0) = \exp(0.1)X + \exp(0.1)Z$.
Percentage error and root mean square error in correct decision are underlined.
OR_X: Percentage of OR of X where Z was not adjusted and the criterion has indicated a confounding effect.
$OR_{X,Z}$: Percentage of OR of X where Z was adjusted and the criterion has indicated a confounding effect.
OR_X': Percentage of OR of X where Z was not adjusted and the criterion has not indicated a confounding effect.
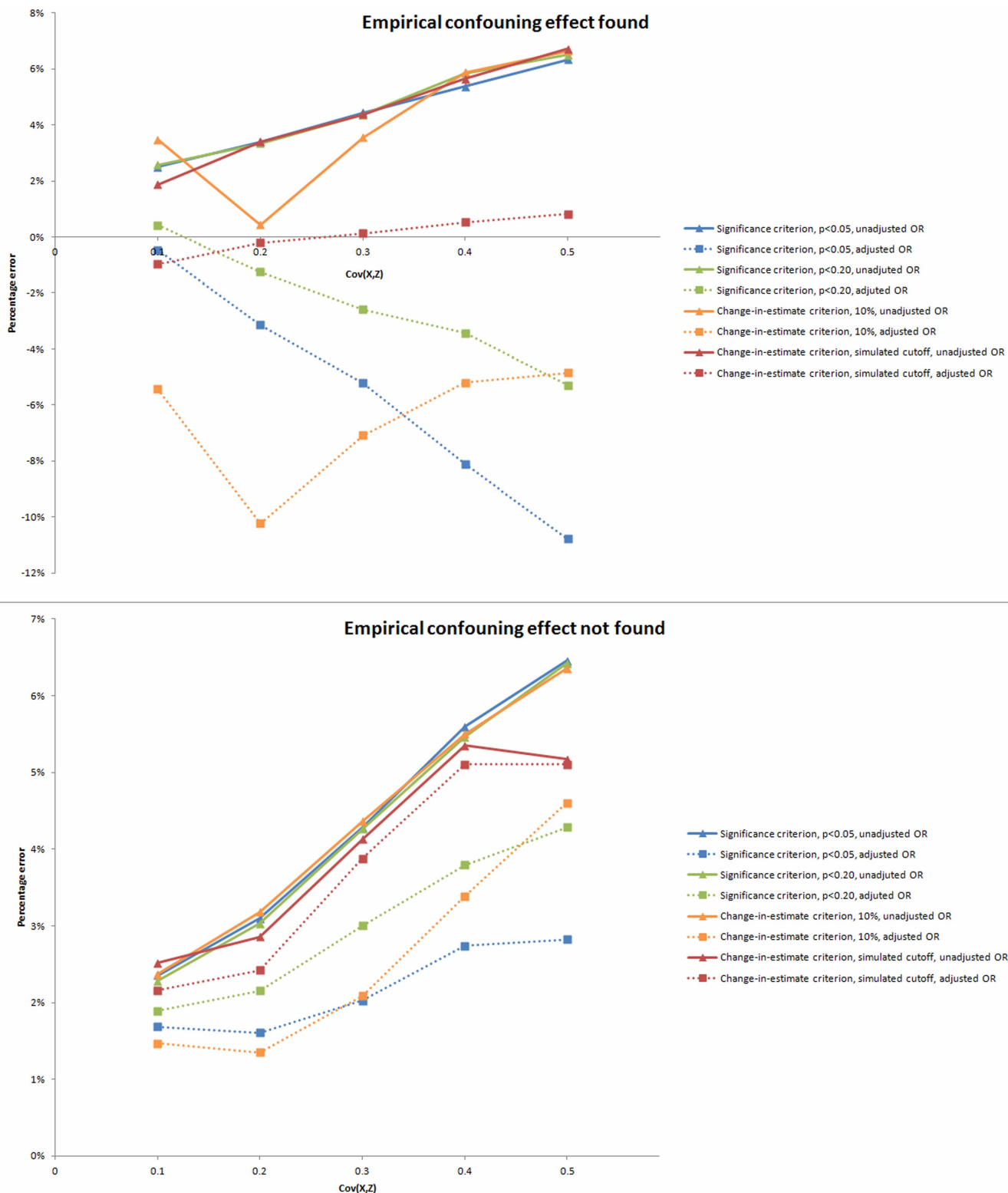$OR_{X,Z}'$: Percentage of OR of X where Z was adjusted and the criterion has not indicated a confounding effect.

Figure 1 | Percentage error in the simulated odds ratio (OR) with confounder (simulation size = 10,000, sample size = 200).

sample size, as expected. Our simulation results agreed with previous simulation studies showing that a significance criterion using a p-value cutoff of 0.20 yielded better outcomes than that using a cutoff of 0.05[1].

To summarize the results of Simulation 1, theoretical confounders should be adjusted for regardless of the empirical evidence found in the dataset. This is consistent with previous simulation results, in which biases of estimates of exposure-outcome associations with uncontrolled confounders were reported[15,16]. Our simulation results showed that if a well-established confounding effect is not observed in a dataset, the results should be interpreted with caution, because the exposure-outcome association obtained from the dataset is likely to be biased.

In the absence of any confounding effect, all estimates were insensitive to the confounder identification criterion, to the adjustment for potential confounding, and to the exposure-outcome association,

Table 2 | Percentage error (root mean square error) in the simulated odds ratio (OR) without confounding effect (simulation size = 10,000)

Significance criterion (p < 0.05)

Corr (X, Z)	N=200					N=500					N=1,000				
	OR_x	$OR_{x,z}$	OR_x'	$OR_{x,z}'$	% indicating confounding effect	OR_x	$OR_{x,z}$	OR_x'	$OR_{x,z}'$	% indicating confounding effect	OR_x	$OR_{x,z}$	OR_x'	$OR_{x,z}'$	% indicating confounding effect
0.1	0.08 (0.16)	0.80 (0.18)	1.24 (0.17)	1.28 (0.17)	4.73	-0.09 (0.10)	-0.11 (0.10)	0.42 (0.10)	0.42 (0.10)	4.95	0.27 (0.07)	0.36 (0.07)	0.28 (0.07)	0.29 (0.07)	4.81
0.2	1.28 (0.15)	2.05 (0.18)	1.46 (0.17)	1.60 (0.17)	4.89	0.04 (0.11)	0.25 (0.12)	0.44 (0.10)	0.47 (0.10)	4.58	0.29 (0.07)	0.43 (0.08)	0.24 (0.07)	0.27 (0.07)	4.64
0.3	2.20 (0.17)	2.39 (0.22)	1.49 (0.17)	1.63 (0.17)	4.34	0.81 (0.10)	1.06 (0.12)	0.56 (0.10)	0.58 (0.10)	4.92	0.15 (0.07)	0.46 (0.09)	0.25 (0.07)	0.26 (0.07)	4.69
0.4	1.42 (0.17)	3.41 (0.25)	1.47	1.69 (0.18)	4.93	0.53 (0.10)	1.65 (0.14)	0.52 (0.10)	0.63 (0.11)	4.63	-0.33 (0.07)	0.02 (0.10)	0.23 (0.07)	0.29 (0.08)	4.85
0.5	2.68 (0.16)	4.93 (0.30)	1.23 (0.17)	1.69 (0.19)	4.88	0.50 (0.10)	1.65 (0.18)	0.55 (0.10)	0.65 (0.11)	4.53	-0.03 (0.07)	0.30 (0.12)	0.29 (0.07)	0.33 (0.08)	5.38

Significance criterion (p < 0.20)

Corr (X, Z)	N=200					N=500					N=1,000				
	OR_x	$OR_{x,z}$	OR_x'	$OR_{x,z}'$	% indicating confounding effect	OR_x	$OR_{x,z}$	OR_x'	$OR_{x,z}'$	% indicating confounding effect	OR_x	$OR_{x,z}$	OR_x'	$OR_{x,z}'$	% indicating confounding effect
0.1	0.83 (0.16)	1.11 (0.17)	1.27 (0.17)	1.29 (0.17)	19.58	0.47 (0.10)	0.45 (0.11)	0.38 (0.10)	0.38 (0.10)	20.29	0.35 (0.07)	0.39 (0.07)	0.27 (0.07)	0.27 (0.07)	20.09
0.2	1.82 (0.17)	2.33 (0.18)	1.35 (0.17)	1.44 (0.17)	20.18	0.55 (0.10)	0.76 (0.11)	0.39 (0.10)	0.38 (0.10)	19.54	0.35 (0.07)	0.48 (0.08)	0.21 (0.07)	0.23 (0.07)	20.18
0.3	1.71 (0.17)	2.15 (0.18)	1.48 (0.17)	1.54 (0.17)	20.07	0.59 (0.10)	0.82 (0.11)	0.57 (0.10)	0.55 (0.10)	20.03	0.35 (0.07)	0.48 (0.08)	0.21 (0.07)	0.21 (0.07)	20.75
0.4	1.73 (0.17)	2.46 (0.20)	1.40 (0.16)	1.60 (0.17)	20.04	0.38 (0.10)	0.88 (0.13)	0.56 (0.10)	0.63 (0.11)	19.51	-0.08 (0.07)	0.16 (0.09)	0.27 (0.07)	0.30 (0.07)	20.13
0.5	1.24 (0.16)	2.95 (0.25)	1.32 (0.17)	1.57 (0.18)	20.05	0.32 (0.10)	0.74 (0.15)	0.60 (0.10)	0.68 (0.11)	19.65	0.35 (0.07)	0.57 (0.10)	0.25 (0.07)	0.27 (0.07)	20.95

Change-in-estimate criterion (10%)

Corr (X, Z)	N=200					N=500					N=1,000				
	OR_x	$OR_{x,z}$	OR_x'	$OR_{x,z}'$	% indicating confounding effect	OR_x	$OR_{x,z}$	OR_x'	$OR_{x,z}'$	% indicating confounding effect	OR_x	$OR_{x,z}$	OR_x'	$OR_{x,z}'$	% indicating confounding effect
0.1	14.68 (0.17)	24.44 (0.31)	1.18 (0.17)	1.24 (0.17)	0.05	N/A	N/A	0.40 (0.10)	0.40 (0.10)	0	N/A	N/A	0.28 (0.07)	0.29 (0.07)	0
0.2	0.72 (0.17)	-0.11 (0.23)	1.45 (0.17)	1.64 (0.17)	0.68	N/A	N/A	0.42 (0.10)	0.46 (0.10)	0	N/A	N/A	0.24 (0.07)	0.28 (0.07)	0
0.3	1.68 (0.17)	0.68 (0.23)	1.52 (0.17)	1.70 (0.17)	3.89	3.76 (0.11)	-3.12 (0.13)	0.57 (0.10)	0.61 (0.11)	0.13	N/A	N/A	0.24 (0.07)	0.27 (0.07)	0
0.4	1.71 (0.17)	2.06 (0.24)	1.44 (0.16)	1.73 (0.17)	11.76	0.05 (0.09)	-0.33 (0.16)	0.53 (0.10)	0.69 (0.11)	1.41	-2.41 (0.06)	-11.81 (0.14)	0.20 (0.07)	0.28 (0.08)	0.04
0.5	1.14 (0.17)	2.05 (0.25)	1.35 (0.17)	1.78 (0.18)	22.85	0.48 (0.10)	-0.31 (0.18)	0.55 (0.10)	0.75 (0.11)	5.48	1.36 (0.07)	1.64 (0.15)	0.26 (0.07)	0.32 (0.08)	0.85

Table 2 | Continued

Change-in-estimate criterion (simulated)	N=200				% indicating confounding effect/simulated cutoff	N=500				% indicating confounding effect/simulated cutoff	N=1,000				% indicating confounding effect/simulated cutoff
Corr (X, Z)															
0.1	0.73 (0.17)	1.07 (0.18)	1.27 (0.16)	1.29 (0.17)	16.29/2.28%	0.49 (0.10)	0.51 (0.10)	0.36 (0.10)	0.35 (0.10)	28.90/0.93%	0.37 (0.07)	0.40 (0.07)	0.22 (0.07)	0.21 (0.07)	43.07/0.47%
0.2	1.73 (0.17)	2.12 (0.18)	1.24 (0.17)	1.26 (0.17)	42.04/2.22%	0.46 (0.10)	0.53 (0.11)	0.34 (0.10)	0.34 (0.10)	62.70/0.85%	0.31 (0.07)	0.37 (0.07)	0.05 (0.07)	0.04 (0.07)	71.45/0.47%
0.3	1.74 (0.17)	1.95 (0.18)	1.22 (0.17)	1.26 (0.16)	58.47/2.34%	0.63 (0.10)	0.68 (0.11)	0.38 (0.10)	0.37 (0.10)	76.11/0.84%	0.24 (0.07)	0.27 (0.08)	0.26 (0.07)	0.26 (0.07)	83.26/0.43%
0.4	1.59 (0.16)	2.01 (0.19)	1.15 (0.16)	1.14 (0.16)	72.33/2.13%	0.46 (0.10)	0.65 (0.11)	0.78 (0.10)	0.79 (0.10)	81.09/0.91%	0.19 (0.07)	0.27 (0.08)	0.28 (0.07)	0.28 (0.07)	87.39/0.45%
0.5	1.37 (0.17)	2.04 (0.20)	1.03 (0.17)	1.07 (0.17)	79.82/2.13%	0.54 (0.10)	0.71 (0.12)	0.60 (0.10)	0.58 (0.10)	85.72/0.90%	0.29 (0.07)	0.36 (0.08)	0.06 (0.07)	0.05 (0.07)	90.77/0.44%

Data are simulated from the model Prob(Y=1)/Prob(Y=0) = exp(0.1)X.
Percentage error and root mean square error in correct decision are underlined.
OR_x: Percentage of OR of X where Z was not adjusted and the criterion has indicated a confounding effect.
OR_{xz}: Percentage of OR of X where Z was adjusted and the criterion has indicated a confounding effect.
OR_x': Percentage of OR of X where Z was not adjusted and the criterion has not indicated a confounding effect.
OR_{xz}': Percentage of OR of X where Z was adjusted and the criterion has not indicated a confounding effect.
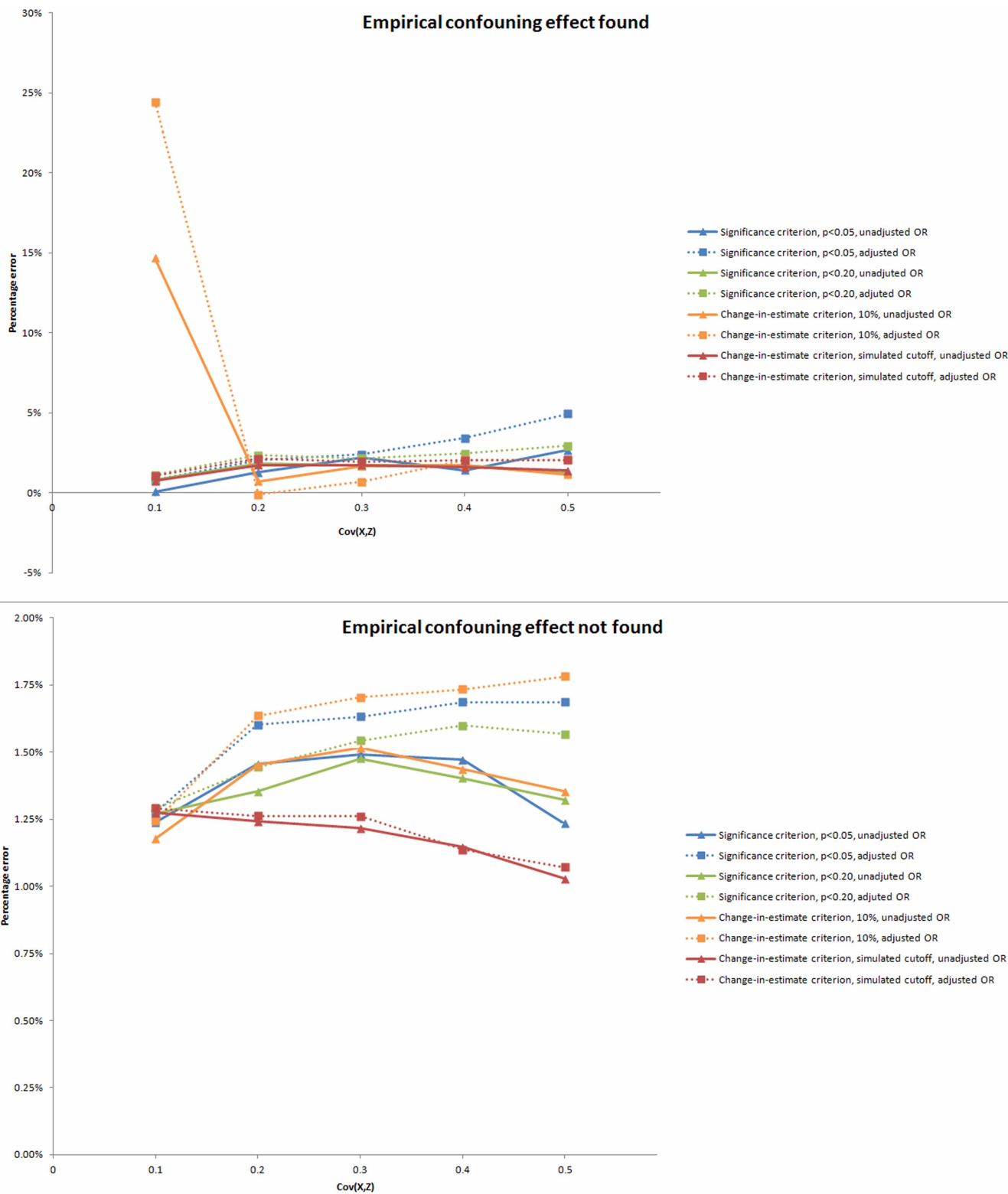
Figure 2 | **Percentage error in the simulated odds ratio (OR) without confounding effect (simulation size = 10,000, sample size = 200).**

although the biases increased with increasing exposure-confounder correlation. However, the RMSE increased when the confounding effect was identified and adjusted for. Under all criteria, the accuracy improved with increasing sample size. Under the significance criteria and the CIE criterion with a 10% cutoff, adjusting for the confounding effect induced marked increases in the RMSE. Under the significance criterion, the likelihood of adjustment was independent of sample size and exposure-confounder correlation, although these

parameters affected the adjustment likelihood under the CIE criterion.

The results of Simulations 2, 3, and 4, in summary, suggested that the adjustment for factors that do not exert a confounding effect has a minimal impact on the accuracy of exposure-outcome estimates. Furthermore, all confounder identification criteria perform equally well in terms of percentage error and RMSE, except for the CIE with a 10% cutoff, which may yield a large percentage error in the

Table 3 | Percentage error (root mean square error) in the simulated odds ratio (OR) without confounding effect (simulation size = 10,000)

	N=200					N=500					N=1,000				
	OR$_x$	OR$_{x,z}$	OR$_x$'	OR$_{x,z}$'	% indicating confounding effect	OR$_x$	OR$_{x,z}$	OR$_x$'	OR$_{x,z}$'	% indicating confounding effect	OR$_x$	OR$_{x,z}$	OR$_x$'	OR$_{x,z}$'	% indicating confounding effect
Significance criterion (p < 0.05)															
Pr(Y=1)/Pr(Y=0) = exp[0.1]|X	2.13 (0.17)	2.48 (0.18)	1.47 (0.17)	1.53 (0.17)	4.91	0.56 (0.11)	0.74 (0.11)	0.56 (0.10)	0.58 (0.10)	4.76	0.18 (0.07)	0.25 (0.07)	0.37 (0.07)	0.38 (0.07)	4.78
exp[0.1]|X + exp[0.1]|Z	1.16 (0.16)	1.61 (0.17)	1.12 (0.16)	1.17 (0.16)	10.36	0.64 (0.10)	0.77 (0.10)	0.57 (0.10)	0.59 (0.11)	19.29	0.11 (0.07)	0.18 (0.07)	0.22 (0.07)	0.24 (0.07)	65.01
Significance criterion (p < 0.20)															
Pr(Y=1)/Pr(Y=0) = exp[0.1]|X	1.63 (0.17)	1.87 (0.17)	1.47 (0.17)	1.50 (0.17)	20.81	0.47 (0.10)	0.57 (0.10)	0.58 (0.10)	0.59 (0.10)	19.57	0.23 (0.07)	0.27 (0.07)	0.39 (0.07)	0.40 (0.07)	19.70
exp[0.1]|X + exp[0.1]|Z	1.49 (0.16)	1.71 (0.17)	0.96 (0.16)	0.99 (0.16)	30.52	0.56 (0.10)	0.65 (0.10)	0.61 (0.10)	0.61 (0.11)	44.94	0.14 (0.07)	0.20 (0.07)	0.25 (0.07)	0.25 (0.07)	61.58
Change-in-estimate criterion (10%)															
Pr(Y=1)/Pr(Y=0) = exp[0.1]|X	N/A	N/A	1.50 (0.17)	1.58 (0.17)	0	N/A	N/A	0.56 (0.10)	0.59 (0.10)	0	N/A	N/A	0.36 (0.07)	0.37 (0.07)	0
exp[0.1]|X + exp[0.1]|Z	N/A	N/A	1.13 (0.16)	1.21 (0.16)	0	N/A	N/A	0.58 (0.10)	0.63 (0.10)	0	N/A	N/A	0.18 (0.07)	0.22 (0.07)	0
Change-in-estimate (simulated) — % indicating confounding effect/simulated cutoff															
Pr(Y=1)/Pr(Y=0) = exp[0.1]|X	0.87 (0.17)	1.64 (0.18)	1.54 (0.17)	1.58 (0.17)	5.34/2.24%	0.95 (0.11)	1.24 (0.11)	0.54 (0.10)	0.55 (0.10)	5.24/0.87%	0.19 (0.07)	0.33 (0.07)	0.37 (0.07)	0.37 (0.07)	4.86/0.44%
exp[0.1]|X + exp[0.1]|Z	1.16 (0.17)	1.81 (0.18)	1.12 (0.16)	1.17 (0.16)	7.08/2.37%	0.54 (0.10)	0.74 (0.10)	0.59 (0.10)	0.61 (0.10)	13.65/0.85%	0.38 (0.07)	0.53 (0.07)	0.13 (0.07)	0.14 (0.07)	20.15/0.44%

Data are simulated such that Corr (X, Z)=0.
Percentage error and root mean square error in correct decision are underlined.
OR$_x$: Percentage of OR of X where Z was not adjusted and the criterion has indicated a confounding effect.
OR$_{x,z}$: Percentage of OR of X where Z was adjusted and the criterion has indicated a confounding effect.
OR$_x$': Percentage of OR of X where Z was not adjusted and the criterion has not indicated a confounding effect.
OR$_{x,z}$': Percentage of OR of X where Z was adjusted and the criterion has not indicated a confounding effect.
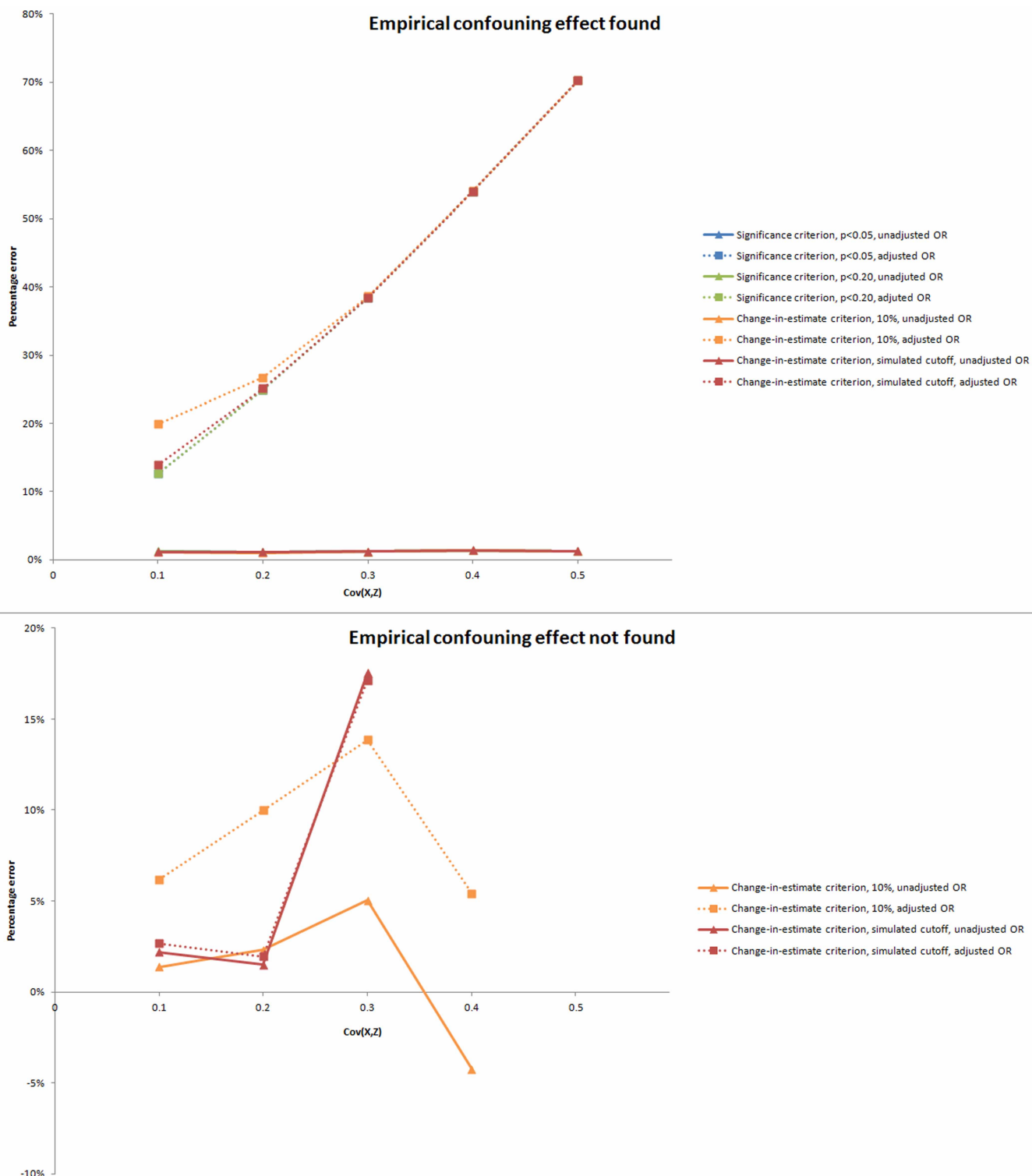
Table 4 | Percentage error (root mean square error) in the simulated odds ratio (OR) without confounding effect (simulation size = 10,000)

	N=200					N=500					N=1,000				
	OR_x	$OR_{x,z}$	OR_x'	$OR_{x,z}'$	% indicating confounding effect	OR_x	$OR_{x,z}$	OR_x'	$OR_{x,z}'$	% indicating confounding effect	OR_x	$OR_{x,z}$	OR_x'	$OR_{x,z}'$	% indicating confounding effect
Significance criterion ($p < 0.05$)															
Z=															
$0.1X+Y+\varepsilon$	1.28 (0.17)	12.61 (0.25)	N/A	N/A	100.00	0.50 (0.10)	11.32 (0.18)	N/A	N/A	100.00	0.10 (0.07)	10.75 (0.15)	N/A	N/A	100.00
$0.2X+Y+\varepsilon$	1.18 (0.16)	25.10 (0.37)	N/A	N/A	100.00	0.33 (0.10)	23.01 (0.29)	N/A	N/A	100.00	0.15 (0.07)	22.62 (0.27)	N/A	N/A	100.00
$0.3X+Y+\varepsilon$	1.23 (0.17)	38.45 (0.51)	N/A	N/A	100.00	0.42 (0.10)	36.31 (0.43)	N/A	N/A	100.00	0.25 (0.07)	35.65 (0.41)	N/A	N/A	100.00
$0.4X+Y+\varepsilon$	1.43 (0.17)	54.03 (0.68)	N/A	N/A	100.00	0.52 (0.10)	50.85 (0.59)	N/A	N/A	100.00	0.25 (0.07)	50.04 (0.57)	N/A	N/A	100.00
$0.5X+Y+\varepsilon$	1.26 (0.17)	70.25 (0.86)	N/A	N/A	100.00	0.45 (0.10)	66.97 (0.77)	N/A	N/A	100.00	0.18 (0.07)	65.76 (0.74)	N/A	N/A	100.00
Significance criterion ($p < 0.20$)															
Z=															
$0.1X+Y+\varepsilon$	1.28 (0.17)	12.61 (0.25)	N/A	N/A	100.00	0.50 (0.10)	11.32 (0.18)	N/A	N/A	100.00	0.10 (0.07)	10.75 (0.15)	N/A	N/A	100.00
$0.2X+Y+\varepsilon$	1.18 (0.16)	25.10 (0.37)	N/A	N/A	100.00	0.33 (0.10)	23.01 (0.29)	N/A	N/A	100.00	0.15 (0.07)	22.62 (0.27)	N/A	N/A	100.00
$0.3X+Y+\varepsilon$	1.23 (0.17)	38.45 (0.51)	N/A	N/A	100.00	0.42 (0.10)	36.31 (0.43)	N/A	N/A	100.00	0.25 (0.07)	35.65 (0.41)	N/A	N/A	100.00
$0.4X+Y+\varepsilon$	1.43 (0.17)	54.03 (0.68)	N/A	N/A	100.00	0.52 (0.10)	50.85 (0.59)	N/A	N/A	100.00	0.25 (0.07)	50.04 (0.57)	N/A	N/A	100.00
$0.5X+Y+\varepsilon$	1.26 (0.17)	70.25 (0.86)	N/A	N/A	100.00	0.45 (0.10)	66.97 (0.77)	N/A	N/A	100.00	0.18 (0.07)	65.76 (0.74)	N/A	N/A	100.00
Change-in-estimate criterion (10%)															
Z=															
$0.1X+Y+\varepsilon$	1.17 (0.17)	19.88 (0.31)	1.37 (0.17)	6.18 (0.19)	46.89	0.47 (0.10)	16.02 (0.22)	0.52 (0.10)	7.48 (0.14)	44.96	0.04 (0.07)	14.02 (0.18)	0.14 (0.07)	8.19 (0.12)	43.90
$0.2X+Y+\varepsilon$	1.04 (0.16)	26.73 (0.38)	2.31 (0.17)	9.99 (0.22)	89.38	0.31 (0.10)	23.35 (0.29)	1.03 (0.10)	10.49 (0.16)	97.67	0.15 (0.07)	22.65 (0.27)	0.53 (0.07)	10.33 (0.14)	99.72
$0.3X+Y+\varepsilon$	1.21 (0.17)	38.62 (0.51)	5.03 (0.16)	13.85 (0.22)	99.33	0.42 (0.10)	36.31 (0.29)	N/A	N/A	100.00	0.25 (0.07)	35.65 (0.27)	N/A	N/A	100.00
$0.4X+Y+\varepsilon$	1.43 (0.17)	54.04 (0.68)	−4.24 (0.20)	5.41 (0.22)	99.98	0.52 (0.10)	50.85 (0.59)	N/A	N/A	100.00	0.25 (0.07)	50.04 (0.57)	N/A	N/A	100.00
$0.5X+Y+\varepsilon$	1.26 (0.17)	70.25 (0.86)	N/A	N/A	100.00	0.45 (0.10)	66.97 (0.77)	N/A	N/A	100.00	0.18 (0.07)	65.76 (0.74)	N/A	N/A	100.00

Table 4 | Continued

Change-in-estimate criterion (simulated)	N=200 OR_x	N=200 $OR_{x,z}$	N=200 OR_x'	N=200 $OR_{x,z}'$	% indicating confounding effect/simulated cutoff	N=500 OR_x	N=500 $OR_{x,z}$	N=500 OR_x'	N=500 $OR_{x,z}'$	% indicating confounding effect/simulated cutoff	N=1,000 OR_x	N=1,000 $OR_{x,z}$	N=1,000 OR_x'	N=1,000 $OR_{x,z}'$	% indicating confounding effect/simulated cutoff
Z=															
0.1X+Y+ε	1.16 [0.17]	13.90 (0.26)	2.19 (0.18)	2.66 (0.18)	88.53/(2.53%)	0.52 [0.10]	11.48 (0.18)	−0.83 (0.10)	−0.75 (0.10)	98.68/0.83%	0.09 [0.07]	10.76 (0.15)	1.35 (0.07)	1.42 (0.07)	99.84/0.45%
0.2X+Y+ε	1.18 [0.16]	25.10 (0.37)	1.51 (0.18)	1.96 (0.18)	99.35/(2.17%)	0.33 [0.10]	23.01 (0.29)	N/A	N/A	100.00/0.88%	0.15 [0.07]	22.62 (0.27)	N/A	N/A	100.00/0.44%
0.3X+Y+ε	1.23 [0.17]	38.46 (0.51)	17.55 (0.22)	17.13 (0.21)	99.98/(2.33%)	0.42 [0.10]	36.31 (0.43)	N/A	N/A	100.00/0.85%	0.25 [0.07]	35.65 (0.41)	N/A	N/A	100.00/0.44%
0.4X+Y+ε	1.43 [0.17]	54.04 (0.68)	N/A	N/A	100.00/(2.38%)	0.52 [0.10]	50.85 (0.59)	N/A	N/A	100.00/0.90%	0.25 [0.07]	50.04 (0.57)	N/A	N/A	100.00/0.46%
0.5X+Y+ε	1.26 [0.17]	70.25 (0.86)	N/A	N/A	100.00/(2.38%)	0.45 [0.10]	66.97 (0.77)	N/A	N/A	100.00/0.88%	0.18 [0.07]	65.76 (0.74)	N/A	N/A	100.00/0.41%

Data are simulated from the model Prob(Y=1)/Prob(Y=0) = $\exp(0.1)X$.

ε denotes a standard normal error.

Percentage error and root mean square error in correct decision are underlined.

OR_x: Percentage of OR of X where Z was not adjusted and the criterion has indicated a confounding effect.

$OR_{x,z}$: Percentage of OR of X where Z was adjusted and the criterion has indicated a confounding effect.

OR_x': Percentage of OR of X where Z was not adjusted and the criterion has not indicated a confounding effect.

$OR_{x,z}'$: Percentage of OR of X where Z was adjusted and the criterion has not indicated a confounding effect.

Figure 3 | **Percentage error in the simulated odds ratio (OR) without confounding effect (simulation size = 10,000, sample size = 200).**

estimation if a confounding effect is mistakenly identified. This minimal impact of adjusting for a factor without no confounding effect is unexpected; according to some researchers, variables that cause neither the exposure nor the outcome should not be adjusted for[9,11], because empirical strategies for confounder identification cannot prove the existence of any casual effect between the confounder and the exposure or the outcome.

If a potential confounder is affected by both exposure and outcome, it is erroneously identified as a true confounder by all empirical identification criteria. This is expected because the potential confounder is associated with both exposure and outcome, which renders it indistinguishable from a true confounder if data-driven identification strategies are used. Therefore, to summarize the results of Simulation 5, we should not adjust for a potential confounder that is affected by both exposure and outcome. Instead, we should implement *a priori* confounder selection by DAGs[9,11].

The simulation results showed us that adjusting for a non-confounder has a minimally adverse impact in the estimation of
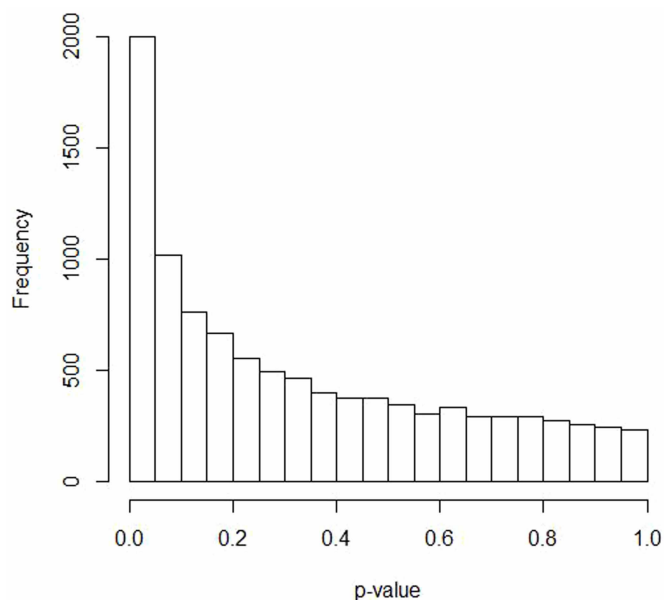
Figure 4 | **Histogram of the p-values of the odds ratios fitted from the generated data (n=10,000).**

the exposure-outcome association. Therefore, we should adjust for a potential confounder even if there is little or no theoretical background knowledge of its confounding effect. However, cautions must be taken in adjusting for variables that are associated with both exposure and outcome because adjusting for such variables will lead to a biased estimation of the exposure-outcome association.

Previous simulation studies have shown that while both significance criteria and CIE criteria can be implemented for confounder identification, CIE criteria works better than significance criteria[8,17]. Similar conclusions can be drawn in this study. Comparing the percentage errors and RMSEs across different simulations shows that the CIE criterion with a simulated cutoff clearly outperformed the other three criteria in detecting a confounding effect while yielding the highest accuracy in estimating the true association between exposure and outcome. The CIE criterion with a fixed cutoff of 10% showed diminishing power for determining confounding effects as the sample size increased; at a sample size of 1,000 and an exposure-confounder correlation of <0.2, the power dropped to almost zero. However, the CIE criterion with a simulated cutoff achieved a power of 80% in most of the scenarios. The cutoff point of 10% appeared to be too high, as the simulations yielded cutoff points of 0.4% to 2.5% for an OR of 1.1, and cutoff points of 0.9% to 5.4% for an OR of 1.6.

This study was not without limitations. One such limitation was that we implemented adjustment by including only a linear term in the logistic regression. Other forms of confounding, for instance confounding by a categorical variable, could have different effects and such effects should be simulated in future studies. Because adjustment for the confounder with a linear effect model yielded satisfactory estimates of the true exposure-outcome association in prior simulation studies[18], we did not consider other forms of confounding here. A second limitation arises from the fact that evidence-based medical studies often involve ordinal, continuous, and survival outcomes. Although only the stimulation results of binary outcomes were studied here, further simulation studies using other types of outcomes and different exposure-outcome effects are warranted, because the confounder adjustment properties differ among regression models[10,19]. Additionally, the conclusions in this study were drawn based on a limited range of parameters, and it remains unknown what the results will be for parameters out of the range this study tested. Such simulations could be performed by modifying the R code provided in Additional file 1. A third limitation is that

several less common empirical confounder identification strategies, including Akaike information criterion and Bayesian information criterion[1], have not been tested in this simulation study. Again, interested readers could test these strategies with a slight modification of the provided R code. A final limitation is that the causal effect of an outcome, especially a health-related outcome, often involves the interaction of several causal exposures and is seldom due to a single exposure[20]. The scenarios simulated in this study were simplifications of real situations, and the conclusions should not be generalized to multiple-exposure effects.

Methods

Confounder adjustment: theory. Here, we denote exposure, outcome and the possible confounder by X, Y, and Z, respectively. Z is a confounder if it satisfies the following conditions[2]: (a) X is associated with Z, (b) Y is associated with Z, and (c) Z is not caused by both X and Y.

In what follows, we will assume the effects of Z and X on Y are linear and compare the estimated effect of X on Y using linear regression with and without adjusting for Z. Note that the relationships between X, Y, and Z can be stated mathematically as $Y = \beta_{X,Y}X + \beta_{Y,Z}Z + \varepsilon$, where ε is the error term. If a linear regression is fitted on Y with only X as the independent variable, that is, $Y = \beta_{X,Y}'X + \varepsilon'$, then the maximum likelihood estimator (MLE) of $\beta_{X,Y}'$, $\beta_{X,Y}'^*$, equals

$$\beta_{X,Y}'^* = \mathrm{Cov}(X,Y)/\mathrm{Var}(X) = \mathrm{Cov}(X, \beta_{X,Y}X + \beta_{Y,Z}Z + \varepsilon)/\mathrm{Var}(X)$$
$$= [\beta_{X,Y}\mathrm{Var}(X) + \beta_{Y,Z}\mathrm{Cov}(X,Z) + \mathrm{Cov}(X,\varepsilon)]/\mathrm{Var}(X)$$
$$= \beta_{X,Y} + \beta_{Y,Z}\mathrm{Cov}(X,Z)/\mathrm{Var}(X),$$

assuming that X and ε are uncorrelated. Hence, $\beta_{X,Y}'^*$ is an unbiased estimator of $\beta_{X,Y}$ if and only if $\mathrm{Cov}(X, Z) = 0$ or $\beta_{Y,Z} = 0$ (making the second term in the right hand side of the equation zero). Thus, $\beta_{X,Y}'^*$ is an unbiased estimator of $\beta_{X,Y}$ if and only if either condition (a) or (b) is not satisfied, or in other words, Z is not a confounder of the association between X and Y. This result shows that a true confounder should be adjusted for, otherwise the estimation of exposure-outcome association will be biased. This result also shows that adjusting for a variable with no association with either the exposure or the outcome has no effect on the bias of the exposure-outcome association estimate. Of course, the standard error of $\beta_{X,Y}'^*$ may be higher than that of $\beta_{X,Y}^*$, which is the MLE of $\beta_{X,Y}$, depending on the correlation between X and Z. It is clear that a potential confounder should not be adjusted for based only on its association with the outcome because if $\mathrm{Cov}(X,Z) = 0$ the unadjusted MLE $\beta_{X,Y}'^*$ is an unbiased estimator of $\beta_{X,Y}$.

The CIE criterion with a 10% cutoff to determine whether a potential confounder should be adjusted for can be restated as $\beta_{Y,Z}\mathrm{Cov}(X, Z)/\beta_{X,Y}\mathrm{Var}(X) > 10\%$. If there exists a confounding effect but the estimate of $\beta_{Y,Z}\mathrm{Cov}(X, Z)/\beta_{X,Y}\mathrm{Var}(X)$ is less than 10%, the bias could be due to the underestimation of $\beta_{Y,Z}$ and/or $\mathrm{Cov}(X, Z)$ or to the overestimation of $\mathrm{Var}(X)$, and we cannot determine which. Hence, we cannot determine which estimator, $\beta_{X,Y}'^*$ or $\beta_{X,Y}^*$, is a better estimate of $\beta_{X,Y}$; therefore we use simulations to answer this question.

Furthermore, the aforementioned theory only applies to linear regressions. However, for logistic regression and other types of regression in which that the MLEs have no closed-form solution, the biased of the MLEs due to an unadjusted confounding effect could not be determined using the above proof, therefore we try to answer the above questions using simulations.

Disagreement of empirical and theoretical confounder adjustment criteria due to random error. This study aims to investigate whether to adjust for a potential confounder or not when empirical and theoretical confounder adjustment criteria disagree due to random error. This is not uncommon, as is illustrated with the following simulation. A total of 10,000 datasets were simulated with sample sizes of $N=500$. In these datasets, a random variable Z, which followed a normal distribution with a mean of zero and a variance of one, was generated, and Y was generated according to the equation $\frac{\Pr(Y=1)}{\Pr(Y=0)} = \exp(0.1)Z$. The p-values of the fitted odds ratios (ORs) are plotted as a histogram in Figure 4. It can be deduced that the power of the logistic regression was small. Even under the least stringent significance criterion of $p < 0.20$, the association between Z and Y was identified empirically in less than 50% of the simulated datasets. While it is obvious that we should adjust for Z (according to the significance criterion and/or prior knowledge that Z and Y are correlated) within those datasets where Z and Y were associated, the decision to adjust for Z in those datasets where Z and Y were not associated, is yet to be determined. Previous studies on confounder adjustment did not consider such a problem, and this study aims to provide guidelines for this situation and for other similar situations where empirical and theoretical confounder adjustment criteria yield conflicting results due to random error.

Model simulation. In the following simulations, we assumed no selection bias, that all errors were random and that there were no systematic errors such as measurement errors. We simulated the case for logistic regression, and the accuracy of the estimated OR for the effect of X on Y with and without confounder adjustment was compared in

five simulations with different correlation patterns between X and Z and between Y and Z imposing different objective criteria. In the first four simulations, X and Z were both normally distributed with zero mean and variance of one. X and Z were drawn to satisfy specific Pearson correlations. In Simulation 1, Y was a binary variable generated according to the equation $\frac{\Pr(Y=1)}{\Pr(Y=0)} = \exp(0.1)X + \exp(0.1)Z$. In Simulation 2, Y was generated according to the equation $\frac{\Pr(Y=1)}{\Pr(Y=0)} = \exp(0.1)X$. In both simulations, X and Z were correlated to varying extents (0.1, 0.2, 0.3, 0.4, and 0.5). The data for Simulations 3 and 4 were generated similarly but with no correlation between X and Z. In Simulation 5, X was normally distributed with a mean of zero and a variance of one, Y was generated according to $\frac{\Pr(Y=1)}{\Pr(Y=0)} = \exp(0.1)X$, and Z was specified by $\beta X + Y + \varepsilon$, where ε is a random variable drawn from a normal distribution with a mean and variance of zero and one, respectively. Different levels of β (0.1, 0.2, 0.3, 0.4, and 0.5) were simulated. We noted that Z was a true confounder in Simulation 1, while Simulations 2, 3 and 4 mimicked a logistic regression in which Z was not a confounder. Z was affected by both X and Y in Simulation 5. Simulations were conducted at different sample sizes (200, 500, and 1,000), and 10,000 datasets were simulated in each scenario.

The OR of the above simulations were set to $\exp(0.1) \sim 1.1$. Additionally, both the exposure and the confounder were continuous. An additional set of simulations were performed with ORs = $\exp(0.5) \sim 1.6$, and both the exposure and the confounder were dichotomized with positive values and negative values set to 1 and 0, respectively. The conclusions drawn from the two sets of simulations were similar, therefore only the first set of simulations with ORs = $\exp(0.1)$ were reported here. The results of the additional set of simulations were provided in the Supplementary Information.

Estimation of the OR. Under all five scenarios, the 10,000 simulated datasets were divided into two groups. The first group was composed of those simulated datasets that showed empirical evidence of a need for confounder adjustment. This evidence was based on significance (a potential confounder was adjusted for if the p value of its OR was below 0.05 or 0.20), on the CIE criterion with a 10% cutoff (a potential confounder was adjusted for if the adjustment changed the OR by more than 10% of the unadjusted OR), or on the CIE criterion with a simulated cutoff[10] yielding a 5% level of type I error (that is, a variable without confounding effect has a probability of 5% to be identified as a confounder). Each of these adjustments was performed separately. Those datasets that showed no empirical evidence of a need for confounder adjustment belonged to the second group. We reported the percentage of the simulated datasets showing evidence of confounder adjustment. Two regression models were fitted to both groups. Both models treated Y as the dependent variable (denoted by OR_x and OR_x' for datasets in the first and second group respectively). However, one model assumed that X was the sole independent variable, while the other further adjusted for Z (denoted by $OR_{x,z}$ and $OR_{x,z}'$ for datasets in the first and second group, respectively).

Performance assessment. The estimation accuracy was assessed from the percentage error, given by $\dfrac{\sum_{i=1}^{k} \hat{OR}_i - \exp(0.1)}{\exp(0.1)}$, and the root-mean-square error (RMSE), given by $\sum_{i=1}^{k} \dfrac{\sqrt{(\hat{OR}_i - \exp(0.1))^2}}{k}$, where k and $\hat{OR}_i$ are the total number of simulations and the estimated OR for the ith simulation respectively. All simulations were carried out using R version 3.0.0 and the simulation syntax is given in the Supplementary Information.

1. Budtz-Jorgensen, E., Keiding, N., Grandjean, P. & Weihe, P. Confounder selection in environmental epidemiology: Assessment of health effects of prenatal mercury exposure. *Ann Epidemiol* **17**, 27–35 (2007).
2. Rothman, K. J., Greenland, S. & Lash, T. L. *Modern Epidemiology.* (Lippincott Williams & Wilkins, 2008).
3. Greenland, S. & Morgenstern, H. Confounding in health research. *Annu Rev Public Health* **22**, 189–212 (2001).
4. McNamee, R. Regression modelling and other methods to control confounding. *Occup Environ Med* **62**, 500–506 (2005).
5. Hernandez, A. V., Eijkemans, M. J. C. & Steyerberg, E. W. Randomized controlled trials with time-to-event outcomes: How much does prespecified covariate adjustment increase power? *Ann Epidemiol* **16**, 41–48 (2006).
6. Hernandez, A. V., Steyerberg, E. W. & Habbema, J. D. F. Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. *J Clin Epidemiol* **57**, 454–460 (2004).
7. Maldonado, G. & Greenland, S. Simulation study of confounder-selection strategies. *Am J Epidemiol* **138**, 923–936 (1993).
8. Mickey, R. M. & Greenland, S. The impact of confounder selection criteria on effect estimation. *Am J Epidemiol* **129**, 125–137 (1989).
9. Hernan, M. A., Hernandez-Diaz, S., Werler, M. M. & Mitchell, A. A. Causal knowledge as a prerequisite for confounding evaluation: An application to birth defects epidemiology. *Am J Epidemiol* **155**, 176–184 (2002).
10. Lee, P. H. Is the cutoff of 10% appropriate for the change-in-estimate confounder identification criterion? *J Epidemiol* **24**, 161–167 (2014).
11. Greenland, S., Pearl, J. & Robins, J. M. Casual diagrams for epidemiologic research. *Epidemiology* **10**, 37–48 (1999).
12. McNamee, R. Confounding and confounders. *Occup Environ Med* **60**, 227–234 (2003).
13. Evans, D., Chaix, B., Lobbedez, T., Verger, C. & Flahault, A. Combining directed acyclic graphs and the change-in-estimate procedure as a novel approach to adjustment-variable selection in epidemiology. *BMC Med Res Methodol* **12**, 156 (2012).
14. Weng, H.-Y., Hsueh, Y.-H., Messam, L. L. M. & Hertz-Picciotto, I. Methods of covariate selection: Directed Acyclic Graphs and the Change-in-Estimate procedure. *Am J Epidemiol* **169**, 1182–1190 (2009).
15. Arah, O. A., Chiba, Y. & Greenland, S. Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders. *Ann Epidemiol* **18**, 637–646, doi:10.1016/j.annepidem.2008.04.003 (2008).
16. Fewell, Z., Davey Smith, G. & Sterne, J. A. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *Am J Epidemiol* **166**, 646–655, doi:10.1093/aje/kwm165 (2007).
17. Tong, S. & Lu, Y. Identification of confounders in the assessment of the relationship between lead exposure and child development. *Ann Epidemiol* **11**, 38–45 (2001).
18. Brenner, H. & Blettner, M. Controlling for continuous confounders in epidemiologic research. *Epidemiology* **8**, 429–434 (1997).
19. Robinson, L. D. & Jewell, N. P. Some surprising results about covariate adjustment in logistic regression models. *Int Stat Rev* **59**, 227–240 (1991).
20. Maldonado, G. Toward a clearer understanding of causal concepts in epidemiology. *Ann Epidemiol* **23**, 743–749 (2013).

Author contributions

The sole author was responsible for all parts of the manuscript.

Additional information