

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

A Sensitivity Analysis for Mistaking Mediators as Confounders in the Perspective of Causal Diagrams: a Simulation Study

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2016-015640
Article Type:	Research
Date Submitted by the Author:	22-Dec-2016
Complete List of Authors:	Wang, Tingting; Shandong University, Department of Epidemiology and Biostatistics Li, Hongkai; Shandong University, Department of Epidemiology and Biostatistics Su, Ping; Shandong University, Department of Epidemiology and Biostatistics Yu, Yuanyuan; Shandong University, Department of Epidemiology and Biostatistics Sun, Xiaoru; Shandong University, Department of Epidemiology and Biostatistics Liu, Yi; Shandong University, Department of Epidemiology and Biostatistics Yuan, Zhongshang; Shandong University, Department of Epidemiology and Biostatistics Xue, Fuzhong; Shandong University, Department of Epidemiology and Biostatistics
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Public health
Keywords:	EPIDEMIOLGY, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS

SCHOLARONE™
Manuscripts

A Sensitivity Analysis for Mistaking Mediators as Confounders in the Perspective of Causal Diagrams: a Simulation Study

Tingting Wang[†], Hongkai Li[†], Ping Su, Yuanyuan Yu, Xiaoru Sun, Yi Liu, Zhongshang Yuan, Fuzhong Xue*
Department of biostatistics, School of Public Health, Shandong University, Jinan City, Shandong Province, P.R.China

Email addresses:

TTW: wangtingting161780@163.com

HKL: lihongkaiyouxiang@163.com

PS: sdsuping@163.com

YYY: yu_yy_1993@163.com

XRS: sunny150139@126.com

YL: liuyi238@sdu.edu.cn

ZSY: yuanzhongshang@sdu.edu.cn

*Corresponding Author

xuefzh@sdu.edu.cn (FZX)

Word Count: 2899

Abstract

Objectives: In observational studies, when the underlying structure is unknown and only limited knowledge is available, a sensitivity analysis between the effect of exposure-mediator and the effect of mediator-outcome was dissected under mistakenly controlling for mediators to estimate the total effect of exposure on outcome. Through simulation, we focused on six causal diagrams concerning different roles of mediators to compare the sensitivity of the effect of exposure-mediator with the effect of mediator-outcome in adjusting for mediator under the framework of logistic regression model.

Setting: Based on the causal relationships in real world, we generated the simulation data by varying across the effect of exposure-mediator and the effect of mediator-outcome. And compared the bias of varying across the effect exposure-mediator with the bias of varying across the effect mediator-outcome mistakenly adjusting for mediator. The magnitude of bias was defined by the

1
2
3 difference between the estimated causal effect by logistic regression models and the
4 true causal effect based on *do calculus*.
5
6

7 **Results:** Simulation results revealed that, when there are only a single mediator, two
8 series mediators, two independent parallel mediators and two correlated parallel
9 mediators, the bias that varied across the effect exposure-mediator was larger than the
10 one that varied across the effect mediator-outcome under adjusting for the mediator.
11 However, the bias performances were opposite result in scenarios of a single mediator
12 and two independent parallel mediators in the presence of unobserved confounders.
13
14

15 **Conclusions:** We concluded that the sensitivity between the effect exposure-mediator
16 and the effect mediator-outcome was related to whether there is unobserved
17 confounder in causal diagrams.
18
19

20 **Keywords:** observational study; mediator; confounder; causal diagram; sensitivity
21 analysis
22
23

24 **Strengths and limitations of this study**

25 Based on the *do calculus* calculated the total causal effect of exposure on outcome in
26 perspective of causal diagram.
27

28 Various simulations were conducted to assess the consequences between the effect of
29 exposure-mediator ($E \rightarrow M$) and the effect of mediator-outcome ($M \rightarrow D$) under
30 mistakenly adjust for mediator in logistic regression model.
31

32 The simulated parameters were based on the observational study.
33

34 The limitation was only considered the binary variable and was not known the
35 conclusion of continuous variable.
36
37

38 **Introduction**

39 Estimating the total effects of the exposure (E) on the outcome (D) is still a great
40 challenge in the analytic epidemiology study, because researchers often do not fully
41
42
43
44
45
46
47
48
49
50
51

1
2
3 acknowledge the distinction between confounders and mediators.¹⁻³ If confounders
4 and mediators are misclassified, the ability to control confounder in the estimation of
5 the total effect of the exposure on the outcome is hampered. Causal diagrams have
6 provided a formal conceptual framework to identify and select confounders,⁴⁻⁵ so that
7 it can avoid falling into analytic pitfalls.⁶ In practice, even the underlying causal
8 diagrams and the role of covariates (mediator, confounder, collider and instrumental
9 variable) are not all known, investigators usually controlled the covariates that are
10 both associated with the outcome and exposure to estimate the total effect of the
11 exposure on the outcome.⁷⁻¹⁰ Therefore, our paper paid attention to the bias behavior
12 that varied across the effect exposure-mediator ($E \rightarrow M$) and the effect
13 mediator-outcome ($M \rightarrow D$) under mistakenly regraded mediator as confounder in
14 logistic regression model.

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30 Recently, epidemiologists mainly explore the mediator mechanisms of the total
31 effect, direct effect and indirect effect of exposure on outcome.¹¹⁻¹³ Arbitrarily
32 controlling for a mediator would generally obtain biased estimates of the total effect
33 of the exposure on the outcome.^{6, 14-15} Nevertheless, in the perspective of causal
34 diagrams, little attention was paid to the consequences of biases of mistakenly
35 adjusting for mediators in the logistic regression model. Hence, we focused on the
36 sensitivity analysis technique to assess the impact between the effect $E \rightarrow M$ and the
37 effect $M \rightarrow D$ with adjusting for mediator under the framework of logistic regression
38 model.

39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
60 In this paper, six typical causal diagrams corresponding to causal correlation are
given in Figure 1. We considered a study examining the effect of a potentially
beneficial exposure E on outcome D and explored the sensitivity of the effect $E \rightarrow D$

1
2
3 and the effect $M \rightarrow D$. And we performed various quantitative simulations to dissect
4 the bias that varied across the effect $E \rightarrow M$ and the one that varied across the effect
5 $M \rightarrow D$ under the models of adjusting for different mediators. It may provide a guide
6 for studying the importance between the effect magnitude of pathway and the
7 direction from exposure to mediator or from mediator to outcome.
8
9

14 Methods

15
16 A directed acyclic graph (DAG) is composed of variables (nodes) and arrows
17 (directed edges) between nodes such that the graph is acyclic. Pearl formalized causal
18 diagrams as directed acyclic graphs (DAGs), providing investigators with powerful
19 tools for bias assessment.¹⁶ The causal directed acyclic graph theory provides a device
20 for deducing the statistical associations implied by causal relations. Furthermore,
21 given a set of observed statistical associations, a researcher armed with causal
22 diagrams theory can systematically characterize all causal structures compatible with
23 the observations.¹⁷⁻¹⁸
24
25
26
27
28
29
30
31
32
33
34
35
36

37 The total causal effect can be calculated based on the *do-calculus* and *back-door*
38 criterion proposed by Judea Pearl.¹⁹⁻²⁰ For exposure X and outcome Y , a set of
39 variables Z is said to satisfy the backdoor path criterion with respect to (X, Y) if no
40 variable in Z is a descendant of X and if Z blocks all back-door paths from X to Y .
41
42
43
44
45
46
47 Then the causal effect of X on Y is given by the formula,

$$48 \quad P(y | do(x)) = \sum_z P(y | x, z)P(z)$$

49
50 Note that the expression on the right hand side of the equation is simply a
51 standardized mean. The difference $E(Y | do(x')) - E(Y | do(x''))$ is taken as the
52 definition of “causal effect”, where x' and x'' are two distinct realizations of X .¹⁹
53
54
55
56
57
58
59
60

Besides it can be shown that if ignorability holds for $Y(x)$ and X (alternatively if there are no back-door paths from X to Y in the corresponding causal DAGs), then $p(y | do(x)) = p(y | x)$.²¹⁻²² Taking Figure 1a as an example, the true total causal effect (β) of E on D on the scale of logarithm odds ratio was equal to

$$\begin{aligned}\beta &= \text{logit}(P(D=1 | do(E=1))) - \text{logit}(P(D=1 | do(E=0))) \\ &= \text{logit}\left(\sum_M P(D=1 | E=1, M)\right) - \text{logit}\left(\sum_M P(D=1 | E=0, M)\right)\end{aligned}$$

The estimation after adjusting for M was equal to

$$\beta_M = \text{logit}(P(D=1 | E=1, M)) - \text{logit}(P(D=1 | E=0, M))$$

Note that the bias was defined by taking a difference between estimated exposure effect by adjusting for mediator using logistic regression and the true total causal effect based on *do calculus* i.e. $\text{bias} = \beta_M - \beta$. We dissected the biases behavior between the effect $E \rightarrow M$ and the effect $M \rightarrow D$ mistakenly controlling mediator under logistic regression model.

Simulation

As shown in Figure 1, six scenarios were designed to dissect bias behaviors caused by mistaking mediators as confounders using logistic regression model. We made the following assumptions for the simulation: 1) all variables were binary following a Bernoulli distribution; 2) the effect from parent nodes to their child node were positive and log-linearly additive. Taking Figure 1a as an example, we randomly generated the exposure E following a Bernoulli distribution (i.e. let $P(E=1) = \pi$),

Then, $P_M = \exp(\alpha_0 + c_1 E) / (1 + \exp(\alpha_0 + c_1 E))$ for calculating the distribution probability of child node M from its parent node E, similarity, $P_D = \exp(\beta_0 + c_2 M + c_0 E) / (1 + \exp(\beta_0 + c_2 M + c_0 E))$ to generate the distribution

1
2
3
4 probability of D, where the parameters α_0 and β_0 denoted the intercept of M and
5
6 D respectively, and effect parameter c_1, c_2, c_0 referred to the effects of the parent node
7
8 on their corresponding child node using log odds ratio scale.
9

10
11 After generating data, we have dissected the biases behavior between the effect
12
13 $E \rightarrow M$ and the effect $M \rightarrow D$ mistakenly controlling mediator under logistic regression
14
15 model. In scenario 1 (Figure 1a), we compared the bias that varied across the effect
16
17 $E \rightarrow M$ with the one that varied across the effect $M \rightarrow D$ with adjusting for mediator M
18
19 under the logistic regression model. Similarly, in scenario 2 (Figure 1b), the bias that
20
21 varied across the effect $E \rightarrow M_1$ and the effect $M_1 \rightarrow M_2$ with the bias that varied across
22
23 the effect $M_2 \rightarrow D$ were explored with adjustment for M_1 , adjustment for M_2 and
24
25 adjustment for $M_1 M_2$ under the logistic regression model, respectively. In scenario 3
26
27 (Figure 1c), we dissected the bias that varied across the effect $E \rightarrow M_1$ with the bias
28
29 that varied across the effect $M_1 \rightarrow D$ and payed close attention to the bias that varied
30
31 across the effect $E \rightarrow M_2$ with the bias that varied across the effect $M_2 \rightarrow D$ with
32
33 adjustment for M_1 , adjustment for M_2 and adjustment for $M_1 M_2$ under the logistic
34
35 regression model, respectively. The biases comparison of scenario 4 (Figure 1d) were
36
37 same as scenario 3 (Figure 1c). In scenario 5 (Figure 1e), we excavated the bias that
38
39 varied across the effect $E \rightarrow M$ and the bias that varied across the effect $M \rightarrow D$ with
40
41 adjustment for M under the logistic regression model. In scenario 6 (Figure 1f), we
42
43 compared the bias that varied across the effect $E \rightarrow M_1$ with the bias that varied across
44
45 the effect $M_1 \rightarrow D$ and explored the bias that varied across the effect $E \rightarrow M_2$ with the
46
47 bias that varied across the effect $M_2 \rightarrow D$ with adjustment for M_1 , adjustment for M_2
48
49
50
51
52
53
54
55
56
57
58
59
60

and adjustment for M_1 M_2 under the logistic regression model, respectively. We explored the biases behavior with adjusting the mediator under logistic regression model and thus identified the sensitivity between effect of exposure-mediator and effect of mediator-outcome.

For each of 6 simulation scenarios, we observed bias performances of varying across distinct effects under adjusting mediator using logistic regression model with 1000 simulations repetitions. All simulations were conducted using software R from CRAN (<http://cran.r-project.org/>).

Results

Scenario 1: one single mediator (Figure 1a)

For Figure 1(a) of the simplest case, E has a direct ($E \rightarrow D$) effect and an indirect ($E \rightarrow M \rightarrow D$) effect on D . Figure 2A depicted that the bias that varied across the effect $E \rightarrow M$ was obviously larger than the bias that varied across the effect $M \rightarrow D$. In particular, if the effect $E \rightarrow M$ was specified to zero in Figure 2B, M became an independent cause of the outcome, and in this case adjusting for M obtained a positive bias. Moreover, Figure 2 indicated that adjusting for mediator M using logistic regression model was indeed biased to the total effect of the exposure on the outcome.

The true total causal effect (β) of E on D was calculated as

$$\begin{aligned} \beta = \log(OR) &= \log\left(\frac{P(D=1 | do(E=1))P(D=0 | do(E=0))}{P(D=0 | do(E=1))P(D=1 | do(E=0))}\right) \\ &= \log\left(\frac{\sum_M P(D=1 | E=1, M)P(M | E=1) \sum_M P(D=0 | E=0, M)P(M | E=0)}{\sum_M P(D=0 | E=1, M)P(M | E=1) \sum_M P(D=1 | E=0, M)P(M | E=0)}\right) \end{aligned}$$

By conditioning on mediator M , the effect of E on D was equal to

$$\beta_M = \log(OR_M) = \log\left(\frac{P(D=1 | E=1, M)P(D=0 | E=0, M)}{P(D=0 | E=1, M)P(D=1 | E=0, M)}\right)$$

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

After a series of derivations (Appendix 1), we obtained $bias = 0$ under condition of $c_2 = 0$ (c_2 of the effect $M \rightarrow D$), suggesting that the estimation of E on D was unbiased under adjusting for M when the effect $M \rightarrow D$ (c_2) was null. Only if the $c_2 \neq 0$ and $c_1 = 0$ (c_1 of the effect $E \rightarrow M$), 1) while $bias = 0$, if $c_0 = 0$ (c_0 the effect $E \rightarrow D$), indicating that the estimation of E on D was unbiased with adjusting for M; 2) the $bias > 0$, if $c_0 > 0$, indicating that the effect of E on D overestimated with adjusting for M; 3) the $bias < 0$, if $c_0 > 0$, indicating that the effect of E on D underestimated with adjusting for M. And we gained $bias > 0$, if $c_1 c_2 < 0$, suggesting that the effect of E on D overestimated with adjusting for M when the effect $E \rightarrow M$ (c_1) and the effect $M \rightarrow D$ (c_2) were opposite (i.e. the effect $E \rightarrow M$ was positive ($c_1 > 0$) and the effect $M \rightarrow D$ was negative ($c_2 < 0$) or the effect $E \rightarrow M$ was negative ($c_1 < 0$) and the effect $M \rightarrow D$ was positive ($c_2 > 0$)). The result was $bias < 0$ under conditions of $c_1 c_2 > 0$, indicating that the effect of E on D underestimated with adjusting for M when the effect $E \rightarrow M$ (c_1) and the effect $M \rightarrow D$ (c_2) were positive or the effect $E \rightarrow M$ (c_1) and the effect $M \rightarrow D$ (c_2) were negative. The detail of theoretical derivation was in Appendix 1.

Scenario 2: two series mediators (Figure 1b)

Figure 1(b) is a depiction through two series mediators, decomposing total effects into direct ($E \rightarrow D$) and indirect ($E \rightarrow M_1 \rightarrow M_2 \rightarrow D$) components. The bias that varied across the effect $E \rightarrow M_1$ was larger than the one that varied across the effect $M_2 \rightarrow D$ under adjustment for M_1 , adjustment for M_2 and adjustment for $M_1 M_2$ in Figure 3, when the correlation of series mediators was strong to avoid M_2 became an independent cause

of the outcome.

Scenario 3: two independent parallel mediators (Figure 1c)

Figure 1c shows that the exposure E independently causes M_1 and M_2 and indirectly influences the outcome D through M_1 and M_2 , forming three causal paths $E \rightarrow D$, $E \rightarrow M_1 \rightarrow D$ and $E \rightarrow M_2 \rightarrow D$. We obtained that the bias that varied across the effect $E \rightarrow M_1$ was distinctly larger than the one that varied across the effect $M_1 \rightarrow D$ under adjustment for M_1 in Figure 4A. However, the bias with the effect $E \rightarrow M_2$ increasing was nearly equal to the one with the effect $M_2 \rightarrow D$ increasing under identical adjustment for M_1 in Figure 4A. Then, an above similar result can be obtained the biases behavior in Figure 4B. In addition, Figure 4C indicated that biases that varied across the effect $E \rightarrow M_1$ and varied across the effect $E \rightarrow M_2$ were obviously larger than the one with the effect $M_1 \rightarrow D$ and the effect $M_2 \rightarrow D$ increasing under adjustment for M_1 M_2 .

Scenario 4: two correlated parallel mediators (Figure 1d)

In Figure 1d, there exists five paths from E to D: $E \rightarrow D$, $E \rightarrow M_1 \rightarrow D$, $E \rightarrow M_2 \rightarrow D$, $E \rightarrow M_1 \leftarrow M_2 \rightarrow D$ and $E \rightarrow M_2 \rightarrow M_1 \rightarrow D$. In particular, the path $E \rightarrow M_1 \leftarrow M_2 \rightarrow D$ is a blocked path, due to the M_1 being a collider node. Figure 5A indicated that the bias of the effect $E \rightarrow M_1$ was obviously larger than the one of the effect $M_1 \rightarrow D$ under the adjustment for M_1 with the OR of effect increasing. However, the bias that varied across the effect $E \rightarrow M_2$ was almost equal to the one that varied across the effect $M_2 \rightarrow D$ under identical adjustment model. Similarly, Figure 5B showed an analogous result for biases behavior. Besides, Figure 5C manifested that biases that varied across

1
2
3
4 the effect $E \rightarrow M_1$ and the effect $E \rightarrow M_2$ were larger than the ones that varied across the
5
6 effect $M_1 \rightarrow D$ and the effect $M_2 \rightarrow D$ under adjusting for M_1 and M_2 . Simultaneously,
7
8 the effect $E \rightarrow M_2$ was more sensitive than the effect $E \rightarrow M_1$, which adjustment for the
9
10 collider node M_1 would partially open the path $E \rightarrow M_1 \leftarrow M_2 \rightarrow D$.

11
12
13
14 ***Scenario 5: a single mediator with an unobserved confounder (Figure 1e)***

15
16 For Figure 1e, in the framework of a causal diagram with exposure E , outcome D ,
17
18 mediator M and unobserved confounder U , it revealed that the bias that varied across
19
20 the effect $E \rightarrow M$ was lower than the one that varied across the effect $M \rightarrow D$ in the
21
22 presence of unobserved confounder that distorts the association between the exposure
23
24 and outcome ($E \leftarrow U \rightarrow D$) in Figure 6.

25
26
27
28
29 ***Scenario 6: two parallel mediators with an unobserved confounder (Figure 1f)***

30
31 As described above, Figure 1f is a depiction of two parallel mediators M_1 and M_2 with
32
33 an unobserved confounder U . Figure 7A indicated that the bias that varied across the
34
35 effect $E \rightarrow M_1$ was obviously less than the one that varied across the effect $M_1 \rightarrow D$
36
37 under the adjustment for M_1 , while the bias with the effect $E \rightarrow M_2$ increasing was
38
39 larger than the bias with the effect $M_2 \rightarrow D$ increasing under the identical adjustment
40
41 for M_1 . A similar result can also be obtained in Figure 7B. Besides, biases that varied
42
43 across the effect $E \rightarrow M_1$ and varied across the effect $E \rightarrow M_2$ were distinctly less than
44
45 the ones with the effect $M_1 \rightarrow D$ and the effect $M_2 \rightarrow D$ increasing under common
46
47 model of adjusting for M_1 and M_2 (Figure 7C).

48
49
50
51
52
53
54 **Discussion**

55
56 In this study, we dissected the sensitivity of the bias that varied across the effect
57
58
59
60

1
2
3
4 exposure-mediator and the one that varied across the effect mediator-outcome with
5
6 adjusting for mediators under the framework of logistic regression model. When there
7
8 are a single mediator (Figure 1a in scenario 1), two series mediators (Figure 1b in
9
10 scenario 2), two independent parallel (Figure 1c in scenario 3) and two correlated
11
12 parallel mediators (Figure 1d in scenario 4), the bias that varied across the effect
13
14 exposure-mediator was larger than the one that varied across the effect
15
16 mediator-outcome under adjusting for the mediator (Figure 2, Figure 3, Figure 4 &
17
18 Figure 5). However, there are a single mediator and two independent parallel
19
20 mediators in the presence of the unobserved confounder (Figure 1e in scenario 5 &
21
22 Figure 1f in scenario 6), which the opposite result was presented that the bias that
23
24 varied across the effect mediator-outcome was larger than the one that varied across
25
26 the effect exposure-mediator under adjusting for the mediator (Figure 6 & Figure 7).
27
28
29
30
31
32

33
34 Obviously, adjustment for mediator indeed led to bias for estimating the total effect
35
36 of the exposure on outcome.^{6, 14-15} Unfortunately, mediators and confounders were
37
38 indistinguishable in terms of statistical association and conceptual grounds³.
39
40 Investigators also paid little attention to the consequences of biases caused by
41
42 mistaking mediators as confounders to estimate the total effect of exposure on
43
44 outcome under logistic regression model. Most of the studies focused on the
45
46 mediation effect analysis such as the calculation of direct effects and indirect effect.¹⁵
47
48
49
50
51
52
53
54
55
56
57
58
59
60
23-25 Our study results revealed that the effect exposure-mediator was more sensitive
than the effect mediator-outcome under adjusting for the mediator in the absence of
the unobserved confounder in causal diagrams (Figure 1a, Figure 1b, Figure 1c &

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 1d). Nevertheless, the opposite result that was presented that the effect mediator-outcome was more sensitive than the effect exposure-mediator in the presence of the unobserved confounder in causal diagrams (Figure 1e & Figure 1f). Therefore, the biases that varied across different effects depended on the causal diagrams framework whether there existed unobserved confounder.

Note that, in the perspective of diagrams, our simulation study was not comprehensive to evaluate the bias behavior of adjusting for the mediator in logistic regression, since it only considered binary variables, the certain scenarios of effect size and the common type of models. The present work ought to reinforce the mechanisms and conceptual frameworks of confounder and mediator form causal diagrams so as to assess the total effect, indirect effect and direct effect, hence avoid falling into analytic pitfalls.

Conclusion

In conclusion, we showed that the sensitivity between the effect exposure-mediator and the effect mediator-outcome was related to whether there is confounder in causal diagrams. The effect exposure-mediator was more sensitive than the effect mediator-outcome under adjusting for the mediator in the absence of unobserved confounder, however, the sensitivity was opposite in the presence of unobserved confounder.

Statements

Ethics approval and materials

Not applicable

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by grants from National Natural Science Foundation of China (grant number 81573259)

Provenance and peer review

Not commissioned; externally peer reviewed.

Data sharing statement

No additional data are available.

Authors' contributions

TTW and HKL jointly conceived the idea behind the article and designed the study. TTW helped conduct the literature review, performed the simulation and prepared the first draft of the manuscript. PS, YYY, XRS, YL and ZSY participated in the design of the study and the revision of the manuscript. FZX advised on critical revision of the manuscript for important intellectual content. All authors read and approved the final manuscript.

Acknowledgments

We would like to thank anonymous reviewers and academic editors for providing us with constructive comments and suggestions and also wish to acknowledge our colleagues for their invaluable work.

Open Access

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work noncommercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

References

- 1 Bulterys M, Morgenstern H. Confounding or intermediate effect? An appraisal of iatrogenic bias in perinatal AIDS research. *Paediatr Perinat Epidemiol* 1993;7:387–394.
- 2 Hernán MA, Hernandez-Diaz S, Werler MM, et al. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol* 2002;155:176–184.
- 3 MacKinnon DP, Krull JL, and Lockwood CM. Equivalence of the mediation, confounding and suppression effect. *Prevention Science* 2000;1:173–181.
- 4 Howards PP, Schisterman EF, Poole C, et al. “Toward a clearer definition of confounding”

- revisited with directed acyclic graphs, *Am J Epidemiol* 2012;176:506–511.
- 5 Williamson EJ, Aitken Z, Lawrie J, et al. Introduction to causal diagrams for confounder selection. *Respirology*. 2014;19:303-311.
 - 6 Wilcox AJ, Weinberg CR, Basso O. On the pitfalls of adjusting for gestational age at birth. *Am J Epidemiol*. 2011;174:1062–1068.
 - 7 Hernán MA, Hernandez-Diaz S, Werler MM, et al. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol* 2002; 155:176–84.
 - 8 HK. Methodological issues of confounding in analytical epidemiology studies. *Caspian J intern Med* 2012;3:488-495.
 - 9 Kamangar F MD. Confounding variable in epidemiology studies: Basics and beyond. *Arch Tran Med*. 2012;15:508-516.
 - 10 Gaskins AJ, Schisterman EF. The effect of lipid adjustment on the analysis of environmental contaminants and the outcome of human health risks. *Methods Mol Biol* 2009;580:371-381.
 - 11 Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 1992;3:143–155.
 - 12 Pearl J. Direct and indirect effects. Seventeenth conference of uncertainty in artificial intelligence, 2001. San Francisco, CA: Morgan Kaufmann, 2001.
 - 13 Huang B, Sivaganesan S, Succop P, et al. Statistical assessment of mediational effects for logistic mediational models. *Statist. Med* 2004;23:2713-2728.
 - 14 Robins J. The control of confounding by intermediate variable. *Statistics in Medicine*. 1989;8:679-701.
 - 15 Schisterman EF, Cole SR, Platt RW. Overadjustment Bias and Unnecessary Adjustment in Epidemiologic Studies. *Epidemiology*. 2009;20:488–495.
 - 16 Pearl J. Causal diagrams for empirical research. *Biometrika*. 1995;82:669-688.
 - 17 Pearl J. Causality: models, reasoning, and inference. *Econometric Theory*. 2003;19: 675-685.
 - 18 Joffe M, Gambhir M, Chadeau-Hyam M, et al. Causal diagrams in systems epidemiology. *Emerg Themes Epidemiol*. 2012; 9:1-18.
 - 19 Pearl, J. Causality: Models, reasoning, and inference. Cambridge: Cambridge University Press; 2000.
 - 20 Pearl J. Causal inference in statistics: an overview. *Statistics Surveys*. 2009; 3:96–146.
 - 21 Rubin DB. Should observational studies be designed to allow lack of balance in covariate distributions across treatment groups? *Statistics in Medicine*. 2009;28:1420-1423.
 - 22 Shpitser I. Counterfactual graphical models for longitudinal mediation analysis with unobserved confounding. *Cognitive Science* 2013;37:1011-1035.
 - 23 Lange T, Rasmussen M, Thygesen LC. Assessing natural direct and indirect effects through multiple pathways. *American Journal of Epidemiology*. 2013;179, 513–518.
 - 24 Shpitser, Ilya and VanderWeele, Tyler J. "A complete graphical criterion for the adjustment formula in mediation analysis," *The International Journal of Biostatistics*. 2011;7,1-24.
 - 25 Pearl J. Direct and indirect effects. Seventeenth conference of uncertainty in artificial intelligence, 2001. San Francisco, CA: Morgan Kaufmann, 2001.

Figure 1: Six causal diagrams were designed for estimating the causal effect of E on D. a) a single mediator M; b) two series mediators M_1 and M_2 ; c) two independent

1
2
3 parallel mediators M_1 and M_2 ; d) two correlated parallel mediators M_1 and M_2 ; e) a
4 single mediator with an unobserved confounder U ; f) two independent parallel
5 mediators M_1 and M_2 with an unobserved confounder U .
6
7

8 **Figure 2:** The biases with the effect $E \rightarrow M$ (red) and the effect $M \rightarrow D$ (blue)
9 increasing, respectively. Comparison of the bias of different effects in adjustment
10 mediator. The OR of target effect (e.g. $E \rightarrow M$) from 1 to 10 given other effects fixed
11 $\ln 2$ in Figure 2A. The OR of the effect $M \rightarrow D$ from 1 to 10 with the effect $E \rightarrow M$
12 being equal to zero in Figure 2B (Color figure online).
13
14

15 **Figure 3:** The biases with the effect $E \rightarrow M_1$ (red), the effect $M_1 \rightarrow M_2$ (blue) and the
16 effect $M_2 \rightarrow D$ (black) increasing, respectively. Comparison of the bias of different
17 effects in three adjustment models: A) adjustment for M_1 , B) adjustment for M_2 and C)
18 adjustment for M_1 and M_2 . The OR of target effect (e.g. $E \rightarrow M_1$) from 1 to 10 given
19 the effect $M_1 \rightarrow M_2$ fixed $\ln 8$ and other effects fixed $\ln 2$ in Figure 3 (Color figure
20 online).
21
22
23

24 **Figure 4:** The biases with the effect $E \rightarrow M_1$ (red), the effect $E \rightarrow M_2$ (blue), the effect
25 $M_1 \rightarrow D$ (black) and the effect $M_2 \rightarrow D$ (green) increasing, respectively. Comparison of
26 the bias of different effects in three adjustment models: A) adjustment for M_1 , B)
27 adjustment for M_2 and C) adjustment for M_1 and M_2 . The OR of target effects (e.g. E
28 $\rightarrow M_1$) from 1 to 10 given other edges effects fixed $\ln 2$ in Figure 4 (Color figure
29 online).
30
31
32

33 **Figure 5:** The biases with the effect $E \rightarrow M_1$ (red), the effect $E \rightarrow M_2$ (blue), the effect
34 $M_1 \rightarrow D$ (black), the effect $M_2 \rightarrow D$ (green) and the effect $M_1 \rightarrow M_2$ (purple) increasing,
35 respectively. Comparison of the bias of different effects in three adjustment models: A)
36 adjustment for M_1 , B) adjustment for M_2 and C) adjustment for M_1 and M_2 . The OR
37 of target effects (e.g. $E \rightarrow M_1$) from 1 to 10 given other effects fixed $\ln 2$ in Figure 5
38 (Color figure online).
39
40
41

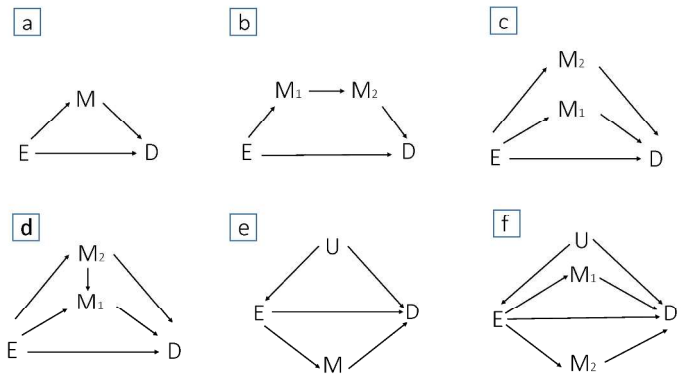
42 **Figure 6:** The biases with the effect $E \rightarrow M$ (red) and the effect $M \rightarrow D$ (blue)
43 respectively. Comparison of the bias of different effects in adjustment mediator M .
44 The OR of target effects (e.g. $E \rightarrow M$) from 1 to 10 given the effects of causal edges
45 fixed $\ln 2$ and the effect of confounder edges fixed $\ln 5$ in Figure $\ln 8$ (Color figure
46 online).
47
48

49 **Figure 7:** The biases with the effect $E \rightarrow M_1$ (red), the effect $E \rightarrow M_2$ (blue), the effect
50 $M_1 \rightarrow D$ (black) and the effect $M_2 \rightarrow D$ (green) respectively. Comparison of the bias of
51 different effects in three adjustment models: A) adjustment for M_1 , B) adjustment for
52 M_2 , and C) adjustment for M_1 and M_2 . The OR of target effects (e.g. $E \rightarrow M_1$) from 1
53 to 10 given the effects of causal edges fixed $\ln 2$ and the effect of confounder edges
54 fixed $\ln 5$ in Figure 7 (Color figure online).
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

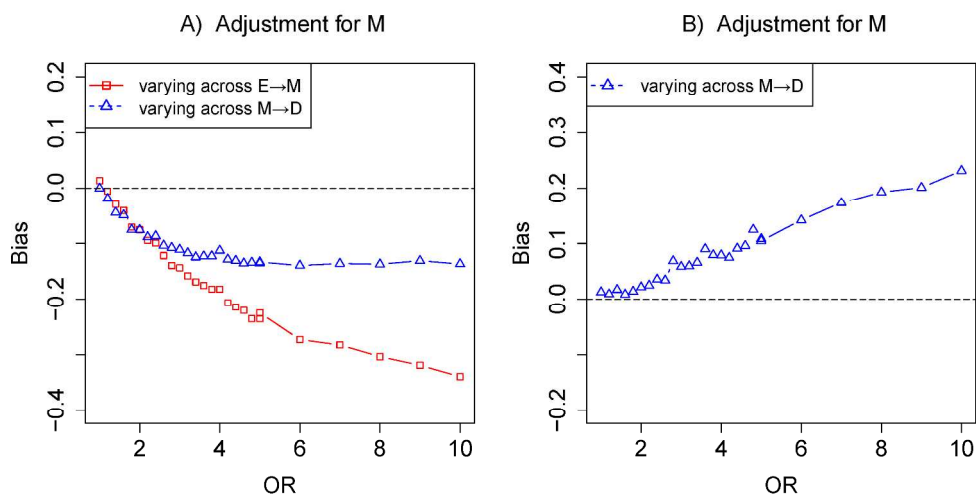
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Six causal diagrams were designed for estimating the causal effect of E on D

338x190mm (300 x 300 DPI)

review only

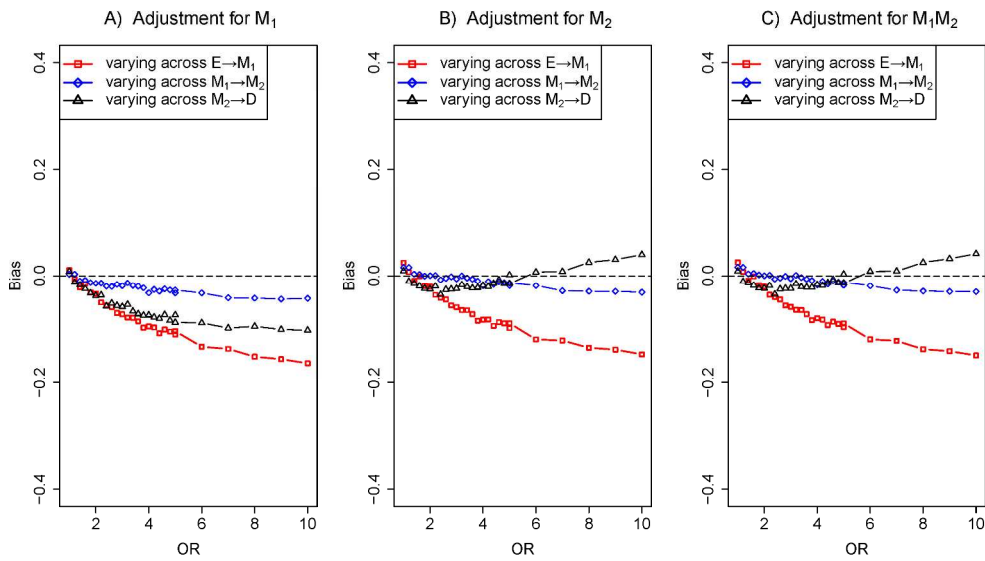


The biases with the effect E→M (red) and the effect M→D (blue) increasing, respectively.

281x148mm (300 x 300 DPI)

review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

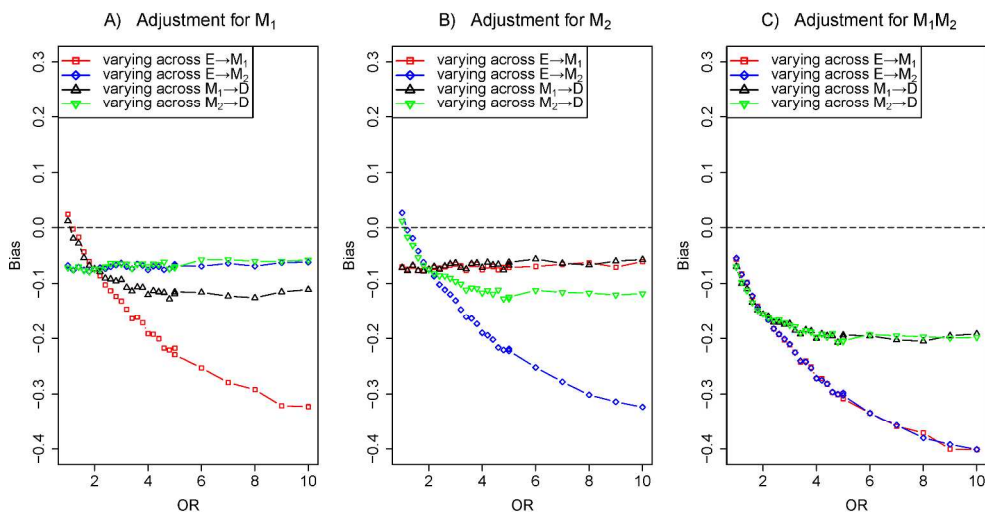


The biases with the effect $E \rightarrow M_1$ (red), the effect $M_1 \rightarrow M_2$ (blue) and the effect $M_2 \rightarrow D$ (black) increasing, respectively.

270x155mm (300 x 300 DPI)

review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

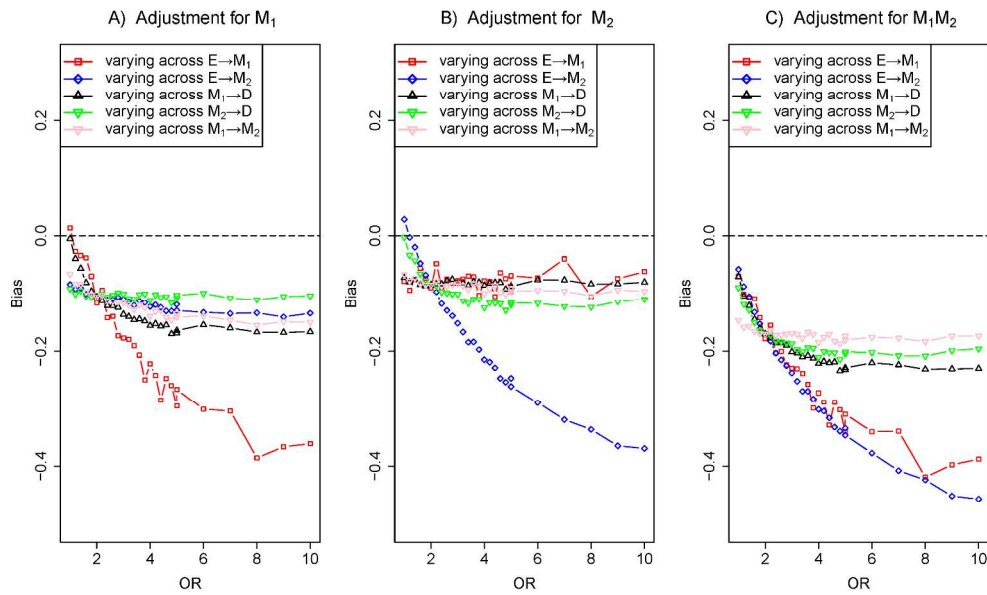


The biases with the effect $E \rightarrow M_1$ (red), the effect $E \rightarrow M_2$ (blue), the effect $M_1 \rightarrow D$ (black) and the effect $M_2 \rightarrow D$ (green) increasing, respectively.

279x147mm (300 x 300 DPI)

review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

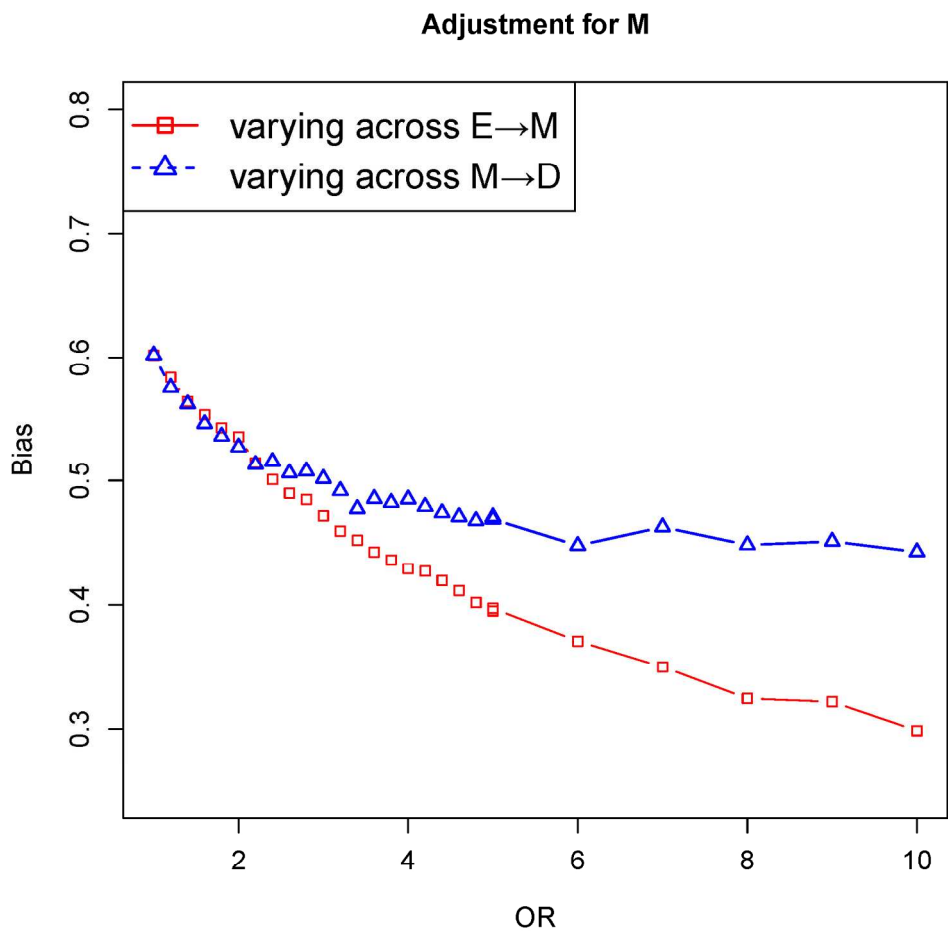


The biases with the effect E→M₁ (red), the effect E→M₂ (blue), the effect M₁→D (black), the effect M₂→D (green) and the effect M₁→M₂ (purple) increasing, respectively.

278x169mm (300 x 300 DPI)

Review only

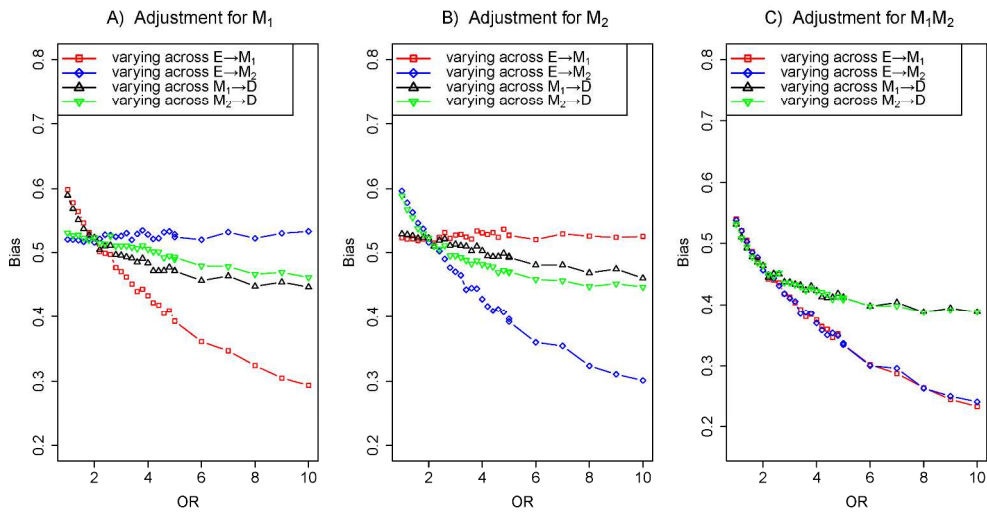
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



The biases with the effect E→M (red) and the effect M→D (blue) respectively.

177x177mm (300 x 300 DPI)





The biases with the effect E→M₁ (red), the effect E→M₂ (blue), the effect M₁→D (black) and the effect M₂→D (green) respectively.

281x148mm (300 x 300 DPI)

review only

Appendix 1:

Theoretical derivation was that adjusting for mediator was biased for estimating the total effect of exposure on outcome using logistic regression model. The bias was defined by taking a difference between estimated exposure effect by adjusting for mediator by logistic regression adjustment model and the true total effect based on *do calculus*.

Deducing the bias of Figure 1a in scenario 1 as follow:

Let E, M and D indicate exposure, mediator and outcome. The effect $E \rightarrow D$, $E \rightarrow M$ and $M \rightarrow D$ defined to c_0 , c_1 and c_2 , respectively.

Suppose the logistic models among them are:

$$E \sim \text{Bernoulli}(1, P_E)$$

$$\text{logit}(P(M = 1 | E)) = \alpha_0 + c_1 E$$

$$\text{logit}(D = 1 | M, E) = \beta_0 + c_2 M + c_0 E$$

The total effect under do calculus ($\ln(OR)$):

$$\begin{aligned} OR &= \frac{P(D = 1 | do(E = 1))P(D = 0 | do(E = 0))}{P(D = 0 | do(E = 1))P(D = 1 | do(E = 0))} \\ &= \frac{\sum_M P(D = 1 | E = 1, M)P(M | E = 1) \sum_M P(D = 0 | E = 0, M)P(M | E = 0)}{\sum_M P(D = 0 | E = 1, M)P(M | E = 1) \sum_M P(D = 1 | E = 0, M)P(M | E = 0)} \end{aligned}$$

The effect of adjusting for mediator M ($\ln(OR_M)$):

$$\begin{aligned} OR_M &= \frac{P(D = 1 | E = 1, M)P(D = 0 | E = 0, M)}{P(D = 0 | E = 1, M)P(D = 1 | E = 0, M)} \\ &= \frac{\exp(c_0 + c_2 \times M + \beta_0)}{1 + \exp(c_0 + c_2 \times M + \beta_0)} \times \frac{1}{1 + \exp(c_2 \times M + \beta_0)} \\ &= \frac{1}{1 + \exp(c_0 + c_2 \times M + \beta_0)} \times \frac{\exp(c_2 \times M + \beta_0)}{1 + \exp(c_2 \times M + \beta_0)} \\ &= \exp(c_0) \end{aligned}$$

where

$$\begin{aligned}
A &= \sum_M P(D=1|E=1, M)P(M|E=1) \\
&= P(D=1|E=1, M=1)P(M=1|E=1) + P(D=1|E=1, M=0)P(M=0|E=1) \\
&= \frac{\exp(c_0 + c_2 + \beta_0)}{1 + \exp(c_0 + c_2 + \beta_0)} \times \frac{\exp(c_1 + \alpha_0)}{1 + \exp(c_1 + \alpha_0)} + \frac{\exp(c_0 + \beta_0)}{1 + \exp(c_0 + \beta_0)} \times \frac{1}{1 + \exp(c_1 + \alpha_0)} \\
B &= \sum_M P(D=0|E=0, M)P(M|E=0) \\
&= P(D=0|E=0, M=1)P(M=1|E=0) + P(D=0|E=0, M=0)P(M=0|E=0) \\
&= \frac{1}{1 + \exp(c_2 + \beta_0)} \times \frac{\exp(\alpha_0)}{1 + \exp(\alpha_0)} + \frac{1}{1 + \exp(\beta_0)} \times \frac{1}{1 + \exp(\alpha_0)} \\
C &= \sum_M P(D=0|E=1, M)P(M|E=1) \\
&= P(D=0|E=1, M=1)P(M=1|E=1) + P(D=0|E=1, M=0)P(M=0|E=1) \\
&= \frac{1}{1 + \exp(c_0 + c_2 + \beta_0)} \times \frac{\exp(c_1 + \alpha_0)}{1 + \exp(c_1 + \alpha_0)} + \frac{1}{1 + \exp(c_0 + \beta_0)} \times \frac{1}{1 + \exp(c_1 + \alpha_0)} \\
D &= \sum_M P(D=1|E=0, M)P(M|E=0) \\
&= P(D=1|E=0, M=1)P(M=1|E=0) + P(D=1|E=0, M=0)P(M=0|E=0) \\
&= \frac{\exp(c_2 + \beta_0)}{1 + \exp(c_2 + \beta_0)} \times \frac{\exp(\alpha_0)}{1 + \exp(\alpha_0)} + \frac{\exp(\beta_0)}{1 + \exp(\beta_0)} \times \frac{1}{1 + \exp(\alpha_0)}
\end{aligned}$$

$$\text{Then } OR = \frac{AB}{CD}$$

Reduction of fractions to a common denominator:

Then the numerators of A, B, C and D were defined by A_1 , B_1 , C_1 and D_1

$$A_1 = \exp(c_0 + c_2 + \beta_0) \times \exp(c_1 + \alpha_0) \times (1 + \exp(c_0 + \beta_0)) + (1 + \exp(c_0 + c_2 + \beta_0)) \times \exp(c_0 + \beta_0)$$

$$B_1 = \exp(\alpha_0) \times (1 + \exp(\beta_0)) + (1 + \exp(c_2 + \beta_0))$$

$$C_1 = \exp(c_1 + \alpha_0) \times (1 + \exp(c_0 + \beta_0)) + (1 + \exp(c_0 + c_2 + \beta_0))$$

$$D_1 = \exp(c_2 + \beta_0) \times \exp(\alpha_0) \times (1 + \exp(\beta_0)) + \exp(\beta_0) \times (1 + \exp(c_2 + \beta_0))$$

Then

$$S_1 = \exp(c_0 + c_2 + \beta_0) \times \exp(c_1 + \alpha_0) \times (1 + \exp(c_0 + \beta_0)) \times \exp(\alpha_0) \times (1 + \exp(\beta_0))$$

$$S_2 = \exp(c_0 + c_2 + \beta_0) \times \exp(c_1 + \alpha_0) \times (1 + \exp(c_0 + \beta_0)) \times (1 + \exp(c_2 + \beta_0))$$

$$S_3 = (1 + \exp(c_0 + c_2 + \beta_0)) \times \exp(c_0 + \beta_0) \times \exp(\alpha_0) \times (1 + \exp(\beta_0))$$

$$S_4 = (1 + \exp(c_0 + c_2 + \beta_0)) \times \exp(c_0 + \beta_0) \times (1 + \exp(c_2 + \beta_0))$$

$$S_5 = \exp(c_1 + \alpha_0) \times (1 + \exp(c_0 + \beta_0)) \times \exp(c_2 + \beta_0) \times \exp(\alpha_0) \times (1 + \exp(\beta_0))$$

$$S_6 = \exp(c_1 + \alpha_0) \times (1 + \exp(c_0 + \beta_0)) \times \exp(\beta_0) \times (1 + \exp(c_2 + \beta_0))$$

$$S_7 = (1 + \exp(c_0 + c_2 + \beta_0)) \times \exp(c_2 + \beta_0) \times \exp(\alpha_0) \times (1 + \exp(\beta_0))$$

$$S_8 = (1 + \exp(c_0 + c_2 + \beta_0)) \times \exp(\beta_0) \times (1 + \exp(c_2 + \beta_0))$$

$$OR = \frac{S_1 + S_2 + S_3 + S_4}{S_5 + S_6 + S_7 + S_8}$$

Comparing the difference of S_1 and S_5 , S_2 and S_6 , S_3 and S_7 , S_4 and S_8 , respectively.

$$OR = \frac{\exp(c_0 + c_2 + \beta_0) \times A_2 + \exp(c_0 + c_2 + \beta_0) \times B_2 + \exp(c_0 + \beta_0) \times C_2 + \exp(c_0 + \beta_0) \times D_2}{\exp(c_2 + \beta_0) \times A_2 + \exp(\beta_0) \times B_2 + \exp(c_2 + \beta_0) \times C_2 + \exp(\beta_0) \times D_2}$$

$$= \exp(c_0) \frac{\exp(c_2) \times A_2 + \exp(c_2) \times B_2 + C_2 + D_2}{\exp(c_2) \times A_2 + B_2 + \exp(c_2) \times C_2 + D_2}$$

$$A_2 = \exp(c_1 + \alpha_0) \times (1 + \exp(c_0 + \beta_0)) \times \exp(\alpha_0) \times (1 + \exp(\beta_0))$$

$$B_2 = \exp(c_1 + \alpha_0) \times (1 + \exp(c_0 + \beta_0)) \times (1 + \exp(c_2 + \beta_0))$$

$$C_2 = (1 + \exp(c_0 + c_2 + \beta_0)) \times \exp(\alpha_0) \times (1 + \exp(\beta_0))$$

$$D_2 = (1 + \exp(c_0 + c_2 + \beta_0)) \times (1 + \exp(c_2 + \beta_0))$$

The effect of adjusting for intermediate M ($\ln(OR_M)$):

$$OR_M = \frac{P(D=1 | E=1, M)P(D=0 | E=0, M)}{P(D=0 | E=1, M)P(D=1 | E=0, M)}$$

$$= \frac{\frac{\exp(c_0 + c_2 \times M + \beta_0)}{1 + \exp(c_0 + c_2 \times M + \beta_0)} \times \frac{1}{1 + \exp(c_2 \times M + \beta_0)}}{\frac{1}{1 + \exp(c_0 + c_2 \times M + \beta_0)} \times \frac{\exp(c_2 \times M + \beta_0)}{1 + \exp(c_2 \times M + \beta_0)}}$$

$$= \exp(c_0)$$

Therefore, we have

$$\text{Bias} = \ln(OR_M) - \ln(OR)$$

$$= \ln\left(\frac{OR_M}{OR}\right)$$

$$= \ln\left(\frac{\exp(c_0)}{\exp(c_0) \frac{\exp(c_2) \times A_2 + \exp(c_2) \times B_2 + C_2 + D_2}{\exp(c_2) \times A_2 + B_2 + \exp(c_2) \times C_2 + D_2}}\right)$$

$$= \ln\left(\frac{\exp(c_2) \times A_2 + B_2 + \exp(c_2) \times C_2 + D_2}{\exp(c_2) \times A_2 + \exp(c_2) \times B_2 + C_2 + D_2}\right)$$

$$\frac{OR_M}{OR} = \frac{\exp(c_0)}{\exp(c_0) \frac{\exp(c_2) \times A_2 + \exp(c_2) \times B_2 + C_2 + D_2}{\exp(c_2) \times A_2 + B_2 + \exp(c_2) \times C_2 + D_2}}$$

$$= \frac{\exp(c_2) \times A_2 + B_2 + \exp(c_2) \times C_2 + D_2}{\exp(c_2) \times A_2 + \exp(c_2) \times B_2 + C_2 + D_2}$$

Focusing on the difference of between $\exp(c_2) \times B_2 + C_2$ and $B_2 + \exp(c_2) \times C_2$

The difference:

$$\begin{aligned}
T(c_1) &= \exp(c_2) \times B_2 + C_2 - (B_2 + \exp(c_2) \times C_2) \\
&= \exp(c_2) \times (B_2 - C_2) - (B_2 - C_2) \\
&= (\exp(c_2) - 1) \times (B_2 - C_2) \\
&= (\exp(c_2) - 1) \times (\exp(c_1 + \alpha_0) \times (1 + \exp(c_0 + \beta_0)) \times (1 + \exp(c_2 + \beta_0)) - \\
&\quad (1 + \exp(c_0 + c_2 + \beta_0)) \times \exp(\alpha_0) \times (1 + \exp(\beta_0))) \\
&= (\exp(c_2) - 1) \times \exp(\alpha_0) \times (\exp(c_1) \times (1 + \exp(c_0 + \beta_0)) \times (1 + \exp(c_2 + \beta_0)) - \\
&\quad (1 + \exp(c_0 + c_2 + \beta_0)) \times (1 + \exp(\beta_0)))
\end{aligned}$$

Then, detailed dissection:

$$1: c_2 = 0, \frac{OR_M}{OR} = 1, OR_M = OR \text{ i.e. } bias = 0$$

$$2: c_2 > 0, \textcircled{1} c_1 = 0, c_0 = 0, \frac{OR_M}{OR} = 1, \text{i.e. } bias = 0$$

$$c_0 > 0, \frac{OR_M}{OR} > 1, \text{i.e. } bias > 0$$

$$c_0 < 0, \frac{OR_M}{OR} < 1, \text{i.e. } bias < 0$$

$$\textcircled{2} c_1 < 0, c_0 = 0, \frac{OR_M}{OR} > 1, \text{i.e. } bias > 0$$

$$c_0 > 0, \frac{OR_M}{OR} > 1, \text{i.e. } bias > 0$$

$$c_0 < 0, \frac{OR_M}{OR} > 1, \text{i.e. } bias > 0$$

In the proof

$$\begin{aligned}
T(c_1) &= (\exp(c_2) - 1) \times \exp(\alpha_0) \times (\exp(c_1) \times (1 + \exp(c_0 + \beta_0)) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0)) - \\
&\quad (1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0))
\end{aligned}$$

$$\text{when } c_0 < 0 \text{ and } c_2 > 0 \Rightarrow \exp(c_0) - 1 < 0 \quad \exp(c_2) - 1 > 0$$

According to $(a-1)(b-1) = ab - a - b + 1$, when $(a-1)(b-1) < 0 \Rightarrow ab + 1 < a + b$

$$\begin{aligned}
1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0) &< 1 + \exp(c_0 + \beta_0) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0) \\
\Rightarrow \exp(c_0 + c_2) + 1 &< \exp(c_0) + \exp(c_2)
\end{aligned}$$

When

$$c_1 < \log\left(\frac{\exp(c_0 + c_2) + 1}{\exp(c_0) + \exp(c_2)}\right) < 0$$

$$c_1 < \log\left(\frac{1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0)}{1 + \exp(c_0 + \beta_0) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0)}\right) < 0$$

$$\Rightarrow \exp(c_1) < \frac{1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0)}{1 + \exp(c_0 + \beta_0) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0)} < 1$$

$$\begin{aligned}
\Rightarrow T(c_1) &= (\exp(c_2) - 1) \times \exp(\alpha_0) \times (\exp(c_1) \times (1 + \exp(c_0 + \beta_0)) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0)) - \\
&\quad (1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0)) \\
&< 0
\end{aligned}$$

Therefore, when $c_2 > 0, c_1 < 0, c_0 < 0$, then $\frac{OR_M}{OR} > 1$, i.e. *bias* > 0

$$\textcircled{3} c_1 > 0, c_0 = 0, \frac{OR_M}{OR} < 1, \text{i.e. } \textit{bias} < 0$$

$$c_0 < 0, \frac{OR_M}{OR} < 1, \text{i.e. } \textit{bias} < 0$$

$$c_0 > 0, \frac{OR_M}{OR} < 1, \text{i.e. } \textit{bias} < 0$$

In the proof

$$T(c_1) = (\exp(c_2) - 1) \times \exp(\beta_0) \times (\exp(c_1) \times (1 + \exp(c_0 + \beta_0) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0)) - (1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0)))$$

$$\text{when } c_0 > 0 \text{ and } c_2 > 0 \Rightarrow \exp(c_0) - 1 > 0 \quad \exp(c_2) - 1 > 0$$

According to $(a - 1)(b - 1) = ab - a - b + 1$, when $ab > 0 \Rightarrow ab + 1 > a + b$

$$1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0) > 1 + \exp(c_0 + \beta_0) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0) \\ \Rightarrow \exp(c_0 + c_2) + 1 > \exp(c_0) + \exp(c_2)$$

When

$$c_1 > \log\left(\frac{\exp(c_0 + c_2) + 1}{\exp(c_0) + \exp(c_2)}\right) > 0$$

$$c_1 > \log\left(\frac{1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0)}{1 + \exp(c_0 + \beta_0) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0)}\right) > 0$$

$$\Rightarrow \exp(c_1) > \frac{1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0)}{1 + \exp(c_0 + \beta_0) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0)} > 1$$

$$\Rightarrow T(c_1) = (\exp(c_2) - 1) \times \exp(\alpha_0) \times (\exp(c_1) \times (1 + \exp(c_0 + \beta_0) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0)) - (1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0))) > 0$$

Therefore, when $c_2 > 0, c_1 > 0, c_0 > 0$, then $\frac{OR_M}{OR} < 1$, i.e. *bias* < 0

$$3: c_2 < 0, \textcircled{1} c_1 = 0, c_0 = 0, \frac{OR_M}{OR} = 1, \text{i.e. } \textit{bias} = 0$$

$$c_0 > 0, \frac{OR_M}{OR} > 1, \text{i.e. } \textit{bias} > 0$$

$$c_0 < 0, \frac{OR_M}{OR} < 1, \text{i.e. } \textit{bias} < 0$$

$$\textcircled{2} c_1 < 0, c_0 = 0, \frac{OR_M}{OR} < 1, \text{i.e. } \textit{bias} < 0$$

$$c_0 < 0, \frac{OR_M}{OR} > 1, \text{i.e. } \textit{bias} < 0$$

$$c_0 > 0, \frac{OR_M}{OR} > 1, \text{i.e. } \textit{bias} < 0$$

In the proof

$$T(c_1) = (\exp(c_2) - 1) \times \exp(\beta_0) \times (\exp(c_1) \times (1 + \exp(c_0 + \beta_0) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0)) - (1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0)))$$

$$\text{when } c_0 > 0 \text{ and } c_2 < 0 \Rightarrow \exp(c_0) - 1 > 0 \quad \exp(c_2) - 1 < 0$$

According to $(a-1)(b-1) = ab - a - b + 1$, when $ab < 0 \Rightarrow ab + 1 < a + b$

$$1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0) < 1 + \exp(c_0 + \beta_0) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0) \\ \Rightarrow \exp(c_0 + c_2) + 1 < \exp(c_0) + \exp(c_2)$$

When

$$c_1 < \log\left(\frac{\exp(c_0 + c_2) + 1}{\exp(c_0) + \exp(c_2)}\right) < 0$$

$$c_1 < \log\left(\frac{1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0)}{1 + \exp(c_0 + \beta_0) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0)}\right) < 0$$

$$\Rightarrow \exp(c_1) < \frac{1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0)}{1 + \exp(c_0 + \beta_0) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0)} < 1$$

$$\Rightarrow T(c_1) = (\exp(c_2) - 1) \times \exp(\alpha_0) \times (\exp(c_1) \times (1 + \exp(c_0 + \beta_0) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0)) - (1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0))) > 0$$

Therefore, when $c_2 < 0, c_1 < 0, c_0 > 0$, then $\frac{OR_M}{OR} < 1$, i.e. *bias* < 0

$$\textcircled{3} c_1 > 0, c_0 = 0, \frac{OR_M}{OR} > 1, \text{ i.e. } \textit{bias} > 0$$

$$c_0 > 0, \frac{OR_M}{OR} > 1, \text{ i.e. } \textit{bias} > 0$$

$$c_0 < 0, \frac{OR_M}{OR} > 1, \text{ i.e. } \textit{bias} > 0$$

In the proof

$$T(c_1) = (\exp(c_2) - 1) \times \exp(\beta_0) \times (\exp(c_1) \times (1 + \exp(c_0 + \beta_0) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0)) - (1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0)))$$

$$\text{when } c_0 < 0 \text{ and } c_2 < 0 \Rightarrow \exp(c_0) - 1 < 0 \quad \exp(c_2) - 1 < 0$$

According to $(a-1)(b-1) = ab - a - b + 1$, when $ab > 0 \Rightarrow ab + 1 > a + b$

$$1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0) > 1 + \exp(c_0 + \beta_0) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0) \\ \Rightarrow \exp(c_0 + c_2) + 1 > \exp(c_0) + \exp(c_2)$$

When

$$c_1 > \log\left(\frac{\exp(c_0 + c_2) + 1}{\exp(c_0) + \exp(c_2)}\right) > 0$$

$$c_1 > \log\left(\frac{1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0)}{1 + \exp(c_0 + \beta_0) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0)}\right) > 0$$

$$\Rightarrow \exp(c_1) > \frac{1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0)}{1 + \exp(c_0 + \beta_0) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0)} > 1$$

$$\Rightarrow T(c_1) = (\exp(c_2) - 1) \times \exp(\alpha_0) \times (\exp(c_1) \times (1 + \exp(c_0 + \beta_0) + \exp(c_2 + \beta_0) + \exp(c_0 + c_2 + 2\beta_0)) - (1 + \exp(c_0 + c_2 + \beta_0) + \exp(\beta_0) + \exp(c_0 + c_2 + 2\beta_0))) < 0$$

Therefore, when $c_2 < 0, c_1 > 0, c_0 < 0$, then $\frac{OR_M}{OR} > 1$, i.e. *bias* > 0

In conclusion:

1: $c_2 = 0, \frac{OR_M}{OR} = 1$, i.e. $OR_M = OR$ i.e. *bias* = 0

2: $c_2 \neq 0, c_1 = 0, c_0 = 0, \frac{OR_M}{OR} = 1$, i.e. *bias* = 0

$c_0 > 0, \frac{OR_M}{OR} > 1$, i.e. *bias* > 0

$c_0 < 0, \frac{OR_M}{OR} < 1$, i.e. *bias* < 0

3: $c_1 c_2 > 0, \frac{OR_M}{OR} < 1$, i.e. *bias* < 0

$c_1 c_2 < 0, \frac{OR_M}{OR} > 1$, i.e. *bias* > 0

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*
Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	1
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any pre-specified hypotheses	3-4
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5-6
Participants	6	(a) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> —Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants	5-6
		(b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5-6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-6
Bias	9	Describe any efforts to address potential sources of bias	5-6
Study size	10	Explain how the study size was arrived at	5-6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Not applicable
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	4-6
		(b) Describe any methods used to examine subgroups and interactions	Not applicable
		(c) Explain how missing data were addressed	Not applicable
		(d) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed	Not applicable

For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

		<i>Cross-sectional study</i> —If applicable, describe analytical methods taking account of sampling strategy	
		(e) Describe any sensitivity analyses	6
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Not applicable
		(b) Give reasons for non-participation at each stage	Not applicable
		(c) Consider use of a flow diagram	Not applicable
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7-10
		(b) Indicate number of participants with missing data for each variable of interest	Not applicable
		(c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	Not applicable
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time	Not applicable
		<i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure	Not applicable
		<i>Cross-sectional study</i> —Report numbers of outcome events or summary measures	7-10
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	7-10
		(b) Report category boundaries when continuous variables were categorized	Not applicable
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not applicable
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	7-10
Discussion			
Key results	18	Summarise key results with reference to study objectives	11-12
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	11-12
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-12
Generalisability	21	Discuss the generalisability (external validity) of the study results	12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	13

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Sensitivity analysis for mistakenly adjusting for mediators in estimating total effect from the perspective of causal diagrams

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2016-015640.R1
Article Type:	Research
Date Submitted by the Author:	08-Mar-2017
Complete List of Authors:	Wang, Tingting; Shandong University, Department of Epidemiology and Biostatistics Li, Hongkai; Shandong University, Department of Epidemiology and Biostatistics Su, Ping; Shandong University, Department of Epidemiology and Biostatistics Yu, Yuanyuan; Shandong University, Department of Epidemiology and Biostatistics Sun, Xiaoru; Shandong University, Department of Epidemiology and Biostatistics Liu, Yi; Shandong University, Department of Epidemiology and Biostatistics Yuan, Zhongshang; Shandong University, Department of Epidemiology and Biostatistics Xue, Fuzhong; Shandong University, Department of Epidemiology and Biostatistics
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Public health
Keywords:	EPIDEMIOLGY, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS

SCHOLARONE™
Manuscripts

1 Sensitivity analysis for mistakenly adjusting for mediators in 2 estimating total effect from the perspective of causal diagrams

3 Tingting Wang[†], Hongkai Li[†], Ping Su, Yuanyuan Yu, Xiaoru Sun, Yi Liu, Zhongshang Yuan,
4 Fuzhong Xue*

5 Department of Biostatistics, School of Public Health, Shandong University, Jinan City, Shandong
6 Province, P.R.China

7 Email addresses:

8 TTW: wangtingting161780@163.com

9 HKL: lihongkaiyouxiang@163.com

10 PS: sdsuping@163.com

11 YYY: yu_yy_1993@163.com

12 XRS: sunny150139@126.com

13 YL: liuyi238@sdu.edu.cn

14 ZSY: yuanzhongshang@sdu.edu.cn

15
16 *Corresponding Author

17 xuefzh@sdu.edu.cn (FZX)

18
19 **Word Count: 3033**

20 Abstract

21 **Objectives:** In observational studies, epidemiologists often attempt to estimate the
22 total effect of exposure on outcome of interest. However, when the underlying
23 diagram is unknown and only limited knowledge is available, it is necessary to dissect
24 bias performances in mistakenly adjusting for mediators under logistic regression in
25 estimating the total effect of exposure on outcome. Through simulation, we focus on
26 six causal diagrams concerning different roles of mediators. Sensitivity analysis was
27 conducted to assess the bias performances of varying across the effects of
28 exposure-mediator and mediator-outcome in adjusting for mediator under the
29 framework of logistic regression model.

30 **Setting:** Based on the causal relationships in real world, we compare the bias of
31 varying across the effect of exposure-mediator with the one of varying across the
32 effect of mediator-outcome in adjusting for mediator. The magnitude of the bias was

1 defined by the difference between the estimated effect using logistic regression and
2 the total effect of the exposure on the outcome.

3 **Results:** In the following four scenarios: a single mediator, two series mediators, two
4 independent parallel mediators or two correlated parallel mediators, the bias of
5 varying across the effect of exposure-mediator was greater than the one of varying
6 across the effect mediator-outcome in adjusting for the mediator. While in other two
7 scenarios: a single mediator or two independent parallel mediators in the presence of
8 unobserved confounders, the bias of varying across the effect of exposure-mediator
9 was less than the one of varying across the effect mediator-outcome in adjusting for
10 the mediator.

11 **Conclusions:** The biases were higher sensitive to the variation of effects of
12 exposure-mediator than effects of mediator-outcome in adjusting for mediator in the
13 absence of unobserved confounders; while the biases were higher sensitive to the
14 variation of effects of mediator-outcome than effects of exposure-mediator in the
15 presence of unobserved confounder.

16 **Strengths and limitations of this study**

17 1) For six different causal diagrams, we compared biases of distinct adjustment
18 strategies with and without adjusting for mediators by conducting simulation studies.

19 2) Sensitivity analysis was conducted to assess the performances of varying across the
20 effects of exposure-mediator and mediator-outcome.

21 3) The simulation schemes and parameters were conducted mainly based on real
22 observational studies.

23 4) Combination of theoretical derivation and simulation studies make the results more
24 credible.

1
2
3 1 5) The limitation of simulation studies was under the framework of logistic regression
4
5 2 and only focused on binary variables.
6

7 **Introduction**

8
9
10 4 Estimating the total effects of the exposure (E) on the outcome (D) is a great
11
12 5 challenge in epidemiology studies, because confounders are commonly confused with
13
14 6 mediators.¹⁻³ If confounders and mediators are misclassified, the ability to control
15
16 7 confounder in the estimation of the total effect of the exposure on the outcome is
17
18 8 hampered. Causal diagrams provides a formal conceptual framework to identify and
19
20 9 select confounders,⁴⁻⁵ so that it can avoid falling into analytic pitfalls.⁶ In practice,
21
22 10 even the underlying causal diagrams and the role of covariates (mediator, confounder,
23
24 11 collider and instrumental variable) are not all learned, investigators usually adjusted
25
26 12 for the covariates that are associated with the outcome and exposure.⁷⁻¹⁰ Therefore,
27
28 13 our paper focuses on the bias of varying across the effects of exposure-mediator
29
30 14 ($E \rightarrow M$) and mediator-outcome ($M \rightarrow D$) in mistakenly adjusting for mediators under
31
32 15 logistic regression model.
33
34
35

36
37 16 The causal inference literature has made a considerable contribution to mediation
38
39 17 analysis by providing definitions for direct and indirect effects that allow for the effect
40
41 18 decomposition of a total effect into a direct and an indirect effect.¹¹⁻¹⁹ Arbitrarily
42
43 19 adjusting for a mediator would generally lead to biased estimate of the total effect of
44
45 20 the exposure on the outcome.^{6, 20-21} Nevertheless, in the perspective of causal
46
47 21 diagrams, little attention was paid to the biases in adjusting for mediators under the
48
49 22 logistic regression model in estimating the total effect of E on D . Hence, we focused
50
51 23 on the sensitivity analysis technique to assess the effects $E \rightarrow M$ and $M \rightarrow D$ in
52
53 24 adjusting for mediator.
54
55
56
57
58
59
60

1 In this paper, six typical causal diagrams corresponding to causal correlation are
2 given in Figure 1: a single mediator M (Figure 1a); two series mediators (Figure 1b);
3 two independent parallel mediators (Figure 1c); two correlated parallel mediators
4 (Figure 1d); a single mediator with an unobserved confounder (Figure 1e); two
5 parallel mediators with an unobserved confounder (Figure 1f). The paper aim to
6 explore the sensitivity of bias to the variation of the effects of $E \rightarrow D$ and $M \rightarrow D$ in
7 adjusting for mediator. Hence, both theoretical proofs and quantitative simulations
8 were performed to dissect the bias of varying across the effect of $E \rightarrow M$ and the one of
9 varying across the effect of $M \rightarrow D$ in adjusting for mediators under logistic model.

10 **Methods**

11 A directed acyclic graph (*DAG*) is composed of variables (nodes) and arrows (directed
12 edges) between nodes such that the graph is acyclic. Pearl formalized causal diagrams
13 as directed acyclic graphs (*DAGs*), providing investigators with powerful tools for
14 bias assessment.²² The causal directed acyclic graph theory provides a device for
15 deducing the statistical associations implied by causal relations. Furthermore, given a
16 set of observed statistical associations, a researcher armed with causal diagrams
17 theory can systematically characterize all causal structures compatible with the
18 observations.²³⁻²⁴

19 The total effect can be calculated based on the *do-calculus* and *back-door* criterion
20 proposed by Judea Pearl.²⁵⁻²⁶ For exposure X and outcome Y , a set of variables Z is
21 said to satisfy the backdoor path criterion with respect to (X, Y) if no variable in Z is a
22 descendant of X and if Z blocks all back-door paths from X to Y . Then the effect of X
23 on Y is given by the formula,

$$P(y | do(x)) = \sum_z P(y | x, z)P(z)$$

Note that the expression on the right hand side of the equation is simply a standardized mean. The difference $E(Y | do(x')) - E(Y | do(x''))$ is taken as the definition of “causal effect”, where x' and x'' are two distinct realizations of X .²¹ The interventional distribution, such as that corresponding to $Y(x)$, namely $P(y | do(x))$, is not necessarily equal to a conditional distribution $P(y | x)$. It stands for the probability of $Y = y$ when the exposure X set to level x . The ignorability assumption $Y(x) \perp X$ states that if we happen to have information on the exposure variable, it does not give us any information about the outcome Y after the intervention $do(x)$ was performed. Besides it can be shown that if ignorability holds for $Y(x)$ and X (alternatively if there are no back-door paths from X to Y in the corresponding causal DAGs), then $p(y | do(x)) = p(y | x)$.²⁷⁻²⁸

Let D_e and M_e denote respectively the values of the outcome and mediator that would have been observed had the exposure E been set to level e . On the odds ratio ($OR_{E \rightarrow D}^{TE}$) scale, the total effect ($\beta_{E \rightarrow D}^{TE} = \log(OR_{E \rightarrow D}^{TE})$), comparing exposure level e

with e^* , is given by $OR_{E \rightarrow D}^{TE} = \frac{P(D_e = 1) / \{1 - P(D_e = 1)\}}{P(D_{e^*} = 1) / \{1 - P(D_{e^*} = 1)\}}$.¹⁸⁻¹⁹ While the effect

($\beta_{ED|M}(m)$) of adjusting for mediator M by logistic regression model can be given

$$\begin{aligned} \beta_{ED|M}(m) &= \text{logit}\{P(D = 1 | e = 1, m)\} - \text{logit}\{P(D = 1 | e^* = 0, m)\} \\ &= \log \left\{ \frac{P(D = 1 | e = 1, m)P(D = 0 | e^* = 0, m)}{P(D = 0 | e = 1, m)P(D = 1 | e^* = 0, m)} \right\} \end{aligned}$$

where $P(D = 1 | e, m)$ denotes the probability of $D = 1$ when the exposure E , and mediator M , have been set to level e , and m , respectively. Taking Figure 1a as an example, the logistic regression is

$$\text{logit}\{P(D = 1 | e, m)\} = \alpha_1 + \beta_0 e + \beta_2 m.$$

1 Therefore, the total effect ($\beta_{E \rightarrow D}^{TE}$) of exposure E on outcome D on the scale of
 2 logarithm odds ratio was equal to

$$\begin{aligned} \beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\ &= \log \left\{ \frac{P(D_e = 1) / \{1 - P(D_e = 1)\}}{P(D_{e^*} = 1) / \{1 - P(D_{e^*} = 1)\}} \right\} \\ &= \text{logit} \{P(D_e = 1)\} - \text{logit} \{P(D_{e^*} = 1)\} \\ &= \text{logit} \{P(D = 1 | e = 1)\} - \text{logit} \{P(D = 1 | e^* = 0)\} \\ &= \text{logit} \left\{ \sum_m P(D = 1 | e = 1, m) P(m | e = 1) \right\} - \text{logit} \left\{ \sum_m P(D = 1 | e^* = 0, m) P(m | e^* = 0) \right\} \end{aligned}$$

3 The effect estimation ($\hat{\beta}_{ED|M}(m)$) of adjusting for mediator M by logistic regression
 4 model was equal to

$$\hat{\beta}_{ED|M}(m) = \text{logit} \{ \hat{P}(D = 1 | e = 1, m) \} - \text{logit} \{ \hat{P}(D = 1 | e^* = 0, m) \}$$

5 where $\hat{P}(D = 1 | e = 1, m)$ denotes the probability of $D = 1$ when the exposure E ,
 6 and mediator M , have been set to level $e = 1$, and m , respectively. And
 7 $\hat{P}(D = 1 | e^* = 0, m)$ denotes the probability of $D = 1$ when the exposure E , and
 8 mediator M , have been set to level $e^* = 0$, and m , respectively.

9 Note that the bias was defined by taking a difference between effect estimation by
 10 adjusting for mediator using logistic regression and the total effect of exposure E on
 11 outcome D i.e. $bias = E[\hat{\beta}_{ED|M}(m)] - \beta_{E \rightarrow D}^{TE}$. We dissected the biases behavior by
 12 varying across the effects $E \rightarrow M$ and $M \rightarrow D$ in mistakenly adjusting for mediator
 13 under the framework of logistic regression model.

14 Simulation

15 As shown in Figure 1, six scenarios are designed to dissect bias behaviors of
 16 mistakenly adjusting for mediators using logistic regression model. We made the
 17 following assumptions for the simulation: 1) all variables were binary following a
 18 Bernoulli distribution; 2) the effect from parent nodes to their child node were

1 positive and log-linearly additive. Taking Figure 1a as an example, we randomly
2 generated the exposure following a Bernoulli distribution (i.e. let $P(e = 1) = \pi$), then,
3 $P_M = \exp(\alpha_0 + \beta_1 e) / \{1 + \exp(\alpha_0 + \beta_1 e)\}$ for calculating the distribution probability of
4 child node M from its parent node E . Similarly,
5 $P_D = \exp(\alpha_1 + \beta_0 e + \beta_2 m) / \{1 + \exp(\alpha_1 + \beta_0 e + \beta_2 m)\}$ generated the distribution
6 probability of D , where the parameters α_0 and α_1 denoted the intercept of M and D
7 respectively, and effect parameter $\beta_0, \beta_1, \beta_2$ referred to the effects of the parent node
8 on their corresponding child node using log odds ratio scale.

9 After generating data, we dissected the biases behavior between the effects of
10 $E \rightarrow M$ and $M \rightarrow D$ in mistakenly adjusting for mediator under logistic regression model.
11 In scenario 1 (Figure 1a), we compared the performances by across varying the effects
12 of $E \rightarrow M$ and $M \rightarrow D$. Similarly, in scenario 2 (Figure 1b), the effects of $E \rightarrow M_1$,
13 $M_1 \rightarrow M_2$ and $M_2 \rightarrow D$ were explored. In scenario 3 (Figure 1c), we dissected the effects
14 of $E \rightarrow M_1$ ($E \rightarrow M_2$) and $M_1 \rightarrow D$ ($M_2 \rightarrow D$). The comparison of scenario 4 (Figure 1d)
15 was the same as scenario 3 (Figure 1c). In scenario 5 (Figure 1e), the effects of $E \rightarrow M$
16 and $M \rightarrow D$ were excavated. The scenario 6 (Figure 1f) was identical to the scenario 3.
17 We explored the biases in adjusting for mediator under logistic regression model and
18 thus identified the sensitivity of bias to the variation of the effects of
19 exposure-mediator and mediator-outcome.

20 For each of the 6 simulation scenarios, we observed bias performances of varying
21 across distinct effects in adjusting for mediator using logistic regression model with
22 1000 simulations repetitions. All simulations were conducted using software R from
23 CRAN (<http://cran.r-project.org/>).

1 Results

2 Scenario 1: one single mediator (Figure 1a)

3 In Figure 1(a) of the simplest case, E has a direct ($E \rightarrow D$) effect and an indirect
 4 ($E \rightarrow M \rightarrow D$) effect on D . Figure 2A depicted that the bias of varying across the effect
 5 of $E \rightarrow M$ was obviously greater than the bias of varying across the effect of $M \rightarrow D$.
 6 That is, the sensitivity of bias to the variation of the effect $E \rightarrow M$ was greater than the
 7 effect of $M \rightarrow D$ in adjusting for the mediator M using logistic regression model. In
 8 particular, if the effect of $E \rightarrow M$ was specified to zero in Figure 2B, M became an
 9 independent cause of the outcome, and in this case adjusting for M obtained a positive
 10 bias. Moreover, Figure 2 indicated that adjusting for mediator M was indeed biased to
 11 the total effect of the exposure on the outcome.

12 The total effect ($\beta_{E \rightarrow D}^{TE}$) of exposure E on outcome D on the scale of logarithm odds
 13 ratio was equal to

$$\begin{aligned}
 \beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) = \log \left\{ \frac{P(D_e = 1) / \{1 - P(D_e = 1)\}}{P(D_{e^*} = 1) / \{1 - P(D_{e^*} = 1)\}} \right\} \\
 &= \log \left\{ \frac{P(D_e = 1) \times \{1 - P(D_{e^*} = 1)\}}{\{1 - P(D_e = 1)\} \times P(D_{e^*} = 1)} \right\} \\
 &= \log \left\{ \frac{P(D = 1 | e = 1) \times P(D = 0 | e^* = 0)}{P(D = 0 | e = 1) \times P(D = 1 | e^* = 0)} \right\} \\
 &= \log \left\{ \frac{[\sum_m P(D = 1 | e = 1, m)P(m | e = 1)] \times [\sum_m P(D = 0 | e^* = 0, m)P(m | e^* = 0)]}{[\sum_m P(D = 0 | e = 1, m)P(m | e = 1)] \times [\sum_m P(D = 1 | e^* = 0, m)P(m | e^* = 0)]} \right\}
 \end{aligned}$$

15 The effect ($\beta_{ED|M}(m)$) of adjusting for mediator M by logistic regression model can
 16 be given

$$\begin{aligned}
 \beta_{ED|M}(m) &= \text{logit} \{P(D = 1 | e = 1, m)\} - \text{logit} \{P(D = 1 | e^* = 0, m)\} \\
 &= \log \left\{ \frac{P(D = 1 | e = 1, m) \times \{1 - P(D = 1 | e^* = 0, m)\}}{\{1 - P(D = 1 | e = 1, m)\} \times P(D = 1 | e^* = 0, m)} \right\} \\
 &= \beta_0
 \end{aligned}$$

1 β_0 denotes coefficient of the E adjusting for M using logistic regression model.
 2
 3
 4
 5
 6 Furthermore, the effect of adjusting for M was equal to the controlled direct effect.¹⁹
 7
 8
 9 Therefore, the bias of adjusting for mediator using logistic regression model could be
 10
 11 obtained i.e. $bias = \beta_{ED|M}(m) - \beta_{E \rightarrow D}^{TE}$. We added signs to the edges of the directed
 12
 13 acyclic graph to indicate the presence of a particular positive or negative effect in the
 14
 15 Figure 3. Therefore, we gained $bias < 0$ under the condition of $\beta_1 * \beta_2 > 0$ (the
 16
 17 effect $E \rightarrow M$ β_1 and the effect $M \rightarrow D$ β_2), indicating that the total effect of E on D
 18
 19 was biased in adjusting for M using logistic regression model in Figure 3a, Figure 3b,
 20
 21 Figure 3e & Figure 3f. And the bias was less than zero when the effect $E \rightarrow M$ (β_1)
 22
 23 and the effect $M \rightarrow D$ (β_2) share same signs. (i.e. the effects $E \rightarrow M$ ($\beta_1 > 0$) and $M \rightarrow D$
 24
 25 ($\beta_2 > 0$) were a positive sign or the effects $E \rightarrow M$ ($\beta_1 < 0$) and $M \rightarrow D$ were a negative
 26
 27 sign ($\beta_2 < 0$)). Furthermore, we obtained $bias > 0$, if $\beta_1 * \beta_2 < 0$, suggesting that
 28
 29 the total effect of E on D was biased in adjusting for M in Figure 3c, Figure 3d, Figure
 30
 31 3g & Figure 3h. And the bias was greater than zero when the signs of the effects
 32
 33 $E \rightarrow M$ (β_1) and $M \rightarrow D$ (β_2) were the opposite. The results illustrated that the bias was
 34
 35 less than zero under the case of the effects of exposure-mediator and
 36
 37 mediator-outcome sharing the same sign; the bias was greater than zero under
 38
 39 circumstances of the effects of exposure-mediator and mediator-outcome having
 40
 41 opposite signs. The more details of theoretical derivation have been put in Appendix.
 42
 43
 44
 45
 46
 47
 48
 49

20 **Scenario 2: two series mediators (Figure 1b)**

21 Figure 1(b) is a depiction through two series mediators, decomposing total effects into
 22
 23 direct ($E \rightarrow D$) and indirect ($E \rightarrow M_1 \rightarrow M_2 \rightarrow D$) components. The bias of varying across

1 the effect of $E \rightarrow M_1$ was greater than the one of varying across the effect of $M_2 \rightarrow D$
2 under adjustment for M_1 , adjustment for M_2 and adjustment for $M_1 M_2$ in Figure 4,
3 respectively. In this situation, the correlation of series mediators was strong enough to
4 avoid M_2 from becoming an independent cause of the outcome.

5 **Scenario 3: two independent parallel mediators (Figure 1c)**

6 Figure 1c shows that the exposure E independently causes M_1 and M_2 and indirectly
7 influences the outcome D through M_1 and M_2 , forming three causal paths $E \rightarrow D$,
8 $E \rightarrow M_1 \rightarrow D$ and $E \rightarrow M_2 \rightarrow D$. We obtained that the bias of varying across the effect of
9 $E \rightarrow M_1$ was considerably greater than the one of varying across the effect of $M_1 \rightarrow D$
10 under adjustment for M_1 in Figure 5A. However, the bias of varying across the effect
11 of $E \rightarrow M_2$ was nearly equal to the one with varying across the effect of $M_2 \rightarrow D$ under
12 the identical adjustment for M_1 in Figure 5A. Then, an above similar result can be
13 obtained in Figure 5B. In addition, Figure 5C indicated that biases of varying across
14 the effects of $E \rightarrow M_1$ and $E \rightarrow M_2$ were obviously greater than the one of varying
15 across the effects of $M_1 \rightarrow D$ and $M_2 \rightarrow D$ under adjustment for $M_1 M_2$.

16 **Scenario 4: two correlated parallel mediators (Figure 1d)**

17 In Figure 1d, there exist five paths from E to D : $E \rightarrow D$, $E \rightarrow M_1 \rightarrow D$, $E \rightarrow M_2 \rightarrow D$,
18 $E \rightarrow M_1 \leftarrow M_2 \rightarrow D$ and $E \rightarrow M_2 \rightarrow M_1 \rightarrow D$. In particular, the path $E \rightarrow M_1 \leftarrow M_2 \rightarrow D$ is a
19 blocked path, due to the M_1 being a collider node. Figure 6A indicated that the bias of
20 varying across the effect of $E \rightarrow M_1$ was obviously greater than the one of varying
21 across the effect of $M_1 \rightarrow D$ under the adjustment for M_1 . However, the bias of varying
22 across the effect of $E \rightarrow M_2$ was almost equal to the one of varying across the effect of
23 $M_2 \rightarrow D$ under the identical adjustment model. Similarly, Figure 6B showed an

1
2
3 1 analogous result of biases behavior. Besides, Figure 6C manifested that biases of
4
5 2 varying across the effects of $E \rightarrow M_1$ and $E \rightarrow M_2$ were greater than the ones of varying
6
7 3 across the effects of $M_1 \rightarrow D$ and $M_2 \rightarrow D$ in adjusting for M_1 and M_2 . Simultaneously,
8
9 4 the bias was higher sensitive to the variation of effect of $E \rightarrow M_2$ than effects of $E \rightarrow M_1$
10
11 5 under the identical model, which adjustment for the collider node M_1 would partially
12
13 6 open the path $E \rightarrow M_1 \leftarrow M_2 \rightarrow D$.

17 ***Scenario 5: a single mediator with an unobserved confounder (Figure 1e)***

18
19 8 Figure 1e provides a causal diagram representing the relationship among exposure E ,
20
21 9 outcome D , mediator M and unobserved confounder U . It revealed that the bias of
22
23 10 varying across the effect of $E \rightarrow M$ was lower than the one of varying across the effect
24
25 11 of $M \rightarrow D$. Unobserved confounder distorts the association between the exposure and
26
27 12 outcome ($E \leftarrow U \rightarrow D$) in Figure 7.

32 ***Scenario 6: two parallel mediators with an unobserved confounder (Figure 1f)***

33
34 14 As described above, Figure 1f is a depiction of two parallel mediators M_1 and M_2 with
35
36 15 an unobserved confounder U . Figure 8A indicated that the bias of varying across the
37
38 16 effect of $E \rightarrow M_1$ was obviously less than the one of varying across the effect of $M_1 \rightarrow D$
39
40 17 under the adjustment for M_1 . However, the bias of varying across the effect of $E \rightarrow M_2$
41
42 18 was greater than the bias of varying across the effect of $M_2 \rightarrow D$ under the identical
43
44 19 model of adjusting for M_1 . A similar result can also obtain in Figure 8B. Besides,
45
46 20 biases of varying across the effects of $E \rightarrow M_1$ and $E \rightarrow M_2$ were distinctly less than
47
48 21 the ones of varying across the effects of $M_1 \rightarrow D$ and $M_2 \rightarrow D$ under the common model
49
50 22 of adjusting for M_1 and M_2 in Figure 8C.

57 **Discussion**

1
2
3 1 In the paper, we dissected the sensitivity of bias to the variation of the effects of
4 exposure-mediator and mediator-outcome in adjusting for mediators under the
5 framework of logistic regression model. In the following four scenarios: a single
6 mediator (Figure 1a in scenario 1), two series mediators (Figure 1b in scenario 2), two
7 independent parallel (Figure 1c in scenario 3) or two correlated parallel mediators
8 (Figure 1d in scenario 4), the bias of varying across the effect of exposure-mediator
9 was greater than the one of varying across the effect mediator-outcome in adjusting
10 for the mediator (Figure 2, Figure 4, Figure 5 & Figure 6). However, in other two
11 scenarios: a single mediator or two independent parallel mediators in the presence of
12 unobserved confounders (Figure 1e in scenario 5 & Figure 1f in scenario 6), the
13 biases were higher sensitive to the variation of effect of mediator-outcome than effects
14 of exposure-mediator in adjusting for mediator (Figure 7 & Figure 8).

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

13 Conditioning on a mediator is of concern in all areas of epidemiologic
14 researches,^{11,17,29} it indeed led to bias in estimating the total effect of the exposure on
15 the outcome.^{6, 20-21} Mediators and confounders were indistinguishable in terms of
16 statistical association and conceptual grounds.³ Most of the studies focused on the
17 mediation effect analysis such as the calculation of direct effects and indirect
18 effect.^{18-19,30-33} Little effort has been made to learn the biases performances in
19 adjusting for mediator in estimating the total effect of exposure on outcome. Our
20 study results revealed that the biases were higher sensitive to the variation of effect of
21 exposure-mediator than effects of mediator-outcome in adjusting for mediator in the
22 absence of the unobserved confounder in causal diagrams (Figure 1a, Figure 1b,
23 Figure 1c & Figure 1d). Nevertheless, for causal diagrams (Figure 1e & Figure 1f),
24 the biases were higher sensitive to the variation of effect of mediator-outcome than
25 effects of exposure-mediator in adjusting for mediator in the presence of the

1
2
3 1 unobserved confounder. Therefore, the biases of varying across different effects
4
5 2 depended on the causal diagrams framework whether there existed unobserved
6
7 3 confounder.
8
9

10 4 We need note that, our simulation study was not comprehensive enough to evaluate
11
12 5 the bias performances in adjusting for the mediator under logistic regression, because
13
14 6 it only considered binary variables, the certain scenarios of effect size and the
15
16 7 common type of models. The work in the further ought to reinforce the mechanisms
17
18 8 and conceptual frameworks of confounder and mediator form causal diagrams so as to
19
20 9 avoid falling into analytic pitfalls.
21
22
23

24 25 **Conclusion** 26

27 11 In conclusion, we showed that the sensitivity of bias to the variation of the effects of
28
29 12 exposure-mediator and mediator-outcome was related to whether there is an
30
31 13 unobserved confounder in causal diagrams. The biases were higher sensitive to the
32
33 14 variation of effects of exposure-mediator than effects of mediator-outcome in
34
35 15 adjusting for mediator in the absence of unobserved confounders; while the biases
36
37 16 were higher sensitive to the variation of effects of mediator-outcome than effects of
38
39 17 exposure-mediator in the presence of unobserved confounder.
40
41
42

43 **Statements** 44

45 **Ethics approval and materials** 46

47 19 Not applicable
48
49

50 **Competing interests** 51

52 21 The authors declare that they have no competing interests.
53
54

55 **Funding** 56

57 24 This work was supported by grants from the National Natural Science Foundation of
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 China (grant number 81573259)

2 **Provenance and peer review**

3 Not commissioned; externally peer reviewed.

4 **Data sharing statement**

5 No additional data are available.

6 **Authors' contributions**

7 TTW and HKL jointly conceived the idea behind the article and designed the study.

8 TTW helped conduct the literature review, performed the simulation and prepared the

9 first draft of the manuscript. PS, YYY, XRS, YL and ZSY participated in the design of

10 the study and the revision of the manuscript. FZX advised on critical revision of the

11 manuscript for important intellectual content. All authors read and approved the final

12 manuscript.

13 **Acknowledgments**

14 We would like to thank the anonymous reviewers and academic editors for providing

15 us with constructive comments and suggestions and also wish to acknowledge our

16 colleagues for their invaluable work.

17 **Open Access**

18 This is an Open Access article distributed in accordance with the Creative Commons

19 Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to

20 distribute, remix, adapt, build upon this work noncommercially, and license their

21 derivative works on different terms, provided the original work is properly cited and

22 the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

23 **References**

24 1 Bulterys M, Morgenstern H. Confounding or intermediate effect? An appraisal of
25 iatrogenic bias in perinatal AIDS research. *Paediatr Perinat Epidemiol*
26 1993;7:387–394.

27 2 Herna'n MA, Hernandez-Diaz S, Werler MM, et al. Causal knowledge as a
28 prerequisite for confounding evaluation: an application to birth defects
29 epidemiology. *Am J Epidemiol* 2002;155:176–184.

30 3 MacKinnon DP, Krull JL, and Lockwood CM. Equivalence of the mediation,

- 1
2
3 1 confounding and suppression effect. *Prevention Science* 2000;1:173–181.
- 4 2 4 Howards PP, Schisterman EF, Poole C, et al. “Toward a clearer definition of
5 3 confounding” revisited with directed acyclic graphs, *Am J Epidemiol*
6 4 2012;176:506–511.
- 7 5 5 Williamson EJ, Aitken Z, Lawrie J, et al. Introduction to causal diagrams for
8 6 confounder selection. *Respirology*. 2014;19:303-311.
- 9 7 6 Wilcox AJ, Weinberg CR, Basso O. On the pitfalls of adjusting for gestational
10 8 age at birth. *Am J Epidemiol*. 2011;174:1062–1068.
- 11 9 7 Psaty BM, Koepsell TD, Lin D, et al. Assessment and control for confounding by
12 10 indication in observational studies. *J Am Geriatr Soc* 1999; 47:749-54.
- 13 11 8 Tilaki K H. Methodological issues of confounding in analytical epidemiology
14 12 studies. *Caspian J intern Med* 2012;3:488-495.
- 15 13 9 Kamangar F MD. Confounding variable in epidemiology studies: Basics and
16 14 beyond. *Arch Tran Med*. 2012;15:508-516.
- 17 15 10 Gaskins AJ, Schisterman EF. The effect of lipid adjustment on the analysis of
18 16 environmental contaminants and the outcome of human health risks. *Methods*
19 17 *Mol Biol* 2009;580:371-381.
- 20 18 11 Robins JM, Greenland S. Identifiability and exchangeability for direct and
21 19 indirect effects. *Epidemiology* 1992;3:143–155.
- 22 20 12 Pearl J. Direct and indirect effects. Seventeenth conference of uncertainty in
23 21 artificial intelligence, 2001. San Francisco, CA: Morgan Kaufmann, 2001.
- 24 22 13 Huang B, Sivaganesan S, Succop P, et al. Statistical assessment of mediational
25 23 effects for logistic mediational models. *Statist. Med* 2004;23:2713-2728.
- 26 24 14 Valeri L, Coull B A. Estimating causal contrasts involving intermediate variables
27 25 in the presence of selection bias [J]. *Statistics in Medicine*, 2016, 35(26):
28 26 4779-4793.
- 29 27 15 MacKinnon D P, Pirlott A G. Statistical approaches for enhancing causal
30 28 interpretation of the M to Y relation in mediation analysis [J]. *Personality and*
31 29 *Social Psychology Review*, 2015, 19(1): 30-43.
- 32 30 16 Imai K, Keele L, Tingley D. A general approach to causal mediation analysis.
33 31 *Psychol Methods*. 2010;15(4):309–334.
- 34 32 17 VanderWeele TJ. Bias formulas for sensitivity analysis for direct and indirect
35 33 effects. *Epidemiology*. 2010;21(4):540–551.
- 36 34 18 VanderWeele TJ, Vansteelandt S. Odds ratios for mediation analysis for a
37 35 dichotomous outcome. *Am J Epidemiol*. 2010;172(12):1339–1348.
- 38 36 19 Valeri L, VanderWeele TJ. Mediation analysis allowing for exposure-mediator
39 37 interactions and causal interpretation: theoretical assumptions and
40 38 implementation with SAS and SPSS macros [J]. *Psychological methods*, 2013,
41 39 18(2): 137.
- 42 40 20 Robins J. The control of confounding by intermediate variable. *Statistics in*
43 41 *Medicine*. 1989;8:679-701.
- 44 42 21 Schisterman EF, Cole SR, Platt RW. Overadjustment Bias and Unnecessary
45 43 Adjustment in Epidemiologic Studies. *Epidemiology*. 2009;20:488–495.
- 46 44 22 Pearl J. Causal diagrams for empirical research. *Biometrika*. 1995;82:669-688.

- 1
2
3 1 23 Pearl J. Causality: models, reasoning, and inference. *Econometric Theory*.
4 2 2003;19: 675-685.
5 3 24 Joffe M, Gambhir M, Chadeau-Hyam M, et al. Causal diagrams in systems
6 4 epidemiology. *Emerg Themes Epidemiol*. 2012; 9:1-18.
7 5 25 Pearl, J. Causality: models, reasoning, and inference. Cambridge: Cambridge
8 6 University Press; 2000.
9 7 26 Pearl J. Causal inference in statistics: an overview. *Statistics Surveys*. 2009;
10 8 3:96–146.
11 9 27 Rubin DB. Should observational studies be designed to allow lack of balance in
12 10 covariate distributions across treatment groups? *Statistics in Medicine*.
13 11 2009;28:1420-1423.
14 12 28 Shpitser I. Counterfactual graphical models for longitudinal mediation analysis
15 13 with unobserved confounding. *Cognitive Science* 2013;37:1011-1035.
16 14 29 VanderWeele TJ, Mumford SL, Schisterman EF. Conditioning on intermediates
17 15 in perinatal epidemiology. *Epidemiology*. 2012;23(1):1–9.
18 16 30 Lange T, Rasmussen M, Thygesen LC. Assessing natural direct and indirect
19 17 effects through multiple pathways. *American Journal of Epidemiology*. 2013;179,
20 18 513–518.
21 19 31 Shpitser, Ilya and VanderWeele, Tyler J. "A complete graphical criterion for the
22 20 adjustment formula in mediation analysis," *The International Journal of*
23 21 *Biostatistics*.2011;7,1-24.
24 22 32 Albert J M, Wang W. Sensitivity analyses for parametric causal mediation effect
25 23 estimation [J]. *Biostatistics*, 2015, 16(2): 339-351.
26 24 33 Wang W, Zhang B. Assessing natural direct and indirect effects for a continuous
27 25 exposure and a dichotomous outcome[J]. *Journal of Statistical Theory and*
28 26 *Practice*, 2016, 10(3): 574-587.
29 27

28 **Figure 1:** Six causal diagrams were designed for estimating the causal effect of E on
29 D . a) a single mediator M ; b) two series mediators M_1 and M_2 ; c) two independent
30 parallel mediators M_1 and M_2 ; d) two correlated parallel mediators M_1 and M_2 ; e) a
31 single mediator with an unobserved confounder U ; f) two independent parallel
32 mediators M_1 and M_2 with an unobserved confounder U .
33

34 **Figure 2:** The biases with the effects $E \rightarrow M$ (red) and $M \rightarrow D$ (blue) increasing,
35 respectively. Comparison of the bias of different effects in adjustment mediator. The
36 OR of target effect (e.g. $E \rightarrow M$) from 1 to 10 given other effects fixed $\ln 2$ in Figure
37 2A. The OR of the effect $M \rightarrow D$ from 1 to 10 with the effect $E \rightarrow M$ being equal to zero
38 in Figure 2B (Color figure online).
39

40 **Figure 3:** Illustrating the use of positive and negative signs on edges $E \rightarrow M$, $M \rightarrow D$
41 and $E \rightarrow D$.
42

43 **Figure 4:** The biases with the effects $E \rightarrow M_1$ (red), $M_1 \rightarrow M_2$ (blue) and $M_2 \rightarrow D$ (black)
44 increasing, respectively. Comparison of the bias of different effects in three

1
2
3 adjustment models: A) adjustment for M_1 , B) adjustment for M_2 and C) adjustment for
4 M_1 and M_2 . The OR of target effect (e.g. $E \rightarrow M_1$) from 1 to 10 given the effect $M_1 \rightarrow$
5 M_2 fixed ln8 and other effects fixed ln2 in Figure 4 (Color figure online).
6
7

8 **Figure 5:** The biases with the effects $E \rightarrow M_1$ (red), $E \rightarrow M_2$ (blue), $M_1 \rightarrow D$ (black) and
9 $M_2 \rightarrow D$ (green) increasing, respectively. Comparison of the bias of different effects in
10 three adjustment models: A) adjustment for M_1 , B) adjustment for M_2 and C)
11 adjustment for M_1 and M_2 . The OR of target effects (e.g. $E \rightarrow M_1$) from 1 to 10 given
12 other edges effects fixed ln2 in Figure 5 (Color figure online).
13
14

15 **Figure 6:** The biases with the effects $E \rightarrow M_1$ (red), $E \rightarrow M_2$ (blue), $M_1 \rightarrow D$ (black), M_2
16 $\rightarrow D$ (green) and the effect $M_2 \rightarrow M_1$ (purple) increasing, respectively. Comparison of
17 the bias of different effects in three adjustment models: A) adjustment for M_1 , B)
18 adjustment for M_2 and C) adjustment for M_1 and M_2 . The OR of target effects (e.g. E
19 $\rightarrow M_1$) from 1 to 10 given other effects fixed ln2 in Figure 6 (Color figure online).
20
21

22 **Figure 7:** The biases with the effects $E \rightarrow M$ (red) and $M \rightarrow D$ (blue) respectively.
23 Comparison of the bias of different effects in adjustment mediator M . The OR of
24 target effects (e.g. $E \rightarrow M$) from 1 to 10 given the effects of causal edges fixed ln2 and
25 the effect of confounder edges fixed ln5 in Figure ln8 (Color figure online).
26
27

28 **Figure 8:** The biases with the effects $E \rightarrow M_1$ (red), $E \rightarrow M_2$ (blue), $M_1 \rightarrow D$ (black) and
29 $M_2 \rightarrow D$ (green) respectively. Comparison of the bias of different effects in three
30 adjustment models: A) adjustment for M_1 , B) adjustment for M_2 , and C) adjustment
31 for M_1 and M_2 . The OR of target effects (e.g. $E \rightarrow M_1$) from 1 to 10 given the effects of
32 causal edges fixed ln2 and the effect of confounder edges fixed ln5 in Figure 8 (Color
33 figure online).
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

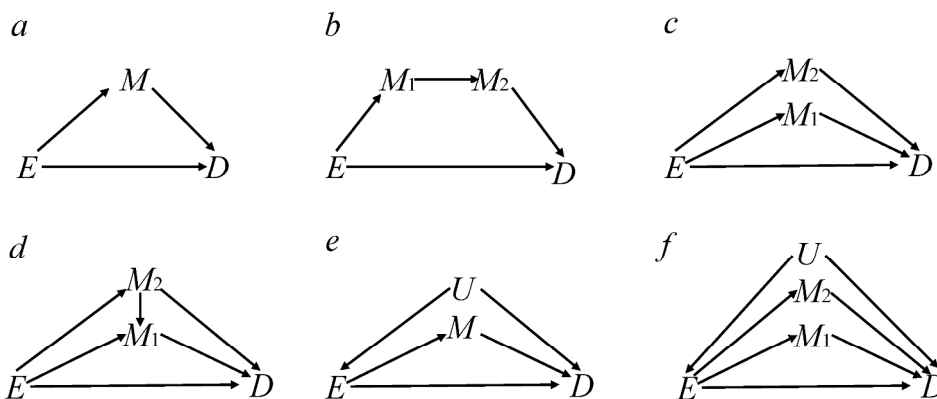


Figure 1: Six causal diagrams were designed for estimating the causal effect of E on D .

252x110mm (300 x 300 DPI)

er review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

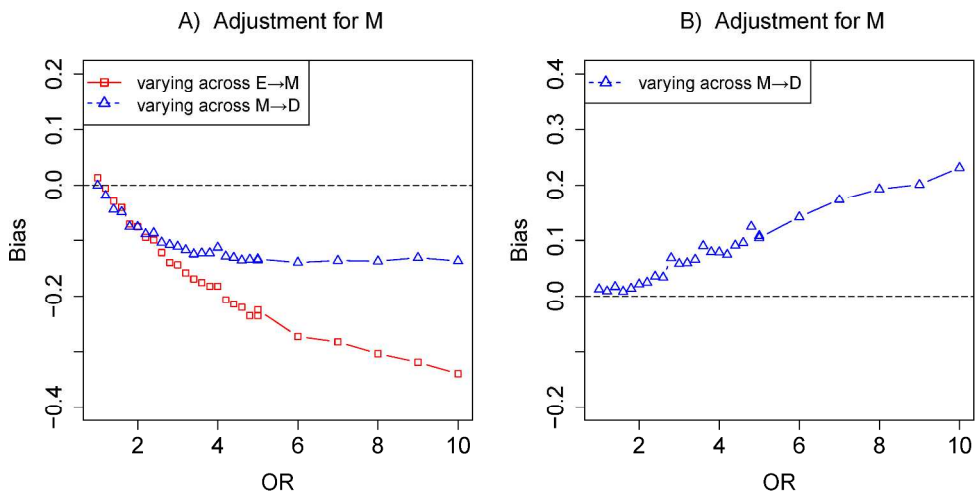


Figure 2 : The biases with the effects E→M (red) and M→D (blue) increasing, respectively.

281x148mm (300 x 300 DPI)

review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

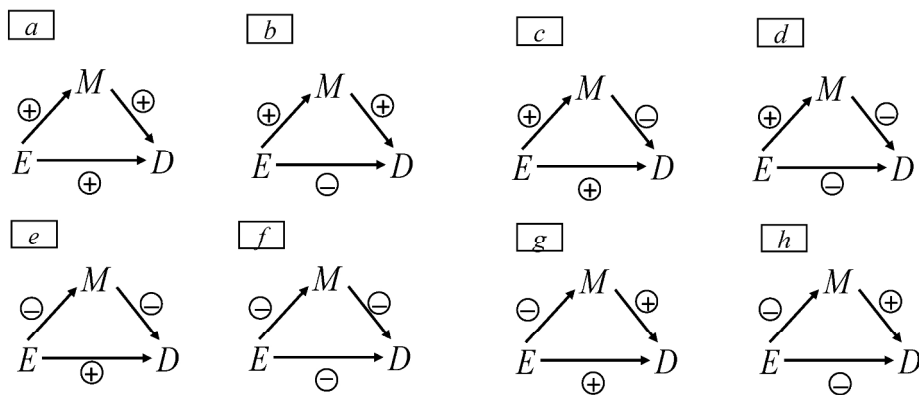


Figure 3: Illustrating the use of positive and negative signs on edges E→M, M→D and E→D.

237x106mm (300 x 300 DPI)

er review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

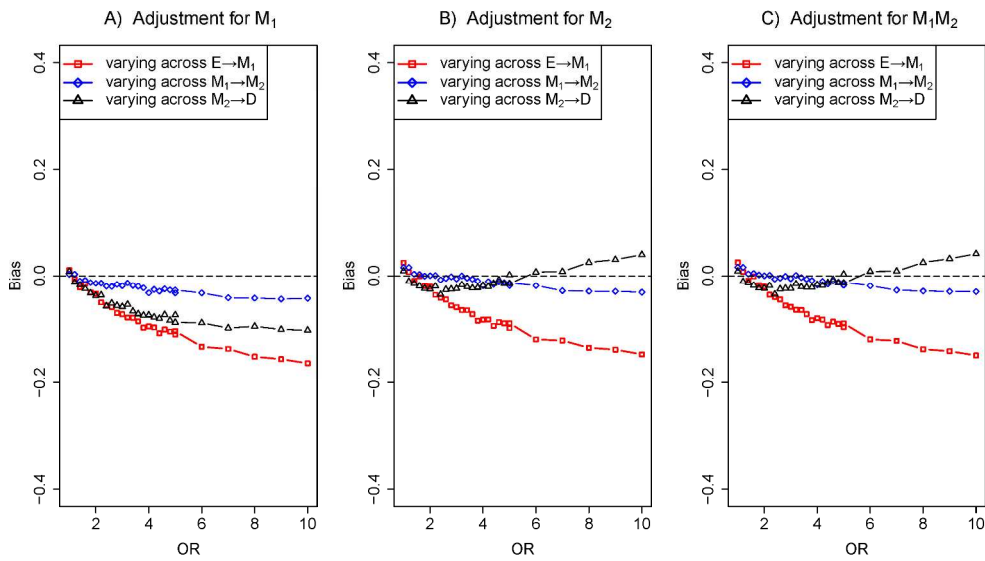


Figure 4: The biases with the effects E→M₁ (red), M₁→M₂ (blue) and M₂→D (black) increasing, respectively.

270x155mm (300 x 300 DPI)

Review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

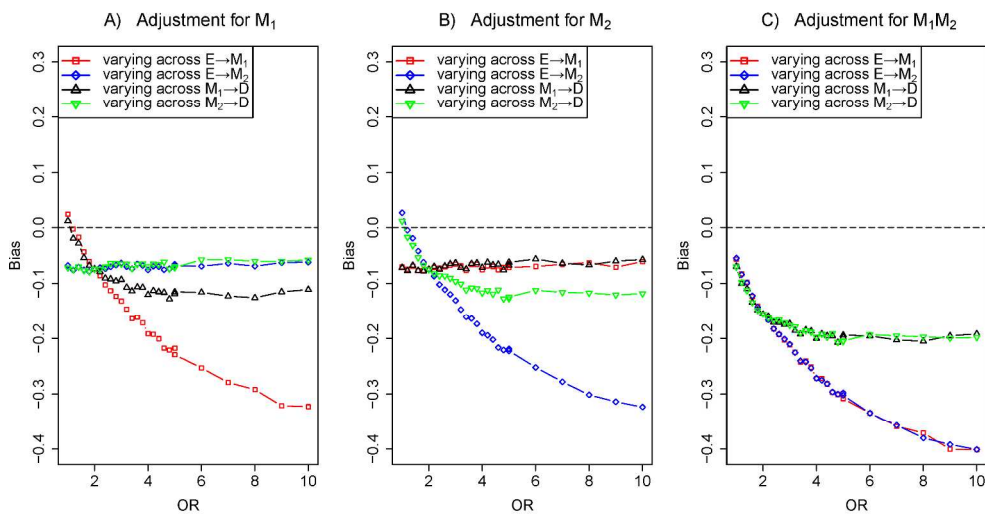


Figure 5 : The biases with the effects $E \rightarrow M_1$ (red), $E \rightarrow M_2$ (blue), $M_1 \rightarrow D$ (black) and $M_2 \rightarrow D$ (green) increasing, respectively.

279x147mm (300 x 300 DPI)

review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

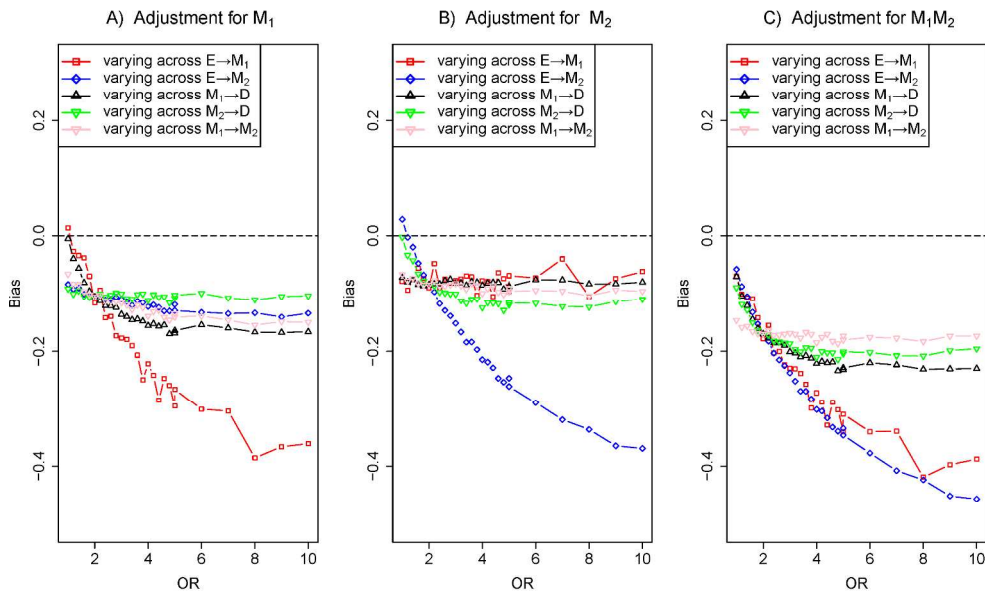


Figure 6: The biases with the effects $E \rightarrow M_1$ (red), $E \rightarrow M_2$ (blue), $M_1 \rightarrow D$ (black), $M_2 \rightarrow D$ (green) and the effect $M_2 \rightarrow M_1$ (purple) increasing, respectively.

278x169mm (300 x 300 DPI)

Review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

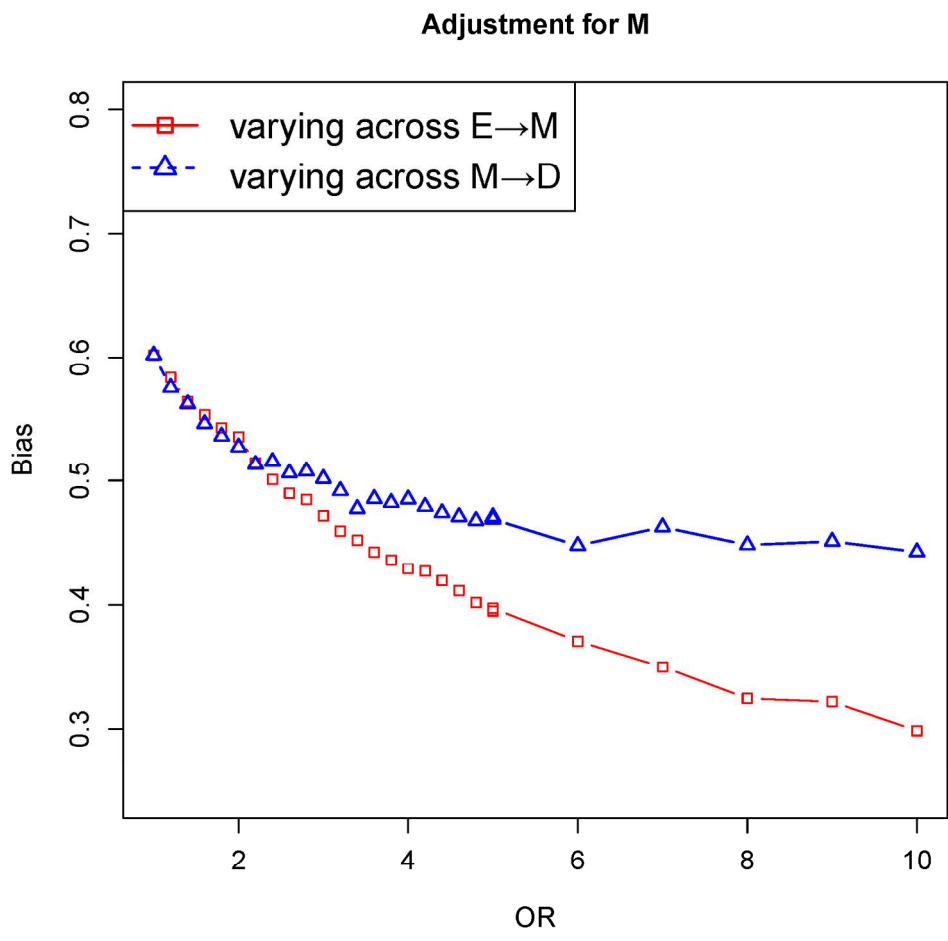


Figure 7: The biases with the effects E→M (red) and M→D (blue) respectively.

177x177mm (300 x 300 DPI)

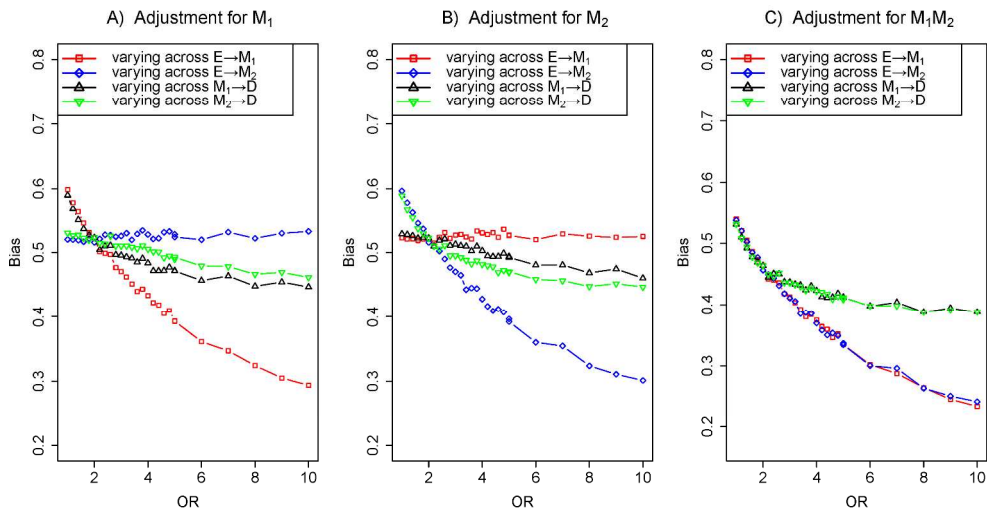


Figure 8 : The biases with the effects E→M₁ (red), E→M₂ (blue), M₁→D (black) and M₂→D (green) respectively.

281x148mm (300 x 300 DPI)

review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Appendix:

The effect of adjusting for mediator was biased for estimating the total effect of exposure on outcome using logistic regression model. Theoretical derivation of Figure 1a as follow:

Suppose the logistic models among E , M and D are:

$$\text{logit}\{P(D=1|e,m)\} = \alpha_1 + \beta_0 e + \beta_2 m,$$

$$\text{logit}\{P(M=1|e)\} = \alpha_0 + \beta_1 e.$$

The total effect ($\beta_{E \rightarrow D}^{TE}$) of exposure E on outcome D on the odds ratio ($OR_{E \rightarrow D}^{TE}$) scale was equal to

$$\begin{aligned} \beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\ &= \log \left\{ \frac{P(D_e = 1) / \{1 - P(D_e = 1)\}}{P(D_{e^*} = 1) / \{1 - P(D_{e^*} = 1)\}} \right\} \\ &= \log \left\{ \frac{P(D_e = 1) \times \{1 - P(D_{e^*} = 1)\}}{\{1 - P(D_e = 1)\} \times P(D_{e^*} = 1)} \right\} \\ &= \log \left\{ \frac{P(D=1|e=1) \times P(D=0|e^*=0)}{P(D=0|e=1) \times P(D=1|e^*=0)} \right\} \\ &= \log \left\{ \frac{\left[\sum_m P(D=1|e=1,m)P(m|e=1) \right] \times \left[\sum_m P(D=0|e^*=0,m)P(m|e^*=0) \right]}{\left[\sum_m P(D=0|e=1,m)P(m|e=1) \right] \times \left[\sum_m P(D=1|e^*=0,m)P(m|e^*=0) \right]} \right\} \end{aligned}$$

The effect ($\beta_{ED|M}(m)$) of adjusting for mediator M by logistic regression model is given

$$\begin{aligned} \beta_{ED|M}(m) &= \text{logit}\{P(D=1|e=1,m)\} - \text{logit}\{P(D=1|e^*=0,m)\} \\ &= \log \left\{ \frac{P(D=1|e=1,m) \times P(D=0|e^*=0,m)}{P(D=0|e=1,m) \times P(D=1|e^*=0,m)} \right\} \\ &= \beta_0 \end{aligned}$$

Therefore,

$$\begin{aligned}
 \text{bias} &= \beta_0 - \log(OR_{E \rightarrow D}^{TE}) \\
 &= \log \left\{ \frac{\exp(\beta_0)}{\exp(\beta_0) \frac{\exp(\beta_2) \times A_1 + \exp(\beta_2) \times B_1 + C_1 + D_1}{\exp(\beta_2) \times A_1 + B_1 + \exp(\beta_2) \times C_1 + D_1}} \right\} \\
 &= \log \left\{ \frac{\exp(\beta_2) \times A_1 + B_1 + \exp(\beta_2) \times C_1 + D_1}{\exp(\beta_2) \times A_1 + \exp(\beta_2) \times B_1 + C_1 + D_1} \right\}
 \end{aligned}$$

where

$$A_1 = \exp(\beta_1 + \alpha_0) \times (1 + \exp(\beta_0 + \alpha_1)) \times \exp(\alpha_0) \times (1 + \exp(\alpha_1))$$

$$B_1 = \exp(\beta_1 + \alpha_0) \times (1 + \exp(\beta_0 + \alpha_1)) \times (1 + \exp(\beta_2 + \alpha_1))$$

$$C_1 = (1 + \exp(\beta_0 + \beta_2 + \alpha_1)) \times \exp(\alpha_0) \times (1 + \exp(\alpha_1))$$

$$D_1 = (1 + \exp(\beta_0 + \beta_2 + \alpha_1)) \times (1 + \exp(\beta_2 + \alpha_1))$$

Focusing on the difference of between $\exp(\beta_2) \times B_1 + C_1$ and $B_1 + \exp(\beta_2) \times C_1$.

$$\begin{aligned}
 T(\beta_1) &= \exp(\beta_2) \times B_1 + C_1 - (B_1 + \exp(\beta_2) \times C_1) \\
 &= \exp(\beta_2) \times (B_1 - C_1) - (B_1 - C_1) \\
 &= (\exp(\beta_2) - 1) \times (B_1 - C_1) \\
 &= (\exp(\beta_2) - 1) \times (\exp(\beta_1 + \alpha_0) \times (1 + \exp(\beta_0 + \alpha_1)) \times (1 + \exp(\beta_2 + \alpha_1)) \\
 &\quad - (1 + \exp(\beta_0 + \beta_2 + \alpha_1)) \times \exp(\alpha_0) \times (1 + \exp(\alpha_1))) \\
 &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \times [\exp(\beta_1) \times (1 + \exp(\beta_0 + \alpha_1)) \times (1 + \exp(\beta_2 + \alpha_1)) \\
 &\quad - (1 + \exp(\beta_0 + \beta_2 + \alpha_1)) \times (1 + \exp(\alpha_1))]
 \end{aligned}$$

Then, detailed dissection:

$$1: \beta_2 = 0, \text{bias} = 0.$$

$$2: \beta_2 > 0,$$

$$\textcircled{1} \beta_1 = 0: \text{(i) } \beta_0 = 0, \text{bias} = 0; \text{(ii) } \beta_0 > 0, \text{bias} > 0; \text{(iii) } \beta_0 < 0, \text{bias} < 0.$$

$$\textcircled{2} \beta_1 < 0: \text{(i) } \beta_0 = 0, \text{bias} > 0; \text{(ii) } \beta_0 > 0, \text{bias} > 0; \text{(iii) } \beta_0 < 0, \text{bias} > 0.$$

proof (iii)

$$\begin{aligned}
 T(\beta_1) &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\
 &\quad \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\
 &\quad - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \}
 \end{aligned}$$

$$\text{when } \beta_0 < 0 \text{ and } \beta_2 > 0 \Rightarrow \exp(\beta_0) - 1 < 0 \quad \exp(\beta_2) - 1 > 0$$

According to $(a-1)(b-1) = ab - a - b + 1$, when $(a-1)(b-1) < 0 \Rightarrow ab + 1 < a + b$

$$\begin{aligned}
 & 1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\
 & < 1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\
 & \Rightarrow \exp(\beta_0 + \beta_2) + 1 < \exp(\beta_0) + \exp(\beta_2)
 \end{aligned}$$

when

$$\beta_1 < \log \left\{ \frac{\exp(\beta_0 + \beta_2) + 1}{\exp(\beta_0) + \exp(\beta_2)} \right\} < 0$$

$$\beta_1 < \log \left\{ \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} \right\} < 0$$

$$\Rightarrow \exp(\beta_1) < \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} < 1$$

$$\begin{aligned}
 \Rightarrow T(\beta_1) &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\
 &\quad \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\
 &\quad - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \} \\
 &< 0
 \end{aligned}$$

Therefore, when $\beta_2 > 0$, $\beta_1 < 0$, $\beta_0 < 0$, then $bias > 0$.

③ $\beta_1 > 0$: (i) $\beta_0 = 0$, $bias < 0$; (ii) $\beta_0 < 0$, $bias < 0$; (iii) $\beta_0 > 0$, $bias < 0$.

proof (iii)

$$\begin{aligned}
 T(\beta_1) &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\
 &\quad \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\
 &\quad - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \}
 \end{aligned}$$

when $\beta_0 > 0$ and $\beta_2 > 0 \Rightarrow \exp(\beta_0) - 1 > 0 \quad \exp(\beta_2) - 1 > 0$

According to $(a-1)(b-1) = ab - a - b + 1$, when $ab > 0 \Rightarrow ab + 1 > a + b$

$$\begin{aligned}
 & 1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\
 & > 1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\
 & \Rightarrow \exp(\beta_0 + \beta_2) + 1 > \exp(\beta_0) + \exp(\beta_2)
 \end{aligned}$$

when

$$\beta_1 > \log \left\{ \frac{\exp(\beta_0 + \beta_2) + 1}{\exp(\beta_0) + \exp(\beta_2)} \right\} > 0$$

$$\beta_1 > \log \left\{ \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} \right\} > 0$$

$$\Rightarrow \exp(\beta_1) > \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} > 1$$

$$\begin{aligned} \Rightarrow T(\beta_1) &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\ &\quad \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\ &\quad - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \} \\ &> 0 \end{aligned}$$

Therefore, when $\beta_2 > 0$, $\beta_1 > 0$, $\beta_0 > 0$, then $bias < 0$.

3: $\beta_2 < 0$,

① $\beta_1 = 0$: (i) $\beta_0 = 0$, $bias = 0$; (ii) $\beta_0 > 0$, $bias > 0$; (iii) $\beta_0 < 0$, $bias < 0$.

② $\beta_1 < 0$: (i) $\beta_0 = 0$, $bias < 0$; (ii) $\beta_0 < 0$, $bias < 0$; (iii) $\beta_0 > 0$, $bias < 0$.

proof (iii)

$$\begin{aligned} T(\beta_1) &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\ &\quad \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\ &\quad - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \} \end{aligned}$$

when $\beta_0 > 0$ and $\beta_2 < 0 \Rightarrow \exp(\beta_0) - 1 > 0$ $\exp(\beta_2) - 1 < 0$

According to $(a-1)(b-1) = ab - a - b + 1$, when $ab < 0 \Rightarrow ab + 1 < a + b$

$$\begin{aligned} 1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\ < 1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\ \Rightarrow \exp(\beta_0 + \beta_2) + 1 < \exp(\beta_0) + \exp(\beta_2) \end{aligned}$$

when

$$\beta_1 < \log \left\{ \frac{\exp(\beta_0 + \beta_2) + 1}{\exp(\beta_0) + \exp(\beta_2)} \right\} < 0$$

$$\beta_1 < \log \left\{ \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} \right\} < 0$$

$$\Rightarrow \exp(\beta_1) < \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} < 1$$

$$\begin{aligned} \Rightarrow T(\beta_1) &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\ &\quad \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\ &\quad - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \} \\ &> 0 \end{aligned}$$

Therefore, when $\beta_2 < 0$, $\beta_1 < 0$, $\beta_0 > 0$, then $bias < 0$.

③ $\beta_1 > 0$: (i) $\beta_0 = 0$, $bias > 0$; (ii) $\beta_0 > 0$, $bias > 0$; (iii) $\beta_0 < 0$, $bias > 0$.

proof (iii)

$$T(\beta_1) = (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\ \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\ - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \}$$

when $\beta_0 < 0$ and $\beta_2 < 0 \Rightarrow \exp(\beta_0) - 1 < 0 \quad \exp(\beta_2) - 1 < 0$

According to $(a-1)(b-1) = ab - a - b + 1$, when $ab > 0 \Rightarrow ab + 1 > a + b$

$$1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\ > 1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\ \Rightarrow \exp(\beta_0 + \beta_2) + 1 > \exp(\beta_0) + \exp(\beta_2)$$

when

$$\beta_1 > \log \left\{ \frac{\exp(\beta_0 + \beta_2) + 1}{\exp(\beta_0) + \exp(\beta_2)} \right\} > 0$$

$$\beta_1 > \log \left\{ \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} \right\} > 0$$

$$\Rightarrow \exp(\beta_1) > \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} > 1$$

$$\Rightarrow T(\beta_1) = (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\ \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\ - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \} \\ < 0$$

Therefore, when $\beta_2 < 0$, $\beta_1 > 0$, $\beta_0 < 0$, then $bias > 0$.

In conclusion:

1: $\beta_2 = 0$, $bias = 0$.

2: $\beta_2 \neq 0$, $\beta_1 = 0$: (i) $\beta_0 = 0$, $bias = 0$; (ii) $\beta_0 > 0$, $bias > 0$; (iii) $\beta_0 < 0$, $bias < 0$.

3: (i) $\beta_1\beta_2 > 0$, $bias < 0$. (ii) $\beta_1\beta_2 < 0$, $bias > 0$.

Supplementary A

The theoretical results of others causal diagrams (Figure 1b-Figure 1f) have been shown in the supplementary of manuscript.

(1) Figure 1(b) is a depiction through two series mediators, decomposing total effects into direct ($E \rightarrow D$) and indirect ($E \rightarrow M_1 \rightarrow M_2 \rightarrow D$) components.

On the odds ratio ($OR_{E \rightarrow D}^{TE}$) scale, the total effect ($\beta_{E \rightarrow D}^{TE} = \log(OR_{E \rightarrow D}^{TE})$), comparing exposure level e with e^* , we could obtain the total effect:

$$\begin{aligned} \beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\ &= \log \left\{ \frac{P(D_e = 1) / \{1 - P(D_e = 1)\}}{P(D_{e^*} = 1) / \{1 - P(D_{e^*} = 1)\}} \right\} \\ &= \log \left\{ \frac{P(D_e = 1) \times \{1 - P(D_{e^*} = 1)\}}{\{1 - P(D_e = 1)\} \times P(D_{e^*} = 1)} \right\} \\ &= \log \left\{ \frac{P(D = 1 | e = 1) \times P(D = 0 | e^* = 0)}{P(D = 0 | e = 1) \times P(D = 1 | e^* = 0)} \right\} \\ &= \log \left\{ \frac{\xi_1}{\xi_2} \right\} \end{aligned}$$

$$\xi_1 = \left[\sum_{m_1 m_2} P(D = 1 | e = 1, m_2) P(m_2 | m_1) P(m_1 | e = 1) \right] \times \left[\sum_{m_1 m_2} P(D = 0 | e^* = 0, m_2) P(m_2 | m_1) P(m_1 | e^* = 0) \right]$$

$$\xi_2 = \left[\sum_{m_1 m_2} P(D = 0 | e = 1, m_2) P(m_2 | m_1) P(m_1 | e = 1) \right] \times \left[\sum_{m_1 m_2} P(D = 1 | e^* = 0, m_2) P(m_2 | m_1) P(m_1 | e^* = 0) \right]$$

The effect ($\beta_{ED|M_1}(m_1)$) of adjusting for mediator M_1 by logistic regression model can

be given

$$\begin{aligned} \beta_{ED|M_1}(m_1) &= \text{logit}\{P(D = 1 | e = 1, m_1)\} - \text{logit}\{P(D = 1 | e^* = 0, m_1)\} \\ &= \log \left\{ \frac{P(D = 1 | e = 1, m_1) P(D = 0 | e^* = 0, m_1)}{P(D = 0 | e = 1, m_1) P(D = 1 | e^* = 0, m_1)} \right\} \\ &= \log \left\{ \frac{\left[\sum_{m_2} P(D = 1 | e = 1, m_2) P(m_2 | m_1) \right] \times \left[\sum_{m_2} P(D = 0 | e^* = 0, m_2) P(m_2 | m_1) \right]}{\left[\sum_{m_2} P(D = 0 | e = 1, m_2) P(m_2 | m_1) \right] \times \left[\sum_{m_2} P(D = 1 | e^* = 0, m_2) P(m_2 | m_1) \right]} \right\} \end{aligned}$$

The effect ($\beta_{ED|M_2}(m_2)$) of adjusting for mediator M_2 by logistic regression model can

be given

$$\begin{aligned} \beta_{ED|M_2}(m_2) &= \text{logit}\{P(D = 1 | e = 1, m_2)\} - \text{logit}\{P(D = 1 | e^* = 0, m_2)\} \\ &= \log \left\{ \frac{P(D = 1 | e = 1, m_2) P(D = 0 | e^* = 0, m_2)}{P(D = 0 | e = 1, m_2) P(D = 1 | e^* = 0, m_2)} \right\} \end{aligned}$$

The effect ($\beta_{ED|M_1, M_2}(m_1, m_2)$) of adjusting for mediator M_1 M_2 by logistic regression model can be given

$$\begin{aligned}\beta_{ED|M_1, M_2}(m_1, m_2) &= \text{logit}\{P(D=1|e=1, m_1, m_2)\} - \text{logit}\{P(D=1|e^*=0, m_1, m_2)\} \\ &= \log\left\{\frac{P(D=1|e=1, m_1, m_2)P(D=0|e^*=0, m_1, m_2)}{P(D=0|e=1, m_1, m_2)P(D=1|e^*=0, m_1, m_2)}\right\} \\ &= \log\left\{\frac{P(D=1|e=1, m_2)P(D=0|e^*=0, m_2)}{P(D=0|e=1, m_2)P(D=1|e^*=0, m_2)}\right\}\end{aligned}$$

Therefore, we could evaluate the biases that contains three adjustment models: A) adjustment for M_1 , $\text{bias}(m_1) = \beta_{ED|M_1}(m_1) - \beta_{E \rightarrow D}^{TE}$; B) adjustment for M_2 , $\text{bias}(m_2) = \beta_{ED|M_2}(m_2) - \beta_{E \rightarrow D}^{TE}$ and C) adjustment for M_1 and M_2 , $\text{bias}(m_1, m_2) = \beta_{ED|M_1, M_2}(m_1, m_2) - \beta_{E \rightarrow D}^{TE}$.

(2) Figure 1c shows that the exposure E independently causes M_1 and M_2 and indirectly influences the outcome D through M_1 and M_2 , forming three causal paths $E \rightarrow D$, $E \rightarrow M_1 \rightarrow D$ and $E \rightarrow M_2 \rightarrow D$.

On the odds ratio ($OR_{E \rightarrow D}^{TE}$) scale, the total effect ($\beta_{E \rightarrow D}^{TE} = \log(OR_{E \rightarrow D}^{TE})$), comparing exposure level e with e^* , we could obtain the total effect:

$$\begin{aligned}\beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\ &= \log\left\{\frac{P(D_e=1)/\{1-P(D_e=1)\}}{P(D_{e^*}=1)/\{1-P(D_{e^*}=1)\}}\right\} \\ &= \log\left\{\frac{P(D_e=1) \times \{1-P(D_{e^*}=1)\}}{\{1-P(D_e=1)\} \times P(D_{e^*}=1)}\right\} \\ &= \log\left\{\frac{P(D=1|e=1) \times P(D=0|e^*=0)}{P(D=0|e=1) \times P(D=1|e^*=0)}\right\} \\ &= \log\left\{\frac{\xi_1}{\xi_2}\right\}\end{aligned}$$

$$\xi_1 = \left[\sum_{m_1, m_2} P(D=1|e=1, m_1, m_2)P(m_2|e=1)P(m_1|e=1) \right] \times \left[\sum_{m_1, m_2} P(D=0|e^*=0, m_1, m_2)P(m_2|e^*=0)P(m_1|e^*=0) \right]$$

$$\xi_2 = \left[\sum_{m_1, m_2} P(D=0|e=1, m_1, m_2)P(m_2|e=1)P(m_1|e=1) \right] \times \left[\sum_{m_1, m_2} P(D=1|e^*=0, m_1, m_2)P(m_2|e^*=0)P(m_1|e^*=0) \right]$$

The effect ($\beta_{ED|M_1}(m_1)$) of adjusting for mediator M_1 by logistic regression model can be given

$$\begin{aligned}
\beta_{ED|M_1}(m_1) &= \text{logit}\{P(D=1|e=1, m_1)\} - \text{logit}\{P(D=1|e^*=0, m_1)\} \\
&= \log \left\{ \frac{P(D=1|e=1, m_1)P(D=0|e^*=0, m_1)}{P(D=0|e=1, m_1)P(D=1|e^*=0, m_1)} \right\} \\
&= \log \left\{ \frac{\left[\sum_{m_2} P(D=1|e=1, m_1, m_2)P(m_2|e=1) \right] \times \left[\sum_{m_2} P(D=0|e^*=0, m_1, m_2)P(m_2|e^*=0) \right]}{\left[\sum_{m_2} P(D=0|e=1, m_1, m_2)P(m_2|e=1) \right] \times \left[\sum_{m_2} P(D=1|e^*=0, m_1, m_2)P(m_2|e^*=0) \right]} \right\}
\end{aligned}$$

The effect ($\beta_{ED|M_2}(m_2)$) of adjusting for mediator M_2 by logistic regression model can

be given

$$\begin{aligned}
\beta_{ED|M_2}(m_2) &= \text{logit}\{P(D=1|e=1, m_2)\} - \text{logit}\{P(D=1|e^*=0, m_2)\} \\
&= \log \left\{ \frac{P(D=1|e=1, m_2)P(D=0|e^*=0, m_2)}{P(D=0|e=1, m_2)P(D=1|e^*=0, m_2)} \right\} \\
&= \log \left\{ \frac{\left[\sum_{m_1} P(D=1|e=1, m_1, m_2)P(m_1|e=1) \right] \times \left[\sum_{m_1} P(D=0|e^*=0, m_1, m_2)P(m_1|e^*=0) \right]}{\left[\sum_{m_1} P(D=0|e=1, m_1, m_2)P(m_1|e=1) \right] \times \left[\sum_{m_1} P(D=1|e^*=0, m_1, m_2)P(m_1|e^*=0) \right]} \right\}
\end{aligned}$$

The effect ($\beta_{ED|M_1, M_2}(m_1, m_2)$) of adjusting for mediator M_1 M_2 by logistic regression

model can be given

$$\begin{aligned}
\beta_{ED|M_1, M_2}(m_1, m_2) &= \text{logit}\{P(D=1|e=1, m_1, m_2)\} - \text{logit}\{P(D=1|e^*=0, m_1, m_2)\} \\
&= \log \left\{ \frac{P(D=1|e=1, m_1, m_2)P(D=0|e^*=0, m_1, m_2)}{P(D=0|e=1, m_1, m_2)P(D=1|e^*=0, m_1, m_2)} \right\}
\end{aligned}$$

Therefore, we could evaluate the biases that contains three adjustment models: A)

adjustment for M_1 , $bias(m_1) = \beta_{ED|M_1}(m_1) - \beta_{E \rightarrow D}^{TE}$; B) adjustment for M_2 ,

$bias(m_2) = \beta_{ED|M_2}(m_2) - \beta_{E \rightarrow D}^{TE}$ and C) adjustment for M_1 and M_2 ,

$bias(m_1, m_2) = \beta_{ED|M_1, M_2}(m_1, m_2) - \beta_{E \rightarrow D}^{TE}$.

(3) In Figure 1d, there exists five paths from E to D : $E \rightarrow D$, $E \rightarrow M_1 \rightarrow D$, $E \rightarrow M_2 \rightarrow D$, $E \rightarrow M_1 \leftarrow M_2 \rightarrow D$ and $E \rightarrow M_2 \rightarrow M_1 \rightarrow D$. In particular, the path $E \rightarrow M_1 \leftarrow M_2 \rightarrow D$ is a blocked path, due to the M_1 being a collider node.

On the odds ratio ($OR_{E \rightarrow D}^{TE}$) scale, the total effect ($\beta_{E \rightarrow D}^{TE} = \log(OR_{E \rightarrow D}^{TE})$), comparing exposure level e with e^* , we could obtain the total effect:

$$\begin{aligned} \beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\ &= \log \left\{ \frac{P(D_e = 1) / \{1 - P(D_e = 1)\}}{P(D_{e^*} = 1) / \{1 - P(D_{e^*} = 1)\}} \right\} \\ &= \log \left\{ \frac{P(D_e = 1) \times \{1 - P(D_{e^*} = 1)\}}{\{1 - P(D_e = 1)\} \times P(D_{e^*} = 1)} \right\} \\ &= \log \left\{ \frac{P(D = 1 | e = 1) \times P(D = 0 | e^* = 0)}{P(D = 0 | e = 1) \times P(D = 1 | e^* = 0)} \right\} \\ &= \log \left\{ \frac{\xi_1}{\xi_2} \right\} \end{aligned}$$

$$\begin{aligned} \xi_1 &= \left[\sum_{m_1, m_2} P(D = 1 | e = 1, m_1, m_2) P(m_2 | e = 1) P(m_1 | e = 1, m_2) \right] \\ &\quad \times \left[\sum_{m_1, m_2} P(D = 0 | e^* = 0, m_1, m_2) P(m_2 | e^* = 0) P(m_1 | e^* = 0, m_2) \right] \\ \xi_2 &= \left[\sum_{m_1, m_2} P(D = 0 | e = 1, m_1, m_2) P(m_2 | e = 1) P(m_1 | e = 1, m_2) \right] \\ &\quad \times \left[\sum_{m_1, m_2} P(D = 1 | e^* = 0, m_1, m_2) P(m_2 | e^* = 0) P(m_1 | e^* = 0, m_2) \right] \end{aligned}$$

The effect ($\beta_{ED|M_1}(m_1)$) of adjusting for mediator M_1 by logistic regression model can be given

$$\begin{aligned} \beta_{ED|M_1}(m_1) &= \text{logit}\{P(D = 1 | e = 1, m_1)\} - \text{logit}\{P(D = 1 | e^* = 0, m_1)\} \\ &= \log \left\{ \frac{P(D = 1 | e = 1, m_1) P(D = 0 | e^* = 0, m_1)}{P(D = 0 | e = 1, m_1) P(D = 1 | e^* = 0, m_1)} \right\} \\ &= \log \left\{ \frac{\left[\sum_{m_2} P(D = 1 | e = 1, m_1, m_2) P(m_2 | e = 1, m_1) \right] \times \left[\sum_{m_2} P(D = 0 | e^* = 0, m_1, m_2) P(m_2 | e^* = 0, m_1) \right]}{\left[\sum_{m_2} P(D = 0 | e = 1, m_1, m_2) P(m_2 | e = 1, m_1) \right] \times \left[\sum_{m_2} P(D = 1 | e^* = 0, m_1, m_2) P(m_2 | e^* = 0, m_1) \right]} \right\} \\ &= \log \left\{ \frac{\xi_1}{\xi_2} \right\} \end{aligned}$$

$$\begin{aligned} \xi_1 &= \left[\sum_{m_2} P(D = 1 | e = 1, m_1, m_2) \frac{P(m_1 | e = 1, m_2) P(m_2 | e = 1)}{\sum_{m_2} P(m_1 | e = 1, m_2) P(m_2 | e = 1)} \right] \\ &\quad \times \left[\sum_{m_2} P(D = 0 | e^* = 0, m_1, m_2) \frac{P(m_1 | e^* = 0, m_2) P(m_2 | e^* = 0)}{\sum_{m_2} P(m_1 | e^* = 0, m_2) P(m_2 | e^* = 0)} \right] \\ \xi_2 &= \left[\sum_{m_2} P(D = 0 | e = 1, m_1, m_2) \frac{P(m_1 | e = 1, m_2) P(m_2 | e = 1)}{\sum_{m_2} P(m_1 | e = 1, m_2) P(m_2 | e = 1)} \right] \\ &\quad \times \left[\sum_{m_2} P(D = 1 | e^* = 0, m_1, m_2) \frac{P(m_1 | e^* = 0, m_2) P(m_2 | e^* = 0)}{\sum_{m_2} P(m_1 | e^* = 0, m_2) P(m_2 | e^* = 0)} \right] \end{aligned}$$

The effect ($\beta_{ED|M_2}(m_2)$) of adjusting for mediator M_2 by logistic regression model can be given

$$\begin{aligned}
\beta_{ED|M_2}(m_2) &= \text{logit}\{P(D=1|e=1,m_2)\} - \text{logit}\{P(D=1|e^*=0,m_2)\} \\
&= \log \left\{ \frac{P(D=1|e=1,m_2)P(D=0|e^*=0,m_2)}{P(D=0|e=1,m_2)P(D=1|e^*=0,m_2)} \right\} \\
&= \log \left\{ \frac{\left[\sum_{m_1} P(D=1|e=1,m_1,m_2)P(m_1|e=1,m_2) \right] \times \left[\sum_{m_1} P(D=0|e^*=0,m_1,m_2)P(m_1|e^*=0,m_2) \right]}{\left[\sum_{m_1} P(D=0|e=1,m_1,m_2)P(m_1|e=1,m_2) \right] \times \left[\sum_{m_1} P(D=1|e^*=0,m_1,m_2)P(m_1|e^*=0,m_2) \right]} \right\}
\end{aligned}$$

The effect ($\beta_{ED|M_1,M_2}(m_1,m_2)$) of adjusting for mediator M_1 M_2 by logistic regression model can be given

$$\begin{aligned}
\beta_{ED|M_1,M_2}(m_1,m_2) &= \text{logit}\{P(D=1|e=1,m_1,m_2)\} - \text{logit}\{P(D=1|e^*=0,m_1,m_2)\} \\
&= \log \left\{ \frac{P(D=1|e=1,m_1,m_2)P(D=0|e^*=0,m_1,m_2)}{P(D=0|e=1,m_1,m_2)P(D=1|e^*=0,m_1,m_2)} \right\}
\end{aligned}$$

Therefore, we could evaluate the biases that contains three adjustment models: A) adjustment for M_1 , $\text{bias}(m_1) = \beta_{ED|M_1}(m_1) - \beta_{E \rightarrow D}^{TE}$; B) adjustment for M_2 , $\text{bias}(m_2) = \beta_{ED|M_2}(m_2) - \beta_{E \rightarrow D}^{TE}$ and C) adjustment for M_1 and M_2 , $\text{bias}(m_1,m_2) = \beta_{ED|M_1,M_2}(m_1,m_2) - \beta_{E \rightarrow D}^{TE}$.

(4) In Figure 1e, the causal diagrams contained a confounder of exposure-outcome relationship. On the odds ratio ($OR_{E \rightarrow D}^{TE}$) scale, the total effect ($\beta_{E \rightarrow D}^{TE} = \log(OR_{E \rightarrow D}^{TE})$), comparing exposure level e with e^* , we could obtain the total effect:

$$\begin{aligned}
\beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\
&= \log \left\{ \frac{P(D_e=1)/\{1-P(D_e=1)\}}{P(D_{e^*}=1)/\{1-P(D_{e^*}=1)\}} \right\} \\
&= \log \left\{ \frac{P(D_e=1) \times \{1-P(D_{e^*}=1)\}}{\{1-P(D_e=1)\} \times P(D_{e^*}=1)} \right\} \\
&= \log \left\{ \frac{P(D=1|e=1) \times P(D=0|e^*=0)}{P(D=0|e=1) \times P(D=1|e^*=0)} \right\} \\
&= \log \left\{ \frac{\left[\sum_{mu} P(D=1|e=1,m,u)P(m|e=1)P(u) \right] \times \left[\sum_{mu} P(D=0|e^*=0,m,u)P(m|e^*=0)P(u) \right]}{\left[\sum_{mu} P(D=0|e=1,m,u)P(m|e=1)P(u) \right] \times \left[\sum_{mu} P(D=1|e^*=0,m,u)P(m|e^*=0)P(u) \right]} \right\}
\end{aligned}$$

The effect ($\beta_{ED|M}(m)$) of adjusting for mediator M by logistic regression model can be given

$$\begin{aligned}
 \beta_{ED|M}(m) &= \log it(P(D=1|e=1,m)) - \log it(P(D=1|e^*=0,m)) \\
 &= \log \left\{ \frac{P(D=1|e=1,m) \times P(D=0|e^*=0,m)}{P(D=0|e=1,m) \times P(D=1|e^*=0,m)} \right\} \\
 &= \log \left\{ \frac{[\sum_u P(D=1|e=1,m,u)p(u|e=1,m)] \times [\sum_u P(D=0|e^*=0,m,u)p(u|e^*=0,m)]}{[\sum_u P(D=0|e=1,m,u)p(u|e=1,m)] \times [\sum_u P(D=1|e^*=0,m,u)p(u|e^*=0,m)]} \right\} \\
 &= \log \left\{ \frac{[\sum_u P(D=1|e=1,m,u) \frac{p(e=1|u)p(u)}{\sum_u p(e=1|u)p(u)}] \times [\sum_u P(D=0|e^*=0,m,u) \frac{p(e^*=0|u)p(u)}{\sum_u p(e^*=0|u)p(u)}]}{[\sum_u P(D=0|e=1,m,u) \frac{p(e=1|u)p(u)}{\sum_u p(e=1|u)p(u)}] \times [\sum_u P(D=1|e^*=0,m,u) \frac{p(e^*=0|u)p(u)}{\sum_u p(e^*=0|u)p(u)}]} \right\}
 \end{aligned}$$

Therefore, we could evaluate the biases of adjustment models:

$$bias(m) = \beta_{ED|M}(m) - \beta_{E \rightarrow D}^{TE}$$

(5) Figure 1f is a depiction of two parallel mediators M_1 and M_2 with confounder.

On the odds ratio ($OR_{E \rightarrow D}^{TE}$) scale, the total effect ($\beta_{E \rightarrow D}^{TE} = \log(OR_{E \rightarrow D}^{TE})$), comparing exposure level e with e^* , we could obtain the total effect:

$$\begin{aligned}
 \beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\
 &= \log \left\{ \frac{P(D_e=1) / \{1 - P(D_e=1)\}}{P(D_{e^*}=1) / \{1 - P(D_{e^*}=1)\}} \right\} \\
 &= \log \left\{ \frac{P(D_e=1) \times \{1 - P(D_{e^*}=1)\}}{\{1 - P(D_e=1)\} \times P(D_{e^*}=1)} \right\} \\
 &= \log \left\{ \frac{P(D=1|e=1) \times P(D=0|e^*=0)}{P(D=0|e=1) \times P(D=1|e^*=0)} \right\} \\
 &= \log \left\{ \frac{\xi_1}{\xi_2} \right\}
 \end{aligned}$$

$$\begin{aligned}
 \xi_1 &= [\sum_{m_1 m_2 u} P(D=1|e=1, m_1, m_2, u) P(m_2|e=1) P(m_1|e=1) P(u)] \\
 &\quad \times [\sum_{m_1 m_2 u} P(D=0|e^*=0, m_1, m_2, u) P(m_2|e^*=0) P(m_1|e^*=0) P(u)] \\
 \xi_2 &= [\sum_{m_1 m_2 u} P(D=0|e=1, m_1, m_2, u) P(m_2|e=1) P(m_1|e=1) P(u)] \\
 &\quad \times [\sum_{m_1 m_2 u} P(D=1|e^*=0, m_1, m_2, u) P(m_2|e^*=0) P(m_1|e^*=0) P(u)]
 \end{aligned}$$

The effect ($\beta_{ED|M_1}(m_1)$) of adjusting for mediator M_1 by logistic regression model can be given

$$\begin{aligned}\beta_{ED|M_1}(m_1) &= \text{logit}\{P(D=1|e=1,m_1)\} - \text{logit}\{P(D=1|e^*=0,m_1)\} \\ &= \log \left\{ \frac{P(D=1|e=1,m_1)P(D=0|e^*=0,m_1)}{P(D=0|e=1,m_1)P(D=1|e^*=0,m_1)} \right\} \\ &= \log \left\{ \frac{\xi_1}{\xi_2} \right\}\end{aligned}$$

$$\begin{aligned}\xi_1 &= \left[\sum_{m_2 u} P(D=1|e=1,m_1,m_2,u) \frac{P(m_2|e=1)P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right] \\ &\quad \times \left[\sum_{m_2 u} P(D=0|e^*=0,m_1,m_2,u) \frac{P(m_2|e^*=0)P(e^*=0|u)P(u)}{\sum_u P(e^*=0|u)P(u)} \right] \\ \xi_2 &= \left[\sum_{m_2 u} P(D=0|e=1,m_1,m_2,u) \frac{P(m_2|e=1)P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right] \\ &\quad \times \left[\sum_{m_2 u} P(D=1|e^*=0,m_1,m_2,u) \frac{P(m_2|e^*=0)P(e^*=0|u)P(u)}{\sum_u P(e^*=0|u)P(u)} \right]\end{aligned}$$

The effect ($\beta_{ED|M_2}(m_2)$) of adjusting for mediator M_2 by logistic regression model can be given

$$\begin{aligned}\beta_{ED|M_2}(m_2) &= \text{logit}\{P(D=1|e=1,m_2)\} - \text{logit}\{P(D=1|e^*=0,m_2)\} \\ &= \log \left\{ \frac{P(D=1|e=1,m_2)P(D=0|e^*=0,m_2)}{P(D=0|e=1,m_2)P(D=1|e^*=0,m_2)} \right\} \\ &= \log \left\{ \frac{\xi_1}{\xi_2} \right\}\end{aligned}$$

$$\begin{aligned}\xi_1 &= \left[\sum_{m_1 u} P(D=1|e=1,m_1,m_2,u) \frac{P(m_1|e=1)P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right] \\ &\quad \times \left[\sum_{m_1 u} P(D=0|e^*=0,m_1,m_2,u) \frac{P(m_1|e^*=0)P(e^*=0|u)P(u)}{\sum_u P(e^*=0|u)P(u)} \right] \\ \xi_2 &= \left[\sum_{m_1 u} P(D=0|e=1,m_1,m_2,u) \frac{P(m_1|e=1)P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right] \\ &\quad \times \left[\sum_{m_1 u} P(D=1|e^*=0,m_1,m_2,u) \frac{P(m_1|e^*=0)P(e^*=0|u)P(u)}{\sum_u P(e^*=0|u)P(u)} \right]\end{aligned}$$

The effect ($\beta_{ED|M_1,M_2}(m_1,m_2)$) of adjusting for mediator $M_1 M_2$ by logistic regression model can be given

$$\begin{aligned} & \beta_{ED|M_1, M_2}(m_1, m_2) \\ &= \text{logit}\{P(D=1|e=1, m_1, m_2)\} - \text{logit}\{P(D=1|e^*=0, m_1, m_2)\} \\ &= \log\left\{\frac{P(D=1|e=1, m_1, m_2)P(D=0|e^*=0, m_1, m_2)}{P(D=0|e=1, m_1, m_2)P(D=1|e^*=0, m_1, m_2)}\right\} \\ &= \log\left\{\frac{\left[\sum_u P(D=1|e=1, m_1, m_2, u) \frac{P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)}\right] \times \left[\sum_u P(D=0|e^*=0, m_1, m_2, u) \frac{P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)}\right]}{\left[\sum_u P(D=0|e=1, m_1, m_2, u) \frac{P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)}\right] \times \left[\sum_u P(D=1|e^*=0, m_1, m_2, u) \frac{P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)}\right]}\right\} \end{aligned}$$

Therefore, we could evaluate the biases that contains three adjustment models: A) adjustment for M_1 , $bias(m_1) = \beta_{ED|M_1}(m_1) - \beta_{E \rightarrow D}^{TE}$; B) adjustment for M_2 , $bias(m_2) = \beta_{ED|M_2}(m_2) - \beta_{E \rightarrow D}^{TE}$ and C) adjustment for M_1 and M_2 , $bias(m_1, m_2) = \beta_{ED|M_1, M_2}(m_1, m_2) - \beta_{E \rightarrow D}^{TE}$.

Supplementary B

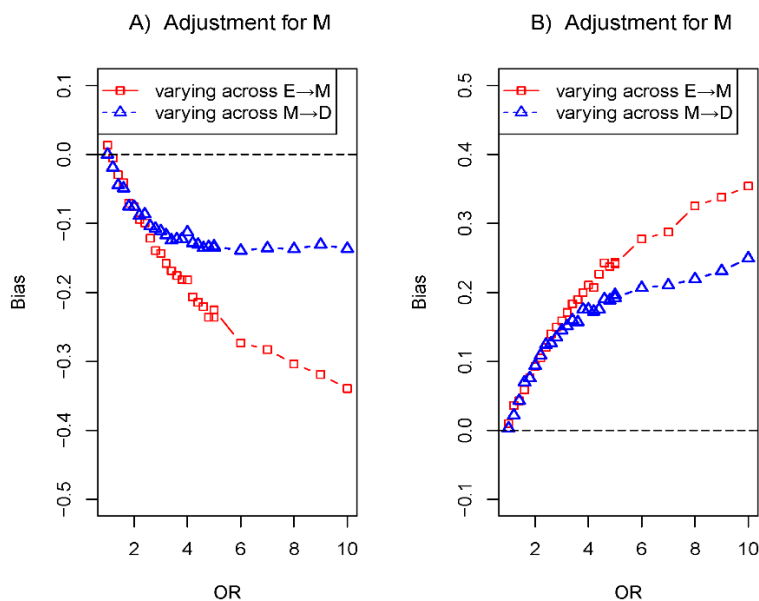


Figure S1: The biases with the effects $E \rightarrow M$ (red) and $M \rightarrow D$ (blue) increasing, respectively. Comparison of the bias of different effects in adjustment mediator.

The Figure S1-A obtained the result $bias < 0$ in Figure 3a with the effects $E \rightarrow M$, $M \rightarrow D$ and $E \rightarrow D$ fixing to $\ln 2$. The Figure S1-B gained the result $bias > 0$ in Figure 3c with the effects $E \rightarrow M$ and $E \rightarrow D$ fixing to $\ln 2$, effect $M \rightarrow D$ fixing to $-\ln 2$. We could obtain the bias performances of varying across the effects of exposure-mediator and mediator-outcome. The effect $E \rightarrow M$ of varying across was more sensitive than the effect $M \rightarrow D$ of varying across in Figure S1.

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*
Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(a) Indicate the study’s design with a commonly used term in the title or the abstract	1
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any pre-specified hypotheses	3-4
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5-6
Participants	6	(a) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> —Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants	5-6
		(b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5-6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-6
Bias	9	Describe any efforts to address potential sources of bias	5-6
Study size	10	Explain how the study size was arrived at	5-6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Not applicable
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	4-6
		(b) Describe any methods used to examine subgroups and interactions	Not applicable
		(c) Explain how missing data were addressed	Not applicable
		(d) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed	Not applicable

		<i>Cross-sectional study</i> —If applicable, describe analytical methods taking account of sampling strategy	
		(e) Describe any sensitivity analyses	6
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Not applicable
		(b) Give reasons for non-participation at each stage	Not applicable
		(c) Consider use of a flow diagram	Not applicable
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7-10
		(b) Indicate number of participants with missing data for each variable of interest	Not applicable
		(c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	Not applicable
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time	Not applicable
		<i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure	Not applicable
		<i>Cross-sectional study</i> —Report numbers of outcome events or summary measures	7-10
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	7-10
		(b) Report category boundaries when continuous variables were categorized	Not applicable
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not applicable
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	7-10
Discussion			
Key results	18	Summarise key results with reference to study objectives	11-12
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	11-12
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-12
Generalisability	21	Discuss the generalisability (external validity) of the study results	12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	13

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Sensitivity analysis for mistakenly adjusting for mediators in estimating total effect: a simulation study

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2016-015640.R2
Article Type:	Research
Date Submitted by the Author:	09-Jun-2017
Complete List of Authors:	Wang, Tingting; Shandong University, Department of Epidemiology and Biostatistics Li, Hongkai; Shandong University, Department of Epidemiology and Biostatistics Su, Ping; Shandong University, Department of Epidemiology and Biostatistics Yu, Yuanyuan; Shandong University, Department of Epidemiology and Biostatistics Sun, Xiaoru; Shandong University, Department of Epidemiology and Biostatistics Liu, Yi; Shandong University, Department of Epidemiology and Biostatistics Yuan, Zhongshang; Shandong University, Department of Epidemiology and Biostatistics Xue, Fuzhong; Shandong University, Department of Epidemiology and Biostatistics
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Public health
Keywords:	EPIDEMIOLOGY, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS

SCHOLARONE™
Manuscripts

1 Sensitivity analysis for mistakenly adjusting for mediators in 2 estimating total effect: a simulation study

3 Tingting Wang[†], Hongkai Li[†], Ping Su, Yuanyuan Yu, Xiaoru Sun, Yi Liu, Zhongshang Yuan,
4 Fuzhong Xue*

5 Department of Biostatistics, School of Public Health, Shandong University, Jinan City, Shandong
6 Province, P.R.China

7 Email addresses:

8 TTW: wangtingting161780@163.com

9 HKL: lihongkaiyouxiang@163.com

10 PS: sdsuping@163.com

11 YYY: yu_yy_1993@163.com

12 XRS: sunny150139@126.com

13 YL: liuyi238@sdu.edu.cn

14 ZSY: yuanzhongshang@sdu.edu.cn

15
16 *Corresponding Author

17 xuefzh@sdu.edu.cn (FZX)

18
19 **Word Count: 3647**

20 Abstract

21 **Objectives:** In observational studies, epidemiologists often attempt to estimate the
22 total effect of exposure on outcome of interest. However, when the underlying
23 diagram is unknown and only limited knowledge is available, dissecting biases
24 performances are essential to estimate the total effect of exposure on outcome in
25 mistakenly adjusting for mediators under logistic regression. Through simulation, we
26 focus on six causal diagrams concerning different roles of mediators. Sensitivity
27 analysis was conducted to assess the bias performances of varying across the effects
28 of exposure-mediator and mediator-outcome in adjusting for mediator.

29 **Setting:** Based on the causal relationships in real world, we compare the biases of
30 varying across the effects of exposure-mediator with the ones of varying across the
31 effects of mediator-outcome under the situation of adjusting for mediator. The

1
2
3 1 magnitude of the bias was defined by the difference between the estimated effect
4
5 2 using logistic regression and the total effect of the exposure on the outcome.
6

7 3 **Results:** In the following four scenarios: a single mediator, two series mediators, two
8
9 4 independent parallel mediators or two correlated parallel mediators, the biases of
10
11 5 varying across the effects of exposure-mediator was greater than the ones of varying
12
13 6 across the effects mediator-outcome in adjusting for the mediator. While in other two
14
15 7 scenarios: a single mediator or two independent parallel mediators in the presence of
16
17 8 unobserved confounders, the biases of varying across the effects of exposure-mediator
18
19 9 was less than the ones of varying across the effects mediator-outcome in adjusting for
20
21 10 the mediator.
22

23 11 **Conclusions:** The biases were higher sensitive to the variation of effects of
24
25 12 exposure-mediator than effects of mediator-outcome in adjusting for mediator in the
26
27 13 absence of unobserved confounders; while the biases were higher sensitive to the
28
29 14 variation of effects of mediator-outcome than effects of exposure-mediator in the
30
31 15 presence of unobserved confounder.
32
33

34 16 **Strengths and limitations of this study**

35
36
37 17 1) For six different causal diagrams, we compared biases of distinct adjustment
38
39 18 strategies with and without adjusting for mediators by conducting simulation studies.
40

41
42
43 19 2) Sensitivity analysis was conducted to assess the performances of varying across the
44
45 20 effects of exposure-mediator and mediator-outcome.
46

47
48 21 3) The simulation schemes and parameters were conducted mainly based on real
49
50 22 observational studies.
51

52
53 23 4) Combination of theoretical derivation and simulation studies make the results more
54
55 24 credible.
56
57
58
59
60

1
2
3 5) The limitation of simulation studies was under the framework of logistic regression
4
5 and only focused on binary variables.
6

7 **Introduction**

8
9
10 Estimating the total effect of the exposure (E) on the outcome (D) is a great challenge
11
12 in epidemiology studies, because confounders are commonly confused with
13
14 mediators.¹⁻³ If confounders and mediators are misclassified, the ability to control
15
16 confounder in the estimation of the total effect of the exposure on the outcome is
17
18 hampered. Actually, various strategies are used to eliminate confounding bias in
19
20 non-randomized controlled studies. The conventional approaches contain multivariate
21
22 regression, stratification, standardization and inverse-probability weighting, etc.⁴⁻⁵
23
24 Furthermore, causal diagrams provides a formal conceptual framework to identify and
25
26 select confounders,⁶⁻⁷ so that it can avoid falling into analytic pitfalls.⁸ In practice,
27
28 even the underlying causal diagrams and the role of covariates (mediator, confounder,
29
30 collider and instrumental variable) are not all learned, investigators usually adjusted
31
32 for the covariates that are associated with the outcome and exposure.⁹⁻¹² Therefore,
33
34 our paper focuses on the biases of varying across the effects of exposure-mediator
35
36 ($E \rightarrow M$) and mediator-outcome ($M \rightarrow D$) in mistakenly adjusting for mediators under
37
38 logistic regression model.
39
40
41
42
43

44 Several causal inference literatures have made a considerable contribution to
45
46 mediation analysis by providing definitions for direct and indirect effects that allow
47
48 for the effect decomposition of a total effect into a direct and an indirect effect.¹³⁻²¹
49
50 Arbitrarily adjusting for a mediator would generally bias the estimate of the total
51
52 effect of the exposure on the outcome.^{8,22-23} Practically, it can mistakenly identify a
53
54 non-confounding risk factor as a confounder. In the perspective of causal diagrams,
55
56
57
58
59
60

1 little attention was paid to the biases in adjusting for mediators under the logistic
2 regression model in estimating the total effect of E on D . Hence, we focused on the
3 sensitivity analysis technique to assess the biases of varying across the effects of
4 $E \rightarrow M$ and $M \rightarrow D$ in adjusting for mediator.

5 In this paper, six typical causal diagrams corresponding to causal correlation are
6 given in Figure 1: a single mediator (Figure 1a); two series mediators (Figure 1b); two
7 independent parallel mediators (Figure 1c); two correlated parallel mediators (Figure
8 1d); a single mediator with an unobserved confounder (Figure 1e); two parallel
9 mediators with an unobserved confounder (Figure 1f). The paper aims to explore the
10 sensitivity of biases to the variation of the effects of $E \rightarrow D$ and $M \rightarrow D$ in adjusting for
11 mediator. Hence, both theoretical proofs and quantitative simulations were performed
12 to dissect the bias of varying across the effect of $E \rightarrow M$ and the one of varying across
13 the effect of $M \rightarrow D$ in adjusting for mediators under logistic model.

14 **Methods**

15 A directed acyclic graph (*DAG*) is composed of variables (nodes) and arrows (directed
16 edges) between nodes such that the graph is acyclic. The causal diagrams formalized
17 as directed acyclic graphs (*DAGs*), providing investigators with powerful tools for
18 bias assessment.²⁴ It provides a device for deducing the statistical associations implied
19 by causal relations. Furthermore, given a set of observed statistical associations, a
20 researcher armed with causal diagrams theory can systematically characterize all
21 causal structures compatible with the observations.²⁵⁻²⁶

22 The total effect of the exposure on the outcome can be calculated based on the
23 *do-calculus* and *back-door* criterion proposed by Judea Pearl.²⁷⁻²⁸ For exposure X and

1 outcome Y , a set of variables Z satisfies the backdoor path criterion with respect to (X ,
 2 Y) if no variable in Z is a descendant of X and Z blocks all back-door paths from X to
 3 Y . Then the effect of X on Y is given by the formula,

$$P(y|do(x)) = \sum_z P(y|x,z)P(z)$$

4 Note that the expression on the right hand side of the equation is simply a
 5 standardized mean. The difference $E(Y|do(x')) - E(Y|do(x''))$ is taken as the
 6 definition of “causal effect”, where x' and x'' are two distinct realizations of X .²³
 7 The interventional distribution, such as that corresponding to $Y(x)$, namely
 8 $P(y|do(x))$, is not necessarily equal to a conditional distribution $P(y|x)$. It stands for
 9 the probability of $Y=y$ when the exposure X set to level x . The ignorability
 10 assumption $Y(x) \perp X$ states that if we happen to have information on the exposure
 11 variable, it does not give us any information about the outcome Y after the
 12 intervention $do(x)$ was performed. Besides it can be shown that if ignorability holds
 13 for $Y(x)$ and X (alternatively if there are no back-door paths from X to Y in the
 14 corresponding causal DAGs), then $P(y|do(x)) = P(y|x)$.²⁹⁻³⁰

15 Let D_e and M_e denote respectively the values of the outcome and mediator that
 16 would have been observed had the exposure E been set to level e . On the odds ratio
 17 ($OR_{E \rightarrow D}^{TE}$) scale, the total effect ($\beta_{E \rightarrow D}^{TE} = \log(OR_{E \rightarrow D}^{TE})$), comparing exposure level e

18 with e^* , is given by $OR_{E \rightarrow D}^{TE} = \frac{P(D_e = 1) / \{1 - P(D_e = 1)\}}{P(D_{e^*} = 1) / \{1 - P(D_{e^*} = 1)\}}$.²⁰⁻²¹ While the effect

19 ($\beta_{ED|M}(m)$) of adjusting for mediator M by logistic regression model can be given

$$\begin{aligned} \beta_{ED|M}(m) &= \text{logit}\{P(D=1|e=1,m)\} - \text{logit}\{P(D=1|e^*=0,m)\} \\ &= \log\left\{\frac{P(D=1|e=1,m)P(D=0|e^*=0,m)}{P(D=0|e=1,m)P(D=1|e^*=0,m)}\right\} \end{aligned}$$

1 where $P(D=1|e,m)$ denotes the probability of $D=1$ when the exposure E , and
 2 mediator M , have been set to level e , and m , respectively. Taking Figure 1a as an
 3 example, the logistic regression is

$$4 \quad \text{logit}\{P(D=1|e,m)\} = \alpha_1 + \beta_0 e + \beta_2 m.$$

5 Therefore, the total effect ($\beta_{E \rightarrow D}^{TE}$) of exposure E on outcome D on the scale of
 6 logarithm odds ratio was equal to

$$\begin{aligned} 7 \quad \beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\ &= \log \left\{ \frac{P(D_e=1) / \{1 - P(D_e=1)\}}{P(D_{e^*}=1) / \{1 - P(D_{e^*}=1)\}} \right\} \\ &= \text{logit}\{P(D_e=1)\} - \text{logit}\{P(D_{e^*}=1)\} \\ &= \text{logit}\{P(D=1|e=1)\} - \text{logit}\{P(D=1|e^*=0)\} \\ &= \text{logit} \left\{ \sum_m P(D=1|e=1,m)P(m|e=1) \right\} - \text{logit} \left\{ \sum_m P(D=1|e^*=0,m)P(m|e^*=0) \right\} \end{aligned}$$

8 The effect estimation ($\hat{\beta}_{ED|M}(m)$) of adjusting for mediator M by logistic regression
 9 model was equal to

$$10 \quad \hat{\beta}_{ED|M}(m) = \text{logit}\{\hat{P}(D=1|e=1,m)\} - \text{logit}\{\hat{P}(D=1|e^*=0,m)\}$$

11 where $\hat{P}(D=1|e=1,m)$ denotes the probability of $D=1$ when the exposure E ,
 12 and mediator M , have been set to level $e=1$, and m , respectively. And
 13 $\hat{P}(D=1|e^*=0,m)$ denotes the probability of $D=1$ when the exposure E , and
 14 mediator M , have been set to level $e^*=0$, and m , respectively. The theoretical results
 15 of other causal diagrams in Figure 1 have been shown in the supplementary A.

16 Note that the bias was defined by taking a difference between effect estimation by
 17 adjusting for mediator using logistic regression and the total effect of exposure E on
 18 outcome D i.e. $\text{bias} = E[\hat{\beta}_{ED|M}(m)] - \beta_{E \rightarrow D}^{TE}$. We dissected the biases behavior by
 19 varying across the effects of $E \rightarrow M$ and $M \rightarrow D$ in mistakenly adjusting for mediator
 20 under the framework of logistic regression model.

21 Simulation

1 Six scenarios are designed to dissect the sensitivity of bias to the variation of the
 2 effects of exposure-mediator and mediator-outcome in adjusting for mediators under
 3 the framework of logistic regression model, these DAGs are shown in Figure 1. We
 4 made the following assumptions for the simulation: 1) all variables were binary
 5 following a Bernoulli distribution; 2) the effect from parent nodes to their child node
 6 were positive and log-linearly additive. Taking Figure 1a as an example, we randomly
 7 generated the exposure following a Bernoulli distribution (i.e. let $P(e = 1) = \pi$), then,
 8 $P_M = \exp(\alpha_0 + \beta_1 e) / \{1 + \exp(\alpha_0 + \beta_1 e)\}$ for calculating the distribution probability of
 9 child node M from its parent node E . Similarly,
 10 $P_D = \exp(\alpha_1 + \beta_0 e + \beta_2 m) / \{1 + \exp(\alpha_1 + \beta_0 e + \beta_2 m)\}$ generated the distribution
 11 probability of D , where the parameters α_0 and α_1 denoted the intercept of M and D
 12 respectively, and effect parameter $\beta_0, \beta_1, \beta_2$ referred to the effects of the parent node
 13 on their corresponding child node using log odds ratio scale.

14 After generating data, we dissected the biases behavior between the effects of
 15 $E \rightarrow M$ and $M \rightarrow D$ in mistakenly adjusting for mediator under logistic regression model.
 16 In scenario 1 (Figure 1a), we compared the performances by across varying the effects
 17 of $E \rightarrow M$ and $M \rightarrow D$. Similarly, in scenario 2 (Figure 1b), the effects of $E \rightarrow M_1$,
 18 $M_1 \rightarrow M_2$ and $M_2 \rightarrow D$ were explored. In scenario 3 (Figure 1c), we dissected the effects
 19 of $E \rightarrow M_1$ ($E \rightarrow M_2$) and $M_1 \rightarrow D$ ($M_2 \rightarrow D$). The comparison of scenario 4 (Figure 1d)
 20 was the same as scenario 3 (Figure 1c). In scenario 5 (Figure 1e), the effects of $E \rightarrow M$
 21 and $M \rightarrow D$ were excavated. The scenario 6 (Figure 1f) was identical to the scenario 3.
 22 We explored the biases in adjusting for mediator under logistic regression model and

1 thus identified the sensitivity of biases to the variation of the effects of
2 exposure-mediator and mediator-outcome.

3 For each of the 6 simulation scenarios, we observed bias performances of varying
4 across distinct effects in adjusting for mediator using logistic regression model with
5 1000 simulations repetitions. All simulations were conducted using software R from
6 CRAN (<http://cran.r-project.org/>).

7 Results

8 *Scenario 1: one single mediator (Figure 1a)*

9 In Figure 1(a), E has a direct ($E \rightarrow D$) effect and an indirect ($E \rightarrow M \rightarrow D$) effect on D .
10 Figure 2A depicted that the bias of varying across the effect of $E \rightarrow M$ was obviously
11 greater than the bias of varying across the effect of $M \rightarrow D$. That is, the sensitivity of
12 bias to the variation of the effect $E \rightarrow M$ was greater than the effect of $M \rightarrow D$ in
13 adjusting for the mediator M using logistic regression model. In particular, if the
14 effect of $E \rightarrow M$ was specified to zero in Figure 2B, M was associated with D
15 conditional on E and unconditionally independent with E , M became an independent
16 risk factor of the outcome, adjusting for M obtained a positive “bias”. Such bias was a
17 consequence of non-collapsibility of odds ratio, and the M-conditional ORs must be
18 far from 1 than the unconditional ORs.³¹⁻³² Actually, both adjustment and
19 non-adjustment for M should yield unbiased causal effect estimates. Certainly, in this
20 case, both marginal OR and conditional OR obtained from standardization and
21 inverse-probability weighting were equals to total effect.³³ Moreover, Figure 2A
22 indicated that adjusting for mediator M was indeed biased to the total effect of the
23 exposure on the outcome.

24 The total effect ($\beta_{E \rightarrow D}^{TE}$) of exposure E on outcome D on the scale of logarithm odds
25 ratio was equal to

$$\begin{aligned}
\beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) = \log \left\{ \frac{P(D_e = 1) / \{1 - P(D_e = 1)\}}{P(D_{e^*} = 1) / \{1 - P(D_{e^*} = 1)\}} \right\} \\
&= \log \left\{ \frac{P(D_e = 1) \times \{1 - P(D_{e^*} = 1)\}}{\{1 - P(D_e = 1)\} \times P(D_{e^*} = 1)} \right\} \\
&= \log \left\{ \frac{P(D = 1 | e = 1) \times P(D = 0 | e^* = 0)}{P(D = 0 | e = 1) \times P(D = 1 | e^* = 0)} \right\} \\
&= \log \left\{ \frac{[\sum_m P(D = 1 | e = 1, m) P(m | e = 1)] \times [\sum_m P(D = 0 | e^* = 0, m) P(m | e^* = 0)]}{[\sum_m P(D = 0 | e = 1, m) P(m | e = 1)] \times [\sum_m P(D = 1 | e^* = 0, m) P(m | e^* = 0)]} \right\}
\end{aligned}$$

The effect ($\beta_{ED|M}(m)$) of adjusting for mediator M by logistic regression model can be given

$$\begin{aligned}
\beta_{ED|M}(m) &= \text{logit} \{P(D = 1 | e = 1, m)\} - \text{logit} \{P(D = 1 | e^* = 0, m)\} \\
&= \log \left\{ \frac{P(D = 1 | e = 1, m) \times \{1 - P(D = 1 | e^* = 0, m)\}}{\{1 - P(D = 1 | e = 1, m)\} \times P(D = 1 | e^* = 0, m)} \right\} \\
&= \beta_0
\end{aligned}$$

β_0 denotes coefficient of the E adjusting for M using logistic regression model.

Furthermore, the effect of adjusting for M was equal to the controlled direct effect.¹⁹

Therefore, the bias of adjusting for mediator using logistic regression model could be obtained i.e. $\text{bias} = \beta_{ED|M}(m) - \beta_{E \rightarrow D}^{TE}$. We added signs to the edges of the directed

acyclic graph to indicate the presence of a particular positive or negative effect in the

Figure 3. Therefore, we gained $\text{bias} < 0$ under the condition of $\beta_1 * \beta_2 > 0$ (the

effect $E \rightarrow M$ β_1 and the effect $M \rightarrow D$ β_2), indicating that the total effect of E on D

was biased in adjusting for M using logistic regression model in Figure 3a, Figure 3b,

Figure 3e & Figure 3f. And the bias was less than zero when the effect $E \rightarrow M$ (β_1)

and the effect $M \rightarrow D$ (β_2) share same signs. (i.e. both the effects $E \rightarrow M$ ($\beta_1 > 0$) and

$M \rightarrow D$ ($\beta_2 > 0$) were a positive sign or both the effects $E \rightarrow M$ ($\beta_1 < 0$) and

$M \rightarrow D$ ($\beta_2 < 0$) were a negative sign). Furthermore, we obtained $\text{bias} > 0$, if

$\beta_1 * \beta_2 < 0$, suggesting that the total effect of E on D was biased in adjusting for M in

1
2
3
4 1 Figure 3c, Figure 3d, Figure 3g & Figure 3h. And the bias was greater than zero when
5
6 2 the signs of the effects $E \rightarrow M$ (β_1) and $M \rightarrow D$ (β_2) were the opposite. The results
7
8
9 3 illustrated that the bias was less than zero under the case of the effects of
10
11 4 exposure-mediator and mediator-outcome sharing the same sign; the bias was greater
12
13 5 than zero under circumstances of the effects of exposure-mediator and
14
15 6 mediator-outcome having opposite signs. We also illustrated the case of the Figure 3c
16
17 7 with the effects $E \rightarrow M$ and $E \rightarrow D$ greater than zero, effect $M \rightarrow D$ less than zero in
18
19 8 supplementary B. More details of theoretical derivation can be found in Appendix.
20
21
22

23
24 9 ***Scenario 2: two series mediators (Figure 1b)***

25
26 10 Figure 1(b) is a depiction through two series mediators, decomposing total effects into
27
28 11 direct effect ($E \rightarrow D$) and indirect effect ($E \rightarrow M_1 \rightarrow M_2 \rightarrow D$). The bias of varying across
29
30 12 the effect of $E \rightarrow M_1$ was greater than the one of varying across the effect of $M_2 \rightarrow D$
31
32 13 under adjustment for M_1 , M_2 and $M_1 M_2$ together in Figure 4, respectively. In this
33
34 14 situation, the correlation of series mediators was strong enough to avoid M_2 from
35
36 15 becoming an independent cause of the outcome.
37
38
39

40
41 16 ***Scenario 3: two independent parallel mediators (Figure 1c)***

42
43 17 Figure 1c shows that the exposure E independently causes M_1 and M_2 and indirectly
44
45 18 influences the outcome D through M_1 and M_2 , forming three causal paths $E \rightarrow D$,
46
47 19 $E \rightarrow M_1 \rightarrow D$ and $E \rightarrow M_2 \rightarrow D$. The results obtained that the bias of varying across the
48
49 20 effect of $E \rightarrow M_1$ was considerably greater than the one of varying across the effect of
50
51 21 $M_1 \rightarrow D$ under adjustment for M_1 in Figure 5A. However, the bias of varying across the
52
53 22 effect of $E \rightarrow M_2$ was nearly equal to the one with varying across the effect of $M_2 \rightarrow D$
54
55
56
57
58
59
60

1
2
3
4 1 under the identical model of adjustment for M_1 in Figure 5A. Then, an above similar
5
6 2 result can be obtained in Figure 5B. In addition, Figure 5C indicated that biases of
7
8 3 varying across the effects of $E \rightarrow M_1$ and $E \rightarrow M_2$ were obviously greater than the ones
9
10 4 of varying across the effects of $M_1 \rightarrow D$ and $M_2 \rightarrow D$ under simultaneously adjusting for
11
12
13
14 5 M_1 and M_2 .

16
17 6 **Scenario 4: two correlated parallel mediators (Figure 1d)**

18 7 In Figure 1d, there exist five paths from E to D : $E \rightarrow D$, $E \rightarrow M_1 \rightarrow D$, $E \rightarrow M_2 \rightarrow D$,
19
20 8 $E \rightarrow M_1 \leftarrow M_2 \rightarrow D$ and $E \rightarrow M_2 \rightarrow M_1 \rightarrow D$. In particular, the path $E \rightarrow M_1 \leftarrow M_2 \rightarrow D$ is a
21
22 9 blocked path, due to the M_1 being a collider node. Figure 6A indicated that the bias of
23
24 10 varying across the effect of $E \rightarrow M_1$ was obviously greater than the one of varying
25
26 11 across the effect of $M_1 \rightarrow D$ under adjustment for M_1 . However, the bias of varying
27
28 12 across the effect of $E \rightarrow M_2$ was almost equal to the one of varying across the effect of
29
30 13 $M_2 \rightarrow D$ under the identical adjustment model. Similarly, an analogous result of biases
31
32 14 behavior was shown in Figure 6B. Besides, the biases of varying across the effects of
33
34 15 $E \rightarrow M_1$ and $E \rightarrow M_2$ were greater than the ones of varying across the effects of $M_1 \rightarrow D$
35
36 16 and $M_2 \rightarrow D$ in adjusting for M_1 and M_2 in Figure 6C. Simultaneously, the bias was
37
38 17 higher sensitive to the variation of effect of $E \rightarrow M_2$ than effect of $E \rightarrow M_1$ under
39
40 18 adjustment for M_1 and M_2 , which adjusting for the collider node M_1 would partially
41
42 19 open the path $E \rightarrow M_1 \leftarrow M_2 \rightarrow D$.

46
47
48 20 **Scenario 5: a single mediator with an unobserved confounder (Figure 1e)**

49
50 21 Figure 1e provides a causal diagram representing the relationship among exposure E ,
51
52 22 outcome D , mediator M and unobserved confounder U . It revealed that the bias of
53
54 23 varying across the effect of $E \rightarrow M$ was lower than the one of varying across the effect
55
56 24 of $M \rightarrow D$. Unobserved confounder distorts the association between the exposure and
57
58
59
60

1 outcome ($E \leftarrow U \rightarrow D$) in Figure 7.

2 **Scenario 6: two parallel mediators with an unobserved confounder (Figure 1f)**

3 As described above, Figure 1f is a depiction of two parallel mediators M_1 and M_2 with
4 an unobserved confounder U . The bias of varying across the effect of $E \rightarrow M_1$ was
5 obviously less than the one of varying across the effect of $M_1 \rightarrow D$ under the
6 adjustment for M_1 in Figure 8A. However, the bias of varying across the effect of
7 $E \rightarrow M_2$ was greater than the one of varying across the effect of $M_2 \rightarrow D$ under the
8 identical model of adjusting for M_1 . A similar result can also obtain in Figure 8B.
9 Besides, biases of varying across the effects of $E \rightarrow M_1$ and $E \rightarrow M_2$ were distinctly
10 less than the ones of varying across the effects of $M_1 \rightarrow D$ and $M_2 \rightarrow D$ under the
11 common model of adjusting for M_1 and M_2 in Figure 8C.

12 **Application**

13 In this analysis, we evaluated two statistical models (unadjusted and M-adjusted) to
14 assess the effect of diabetes on cardiovascular diseases under the scenario 1. The
15 information of 22900 people were collected from the Health Management Center of
16 Shandong Provincial Hospital (HMCSPH). All individuals were Urban Han Chinese
17 with the age above 20 years old and they took the physical examination in 2013.
18 Many studies focused on the associations diabetes with metabolic syndrome,³⁴
19 metabolic syndrome with cardiovascular disease,³⁵ respectively.

20 The exposure indicator E takes the value 1 if people suffer from diabetes, and zero
21 otherwise. The outcome D (cardiovascular diseases) takes the value 1 if the people
22 diagnosed with cardiovascular diseases, and takes the value 0 otherwise. The mediator

1
2
3
4 1 *M* (metabolic syndrome) takes the value 1 if people were the metabolic syndrome and
5
6 2 takes the value 0 otherwise. When adjustment for age and gender by using the logistic
7
8 3 regression model can obtain the total effect of diabetes *E* on cardiovascular diseases *D*
9
10 4 being equal to $\beta = 0.598$ (95% confidence interval (*CI*), 0.307~0.877). Then the effect
11
12 5 of adjusting for metabolic syndrome *M* was equal to $\beta_M = 0.429$ (95% confidence
13
14 6 interval (*CI*), 0.113~0.736). Therefore, the bias was $\beta_M - \beta = -0.169 < 0$, suggesting
15
16 7 that the effect of *E* on *D* was underestimated under adjusting for mediator *M*. This
17
18 8 bias can have negative implication on the interpretation of effect of diabetes on
19
20 9 cardiovascular. The adjustment for mediator produced biased estimates, and thus
21
22 10 adjustment is inappropriate and should be avoided. A particular example was
23
24 11 adjustment for time-varying confounders which are also mediators using methods
25
26 12 including standardization, inverse-probability weighting, and G-estimation.³⁶ That is
27
28 13 to say, investigators should remember to consider biology and clinical information
29
30 14 when specifying a statistical model.

15 **Discussion**

16 In the paper, we dissected the sensitivity of bias to the variation of the effects of
17
18 17 exposure-mediator and mediator-outcome in adjusting for mediators under the
19
20 18 framework of logistic regression model. In the following four scenarios: a single
21
22 19 mediator (Figure 1a in scenario 1), two series mediators (Figure 1b in scenario 2), two
23
24 20 independent parallel (Figure 1c in scenario 3) or two correlated parallel mediators
25
26 21 (Figure 1d in scenario 4), the bias of varying across the effect of exposure-mediator
27
28 22 was greater than the one of varying across the effect mediator-outcome in adjusting
29
30 23 for the mediator (Figure 2, Figure 4, Figure 5 & Figure 6). However, in other two

1 scenarios: a single mediator or two independent parallel mediators in the presence of
2 unobserved confounders (Figure 1e in scenario 5 & Figure 1f in scenario 6), the
3 biases were higher sensitive to the variation of effect of mediator-outcome than effect
4 of exposure-mediator in adjusting for mediator (Figure 7 & Figure 8).

5 Conditioning on a mediator is of concern in all areas of epidemiologic
6 researches,^{13,19,37} it indeed lead to bias in estimating the total effect of the exposure on
7 the outcome.^{8,22-23} Mediators and confounders are indistinguishable in terms of
8 statistical association and conceptual grounds.³ Most of the studies focus on the
9 mediation effect analysis such as the calculation of direct effect and indirect
10 effect.^{20-21,38-41} Recently some authors used causal diagrams described how
11 appropriate handling of the matching variables. And they have proved that matching
12 on mediator M renders M and D independent (by design) in the matched study.
13 Matching on variable that are affected by the exposure and the outcome, or mediators
14 between the exposure and the outcome, would ordinary produce irremediable bias.
15 Furthermore, matching on mediator M blocks the causal path $E \rightarrow M \rightarrow D$ and thus
16 produces unfaithfulness for estimating the total effect E on D .^{31,42} Little effort has
17 been made to learn the biases performances in adjusting for mediator in estimating the
18 total effect of exposure on outcome. Our study results revealed that the biases were
19 higher sensitive to the variation of effects of exposure-mediator than effects of
20 mediator-outcome in adjusting for mediator in the absence of the unobserved
21 confounder in causal diagrams (Figure 1a, Figure 1b, Figure 1c & Figure 1d).
22 Nevertheless, for causal diagrams (Figure 1e & Figure 1f), the biases were higher
23 sensitive to the variation of effects of mediator-outcome than effects of
24 exposure-mediator in adjusting for mediator in the presence of the unobserved

1
2
3 1 confounder. Therefore, the biases of varying across different effects depended on the
4
5 2 causal diagrams framework whether there existed unobserved confounder.
6
7

8 3 The causal diagrams depicted in Figure 1 are indeed very simplistic and concise, as
9
10 4 they all exclude confounding factors of E and M as well as M and D . In practical
11
12 5 application, there exist some confounders in each pair of E , M , and D . Besides our
13
14 6 simulation study was not comprehensive enough to evaluate the bias performances in
15
16 7 adjusting for the mediator under logistic regression, because it only considered binary
17
18 8 variables, the certain scenarios of effect size and the common type of models. In
19
20 9 medical research, regression modeling is commonly used to adjust for covariates
21
22 10 associated with both the outcome and exposure. In this paper, the biases are defined
23
24 11 by the difference between M-adjusted and unadjusted ORs, some of which is
25
26 12 attributable to the non-collapsibility of OR. In the field of causal inference,
27
28 13 standardization and inverse-probability weighting may obtain the different bias
29
30 14 comparing with the regression modeling, and they may be better alternatives to
31
32 15 calculate bias⁴⁻⁵. Therefore, in future research, the methods of standardization and
33
34 16 inverse-probability weighting could be used to calculate the biases of this paper
35
36 17 definition. The work in the further ought to reinforce the mechanisms and conceptual
37
38 18 frameworks of confounder and mediator form causal diagrams so as to avoid falling
39
40 19 into analytic pitfalls.
41
42
43
44
45
46
47
48
49

50 **Conclusion**

51
52 21 In conclusion, the sensitivity of biases to the variation of the effects of
53
54 22 exposure-mediator and mediator-outcome were related to whether there is an
55
56 23 unobserved confounder in causal diagrams. The biases were higher sensitive to the
57
58
59
60

1 variation of effects of exposure-mediator than effects of mediator-outcome in
2 adjusting for mediator in the absence of unobserved confounders; while the biases
3 were higher sensitive to the variation of effects of mediator-outcome than effects of
4 exposure-mediator in the presence of unobserved confounder.

5 **Statements**

6 **Acknowledgments**

7 We would like to thank the anonymous reviewers and academic editors for providing
8 us with constructive comments and suggestions and also wish to acknowledge our
9 colleagues for their invaluable work.

10 **Authors' contributions**

11 TTW and HKL jointly conceived the idea behind the article and designed the study.
12 TTW helped conduct the literature review, performed the simulation and prepared the
13 first draft of the manuscript. PS, YYY, XRS, YL and ZSY participated in the design of
14 the study and the revision of the manuscript. FZX advised on critical revision of the
15 manuscript for important intellectual content. All authors read and approved the final
16 manuscript.

17 **Funding**

18 This work was supported by grants from the National Natural Science Foundation of
19 China (grant number 81573259)

20 **Competing interests**

21 The authors declare that they have no competing interests.

22 **Ethics approval and materials**

23 Ethics Committee of the School of Public Health (20140322), Shandong University.

24 Written informed consent was obtained from all participants.

25 **Provenance and peer review**

26 Not commissioned; externally peer reviewed.

1 Data sharing statement

2 No additional data are available.

3 Open Access

4 This is an Open Access article distributed in accordance with the Creative Commons
5 Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to
6 distribute, remix, adapt, build upon this work noncommercially, and license their
7 derivative works on different terms, provided the original work is properly cited and
8 the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

9 References

- 10 1 Bulterys M, Morgenstern H. Confounding or intermediate effect? An appraisal of
11 iatrogenic bias in perinatal AIDS research. *Paediatr Perinat Epidemiol*
12 1993;7:387–394.
- 13 2 Herna'n MA, Hernandez-Diaz S, Werler MM, et al. Causal knowledge as a
14 prerequisite for confounding evaluation: an application to birth defects
15 epidemiology. *Am J Epidemiol* 2002;155:176–184.
- 16 3 MacKinnon DP, Krull JL, and Lockwood CM. Equivalence of the mediation,
17 confounding and suppression effect. *Prevention Science* 2000;1:173–181.
- 18 4 Mansournia MA, Altman DG. Inverse probability weighting[J]. *Bmj*, 2016,
19 352:i189.
- 20 5 Gharibzadeh S, Mohammad K, Rahimiforoushani A, Amouzegar A, Mansournia
21 MA. Standardization as a tool for causal inference in medical research. *Arch Iran*
22 *Med*. 2016;19:666-70.
- 23 6 Howards PP, Schisterman EF, Poole C, et al. "Toward a clearer definition of
24 confounding" revisited with directed acyclic graphs, *Am J Epidemiol*
25 2012;176:506–511.
- 26 7 Williamson EJ, Aitken Z, Lawrie J, et al. Introduction to causal diagrams for
27 confounder selection. *Respirology*. 2014;19:303-311.
- 28 8 Wilcox AJ, Weinberg CR, Basso O. On the pitfalls of adjusting for gestational
29 age at birth. *Am J Epidemiol*. 2011;174:1062–1068.
- 30 9 Psaty BM, Koepsell TD, Lin D, et al. Assessment and control for confounding by
31 indication in observational studies. *J Am Geriatr Soc* 1999; 47:749-54.
- 32 10 Tilaki K H. Methodological issues of confounding in analytical epidemiology
33 studies. *Caspian J intern Med* 2012;3:488-495.
- 34 11 Kamangar F MD. Confounding variable in epidemiology studies: Basics and
35 beyond. *Arch Tran Med*. 2012;15:508-516.
- 36 12 Gaskins AJ, Schisterman EF. The effect of lipid adjustment on the analysis of
37 environmental contaminants and the outcome of human health risks. *Methods*
38 *Mol Biol* 2009;580:371-381.

- 1
2
3 13 Robins JM, Greenland S. Identifiability and exchangeability for direct and
4 indirect effects. *Epidemiology* 1992;3:143–155.
5
6 14 Pearl J. Direct and indirect effects. Seventeenth conference of uncertainty in
7 artificial intelligence, 2001. San Francisco, CA: Morgan Kaufmann, 2001.
8
9 15 Huang B, Sivaganesan S, Succop P, et al. Statistical assessment of mediational
10 effects for logistic mediational models. *Statist. Med* 2004;23:2713-2728.
11
12 16 Valeri L, Coull B A. Estimating causal contrasts involving intermediate variables
13 in the presence of selection bias [J]. *Statistics in Medicine*, 2016, 35(26):
14 4779-4793.
15
16 17 MacKinnon D P, Pirlott A G. Statistical approaches for enhancing causal
17 interpretation of the M to Y relation in mediation analysis [J]. *Personality and
18 Social Psychology Review*, 2015, 19(1): 30-43.
19
20 18 Imai K, Keele L, Tingley D. A general approach to causal mediation analysis.
21 *Psychol Methods*. 2010;15(4):309–334.
22
23 19 VanderWeele TJ. Bias formulas for sensitivity analysis for direct and indirect
24 effects. *Epidemiology*. 2010;21(4):540–551.
25
26 20 VanderWeele TJ, Vansteelandt S. Odds ratios for mediation analysis for a
27 dichotomous outcome. *Am J Epidemiol*. 2010;172(12):1339–1348.
28
29 21 Valeri L, VanderWeele TJ. Mediation analysis allowing for exposure-mediator
30 interactions and causal interpretation: theoretical assumptions and
31 implementation with SAS and SPSS macros [J]. *Psychological methods*, 2013,
32 18(2): 137.
33
34 22 Robins J. The control of confounding by intermediate variable. *Statistics in
35 Medicine*. 1989;8:679-701.
36
37 23 Schistermana EF, Coleb SR, Platt RW. Overadjustment Bias and Unnecessary
38 Adjustment in Epidemiologic Studies. *Epidemiology*. 2009;20:488–495.
39
40 24 Pearl J. Causal diagrams for empirical research. *Biometrika*. 1995;82:669-688.
41
42 25 Pearl J. Causality: models, reasoning, and inference. *Econometric Theory*.
43 2003;19: 675-685.
44
45 26 Joffe M, Gambhir M, Chadeau-Hyam M, et al. Causal diagrams in systems
46 epidemiology. *Emerg Themes Epidemiol*. 2012; 9:1-18.
47
48 27 Pearl, J. Causality: models, reasoning, and inference. Cambridge: Cambridge
49 University Press; 2000.
50
51 28 Pearl J. Causal inference in statistics: an overview. *Statistics Surveys*. 2009;
52 3:96–146.
53
54 29 Rubin DB. Should observational studies be designed to allow lack of balance in
55 covariate distributions across treatment groups? *Statistics in Medicine*.
56 2009;28:1420-1423.
57
58 30 Shpitser I. Counterfactual graphical models for longitudinal mediation analysis
59 with unobserved confounding. *Cognitive Science* 2013;37:1011-1035.
60
61 31 Mansournia MA, Hernán MA, Greenland S. Matched designs and causal
62 diagrams. *Int J Epidemiol*. 2013;42:860-869.
63
64 32 Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal
65 inference. *Stat Sci* 1999;14:29–46.

- 1
2
3 1 33 Mansournia MA, Greenland S. The relation of collapsibility and confounding to
4 2 faithfulness and stability. *Epidemiology*. 2015;26:466-72.
5 3 34 Clausen TD, Mathiesen ER, Hansen T, et al. Overweight and the metabolic
6 4 syndrome in adult offspring of women with diet-treated gestational diabetes
7 5 mellitus or type 1 diabetes.[J]. *J Clin Endocrinol Metab*, 2013, 94:2464-70.
8 6 35 Cheng JY, Wong VW, Tse YK, et al. Metabolic syndrome increases
9 7 cardiovascular events but not hepatic events and death in patients with chronic
10 8 hepatitis B[J]. *Hepatology*, 2016, 64:1507.
11 9 36 Mansournia MA, Danaei G, Forouzanfar MH, Mahmoudi M, Jamali M,
12 10 Mansournia N, Mohammad K. Effect of physical activity on functional
13 11 performance and knee pain in patients with osteoarthritis: analysis with marginal
14 12 structural models. *Epidemiology*. 2012;23:631-40.
15 13 37 VanderWeele TJ, Mumford SL, Schisterman EF. Conditioning on intermediates
16 14 in perinatal epidemiology. *Epidemiology*. 2012;23(1):1-9.
17 15 38 Lange T, Rasmussen M, Thygesen LC. Assessing natural direct and indirect
18 16 effects through multiple pathways. *American Journal of Epidemiology*. 2013;179,
19 17 513-518.
20 18 39 Shpitser, Ilya and VanderWeele, Tyler J. "A complete graphical criterion for the
21 19 adjustment formula in mediation analysis," *The International Journal of*
22 20 *Biostatistics*.2011;7,1-24.
23 21 40 Albert J M, Wang W. Sensitivity analyses for parametric causal mediation effect
24 22 estimation [J]. *Biostatistics*, 2015, 16(2): 339-351.
25 23 41 Wang W, Zhang B. Assessing natural direct and indirect effects for a continuous
26 24 exposure and a dichotomous outcome[J]. *Journal of Statistical Theory and*
27 25 *Practice*, 2016, 10(3): 574-587.
28 26 42 Li H, Yuan Z, Su P, et al. A simulation study on matched case-control designs in
29 27 the perspective of causal diagrams[J]. *BMC Med Res Methodol*, 2016,16,102 .
30 28

29 **Figure 1:** Six causal diagrams were designed for estimating the causal effect of E on
30 D . a) a single mediator M ; b) two series mediators M_1 and M_2 ; c) two independent
31 parallel mediators M_1 and M_2 ; d) two correlated parallel mediators M_1 and M_2 ; e) a
32 single mediator with an unobserved confounder U ; f) two independent parallel
33 mediators M_1 and M_2 with an unobserved confounder U .
34

35 **Figure 2:** The biases with the effects of $E \rightarrow M$ (red) and $M \rightarrow D$ (blue) increasing,
36 respectively. Comparison of the bias of different effects in adjustment mediator. The
37 OR of target effect (e.g. $E \rightarrow M$) from 1 to 10 given other effects fixed $\ln 2$ in Figure
38 2A. The OR of the effect of $M \rightarrow D$ from 1 to 10 with the effect of $E \rightarrow M$ being equal
39 to zero in Figure 2B (Color figure online).
40

41 **Figure 3:** Illustrating the use of positive and negative signs on edges $E \rightarrow M$, $M \rightarrow D$
42 and $E \rightarrow D$.
43

44 **Figure 4:** The biases with the effects of $E \rightarrow M_1$ (red), $M_1 \rightarrow M_2$ (blue) and $M_2 \rightarrow D$

(black) increasing, respectively. Comparison of the bias of different effects in three adjustment models: A) adjustment for M_1 , B) adjustment for M_2 and C) adjustment for M_1 and M_2 . The OR of target effect (e.g. $E \rightarrow M_1$) from 1 to 10 given the effect of $M_1 \rightarrow M_2$ fixed $\ln 8$ and other effects fixed $\ln 2$ in Figure 4 (Color figure online).

Figure 5: The biases with the effects of $E \rightarrow M_1$ (red), $E \rightarrow M_2$ (blue), $M_1 \rightarrow D$ (black) and $M_2 \rightarrow D$ (green) increasing, respectively. Comparison of the bias of different effects in three adjustment models: A) adjustment for M_1 , B) adjustment for M_2 and C) adjustment for M_1 and M_2 . The OR of target effects (e.g. $E \rightarrow M_1$) from 1 to 10 given other edges effects fixed $\ln 2$ in Figure 5 (Color figure online).

Figure 6: The biases with the effects of $E \rightarrow M_1$ (red), $E \rightarrow M_2$ (blue), $M_1 \rightarrow D$ (black), $M_2 \rightarrow D$ (green) and the effect of $M_2 \rightarrow M_1$ (purple) increasing, respectively. Comparison of the bias of different effects in three adjustment models: A) adjustment for M_1 , B) adjustment for M_2 and C) adjustment for M_1 and M_2 . The OR of target effects (e.g. $E \rightarrow M_1$) from 1 to 10 given other effects fixed $\ln 2$ in Figure 6 (Color figure online).

Figure 7: The biases with the effects of $E \rightarrow M$ (red) and $M \rightarrow D$ (blue) respectively. Comparison of the bias of different effects in adjustment mediator M . The OR of target effects (e.g. $E \rightarrow M$) from 1 to 10 given the effects of causal edges fixed $\ln 2$ and the effect of confounder edges fixed $\ln 5$ in Figure 7 (Color figure online).

Figure 8: The biases with the effects of $E \rightarrow M_1$ (red), $E \rightarrow M_2$ (blue), $M_1 \rightarrow D$ (black) and $M_2 \rightarrow D$ (green) respectively. Comparison of the bias of different effects in three adjustment models: A) adjustment for M_1 , B) adjustment for M_2 , and C) adjustment for M_1 and M_2 . The OR of target effects (e.g. $E \rightarrow M_1$) from 1 to 10 given the effects of causal edges fixed $\ln 2$ and the effect of confounder edges fixed $\ln 5$ in Figure 8 (Color figure online).

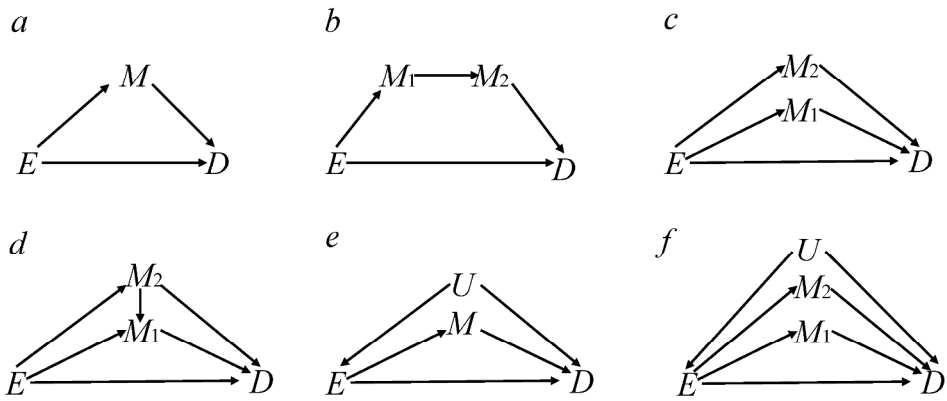


Figure 1: Six causal diagrams were designed for estimating the causal effect of E on D .

252x110mm (300 x 300 DPI)

Peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

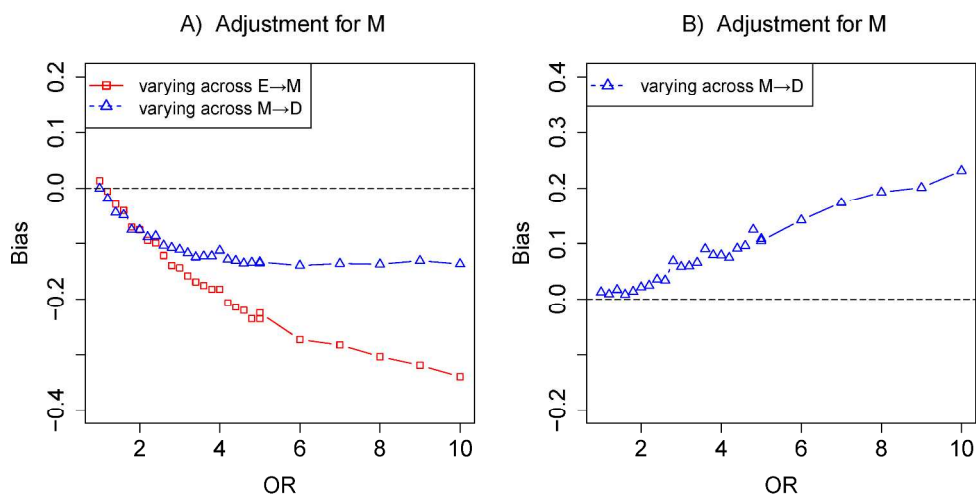


Figure 2 : The biases with the effects E→M (red) and M→D (blue) increasing, respectively.

281x148mm (300 x 300 DPI)

review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

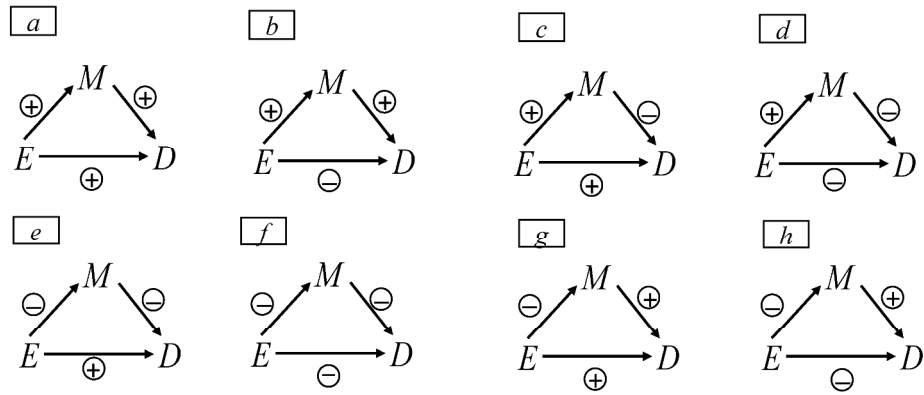


Figure 3: Illustrating the use of positive and negative signs on edges $E \rightarrow M$, $M \rightarrow D$ and $E \rightarrow D$.

237x106mm (300 x 300 DPI)

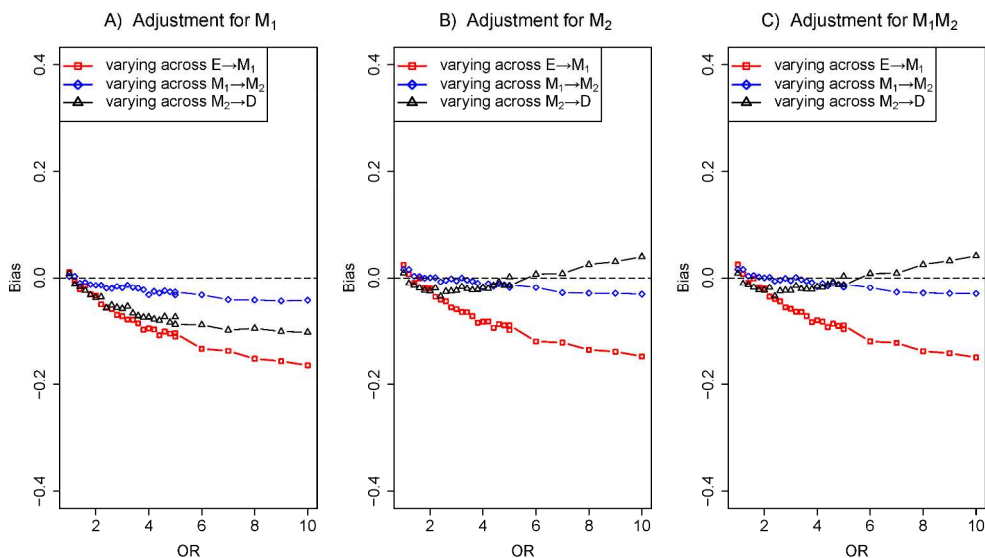


Figure 4: The biases with the effects E→M₁ (red), M₁→M₂ (blue) and M₂→D (black) increasing, respectively.

270x155mm (300 x 300 DPI)

Review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

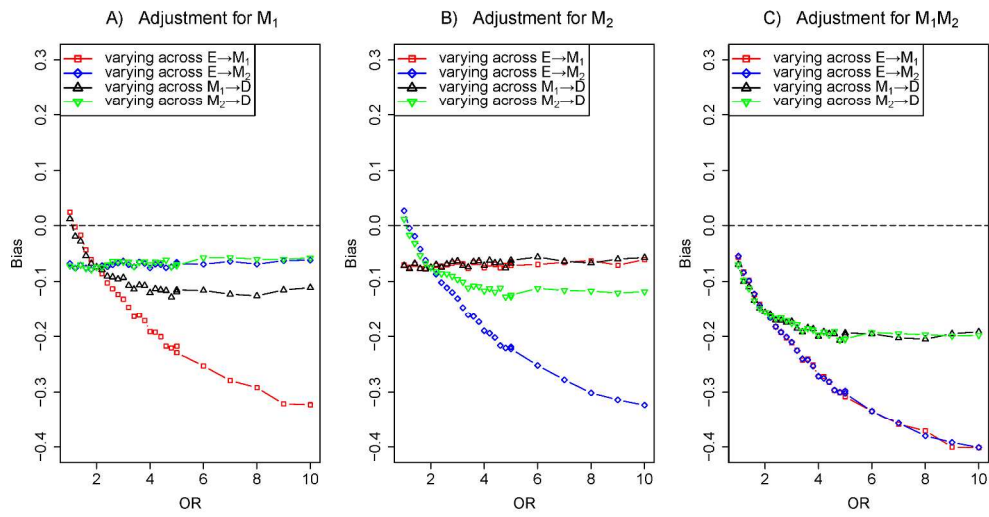


Figure 5 : The biases with the effects $E \rightarrow M_1$ (red), $E \rightarrow M_2$ (blue), $M_1 \rightarrow D$ (black) and $M_2 \rightarrow D$ (green) increasing, respectively.

279x147mm (300 x 300 DPI)

review only

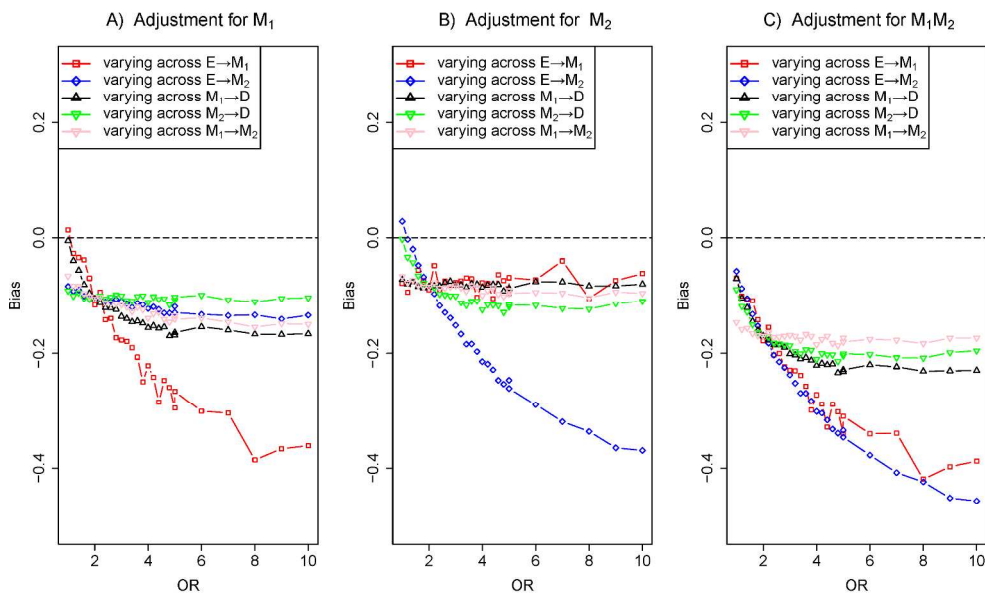


Figure 6: The biases with the effects $E \rightarrow M_1$ (red), $E \rightarrow M_2$ (blue), $M_1 \rightarrow D$ (black), $M_2 \rightarrow D$ (green) and the effect $M_2 \rightarrow M_1$ (purple) increasing, respectively.

278x169mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

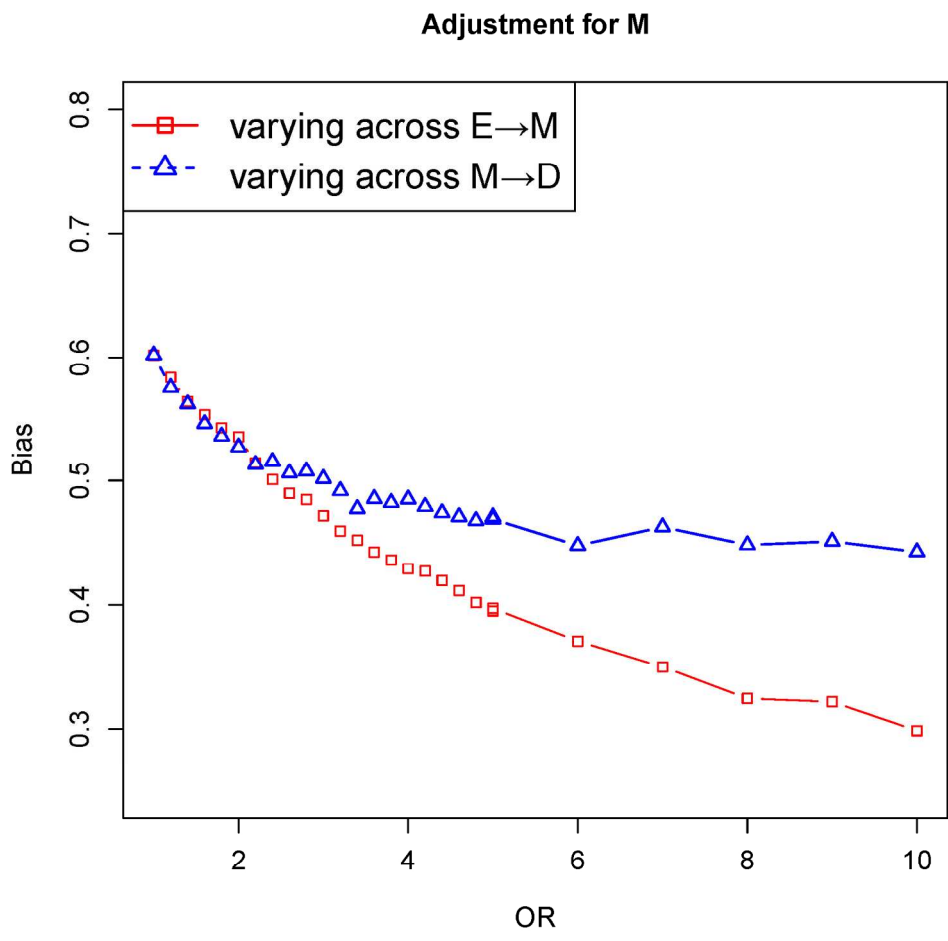


Figure 7: The biases with the effects E→M (red) and M→D (blue) respectively.

177x177mm (300 x 300 DPI)

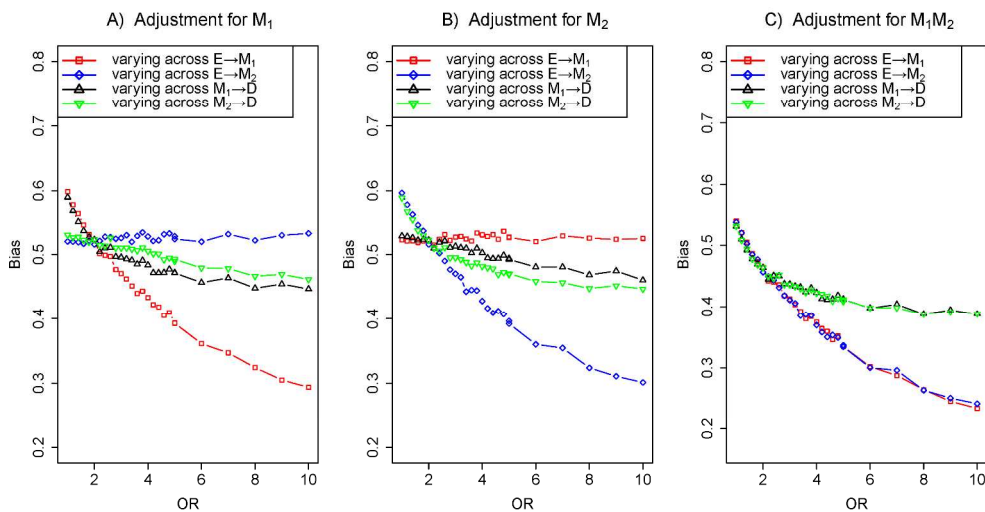


Figure 8 : The biases with the effects E→M₁ (red), E→M₂ (blue), M₁→D (black) and M₂→D (green) respectively.

281x148mm (300 x 300 DPI)

review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Appendix:

The effect of adjusting for mediator was biased for estimating the total effect of exposure on outcome using logistic regression model. Theoretical derivation of Figure 1a as follow:

Suppose the logistic models among E , M and D are:

$$\text{logit}\{P(D=1|e,m)\} = \alpha_1 + \beta_0 e + \beta_2 m,$$

$$\text{logit}\{P(M=1|e)\} = \alpha_0 + \beta_1 e.$$

The total effect ($\beta_{E \rightarrow D}^{TE}$) of exposure E on outcome D on the odds ratio ($OR_{E \rightarrow D}^{TE}$) scale was equal to

$$\begin{aligned} \beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\ &= \log \left\{ \frac{P(D_e = 1) / \{1 - P(D_e = 1)\}}{P(D_{e^*} = 1) / \{1 - P(D_{e^*} = 1)\}} \right\} \\ &= \log \left\{ \frac{P(D_e = 1) \times \{1 - P(D_{e^*} = 1)\}}{\{1 - P(D_e = 1)\} \times P(D_{e^*} = 1)} \right\} \\ &= \log \left\{ \frac{P(D=1|e=1) \times P(D=0|e^*=0)}{P(D=0|e=1) \times P(D=1|e^*=0)} \right\} \\ &= \log \left\{ \frac{\left[\sum_m P(D=1|e=1,m)P(m|e=1) \right] \times \left[\sum_m P(D=0|e^*=0,m)P(m|e^*=0) \right]}{\left[\sum_m P(D=0|e=1,m)P(m|e=1) \right] \times \left[\sum_m P(D=1|e^*=0,m)P(m|e^*=0) \right]} \right\} \end{aligned}$$

The effect ($\beta_{ED|M}(m)$) of adjusting for mediator M by logistic regression model is given

$$\begin{aligned} \beta_{ED|M}(m) &= \text{logit}\{P(D=1|e=1,m)\} - \text{logit}\{P(D=1|e^*=0,m)\} \\ &= \log \left\{ \frac{P(D=1|e=1,m) \times P(D=0|e^*=0,m)}{P(D=0|e=1,m) \times P(D=1|e^*=0,m)} \right\} \\ &= \beta_0 \end{aligned}$$

Therefore,

$$\begin{aligned}
 \text{bias} &= \beta_0 - \log(OR_{E \rightarrow D}^{TE}) \\
 &= \log \left\{ \frac{\exp(\beta_0)}{\exp(\beta_0) \frac{\exp(\beta_2) \times A_1 + \exp(\beta_2) \times B_1 + C_1 + D_1}{\exp(\beta_2) \times A_1 + B_1 + \exp(\beta_2) \times C_1 + D_1}} \right\} \\
 &= \log \left\{ \frac{\exp(\beta_2) \times A_1 + B_1 + \exp(\beta_2) \times C_1 + D_1}{\exp(\beta_2) \times A_1 + \exp(\beta_2) \times B_1 + C_1 + D_1} \right\}
 \end{aligned}$$

where

$$A_1 = \exp(\beta_1 + \alpha_0) \times (1 + \exp(\beta_0 + \alpha_1)) \times \exp(\alpha_0) \times (1 + \exp(\alpha_1))$$

$$B_1 = \exp(\beta_1 + \alpha_0) \times (1 + \exp(\beta_0 + \alpha_1)) \times (1 + \exp(\beta_2 + \alpha_1))$$

$$C_1 = (1 + \exp(\beta_0 + \beta_2 + \alpha_1)) \times \exp(\alpha_0) \times (1 + \exp(\alpha_1))$$

$$D_1 = (1 + \exp(\beta_0 + \beta_2 + \alpha_1)) \times (1 + \exp(\beta_2 + \alpha_1))$$

Focusing on the difference of between $\exp(\beta_2) \times B_1 + C_1$ and $B_1 + \exp(\beta_2) \times C_1$.

$$\begin{aligned}
 T(\beta_1) &= \exp(\beta_2) \times B_1 + C_1 - (B_1 + \exp(\beta_2) \times C_1) \\
 &= \exp(\beta_2) \times (B_1 - C_1) - (B_1 - C_1) \\
 &= (\exp(\beta_2) - 1) \times (B_1 - C_1) \\
 &= (\exp(\beta_2) - 1) \times (\exp(\beta_1 + \alpha_0) \times (1 + \exp(\beta_0 + \alpha_1)) \times (1 + \exp(\beta_2 + \alpha_1)) \\
 &\quad - (1 + \exp(\beta_0 + \beta_2 + \alpha_1)) \times \exp(\alpha_0) \times (1 + \exp(\alpha_1))) \\
 &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \times [\exp(\beta_1) \times (1 + \exp(\beta_0 + \alpha_1)) \times (1 + \exp(\beta_2 + \alpha_1)) \\
 &\quad - (1 + \exp(\beta_0 + \beta_2 + \alpha_1)) \times (1 + \exp(\alpha_1))]
 \end{aligned}$$

Then, detailed dissection:

1: $\beta_2 = 0$, $\text{bias} = 0$.

2: $\beta_2 > 0$,

① $\beta_1 = 0$: (i) $\beta_0 = 0$, $\text{bias} = 0$; (ii) $\beta_0 > 0$, $\text{bias} > 0$; (iii) $\beta_0 < 0$, $\text{bias} < 0$.

② $\beta_1 < 0$: (i) $\beta_0 = 0$, $\text{bias} > 0$; (ii) $\beta_0 > 0$, $\text{bias} > 0$; (iii) $\beta_0 < 0$, $\text{bias} > 0$.

proof (iii)

$$\begin{aligned}
 T(\beta_1) &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\
 &\quad \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\
 &\quad - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \}
 \end{aligned}$$

when $\beta_0 < 0$ and $\beta_2 > 0 \Rightarrow \exp(\beta_0) - 1 < 0 \quad \exp(\beta_2) - 1 > 0$

According to $(a-1)(b-1) = ab - a - b + 1$, when $(a-1)(b-1) < 0 \Rightarrow ab + 1 < a + b$

$$\begin{aligned}
& 1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\
& < 1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\
& \Rightarrow \exp(\beta_0 + \beta_2) + 1 < \exp(\beta_0) + \exp(\beta_2)
\end{aligned}$$

when

$$\beta_1 < \log \left\{ \frac{\exp(\beta_0 + \beta_2) + 1}{\exp(\beta_0) + \exp(\beta_2)} \right\} < 0$$

$$\beta_1 < \log \left\{ \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} \right\} < 0$$

$$\Rightarrow \exp(\beta_1) < \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} < 1$$

$$\begin{aligned}
\Rightarrow T(\beta_1) &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\
&\quad \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\
&\quad - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \} \\
&< 0
\end{aligned}$$

Therefore, when $\beta_2 > 0$, $\beta_1 < 0$, $\beta_0 < 0$, then $bias > 0$.

③ $\beta_1 > 0$: (i) $\beta_0 = 0$, $bias < 0$; (ii) $\beta_0 < 0$, $bias < 0$; (iii) $\beta_0 > 0$, $bias < 0$.

proof (iii)

$$\begin{aligned}
T(\beta_1) &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\
&\quad \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\
&\quad - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \}
\end{aligned}$$

when $\beta_0 > 0$ and $\beta_2 > 0 \Rightarrow \exp(\beta_0) - 1 > 0 \quad \exp(\beta_2) - 1 > 0$

According to $(a-1)(b-1) = ab - a - b + 1$, when $ab > 0 \Rightarrow ab + 1 > a + b$

$$\begin{aligned}
& 1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\
& > 1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\
& \Rightarrow \exp(\beta_0 + \beta_2) + 1 > \exp(\beta_0) + \exp(\beta_2)
\end{aligned}$$

when

$$\beta_1 > \log \left\{ \frac{\exp(\beta_0 + \beta_2) + 1}{\exp(\beta_0) + \exp(\beta_2)} \right\} > 0$$

$$\beta_1 > \log \left\{ \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} \right\} > 0$$

$$\Rightarrow \exp(\beta_1) > \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} > 1$$

$$\begin{aligned} \Rightarrow T(\beta_1) &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\ &\quad \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\ &\quad - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \} \\ &> 0 \end{aligned}$$

Therefore, when $\beta_2 > 0$, $\beta_1 > 0$, $\beta_0 > 0$, then *bias* < 0.

3: $\beta_2 < 0$,

① $\beta_1 = 0$: (i) $\beta_0 = 0$, *bias* = 0; (ii) $\beta_0 > 0$, *bias* > 0; (iii) $\beta_0 < 0$, *bias* < 0.

② $\beta_1 < 0$: (i) $\beta_0 = 0$, *bias* < 0; (ii) $\beta_0 < 0$, *bias* < 0; (iii) $\beta_0 > 0$, *bias* < 0.

proof (iii)

$$\begin{aligned} T(\beta_1) &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\ &\quad \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\ &\quad - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \} \end{aligned}$$

when $\beta_0 > 0$ and $\beta_2 < 0 \Rightarrow \exp(\beta_0) - 1 > 0$ $\exp(\beta_2) - 1 < 0$

According to $(a-1)(b-1) = ab - a - b + 1$, when $ab < 0 \Rightarrow ab + 1 < a + b$

$$\begin{aligned} &1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\ &< 1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\ &\Rightarrow \exp(\beta_0 + \beta_2) + 1 < \exp(\beta_0) + \exp(\beta_2) \end{aligned}$$

when

$$\beta_1 < \log \left\{ \frac{\exp(\beta_0 + \beta_2) + 1}{\exp(\beta_0) + \exp(\beta_2)} \right\} < 0$$

$$\beta_1 < \log \left\{ \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} \right\} < 0$$

$$\Rightarrow \exp(\beta_1) < \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} < 1$$

$$\begin{aligned} \Rightarrow T(\beta_1) &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\ &\quad \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\ &\quad - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \} \\ &> 0 \end{aligned}$$

Therefore, when $\beta_2 < 0$, $\beta_1 < 0$, $\beta_0 > 0$, then *bias* < 0.

③ $\beta_1 > 0$: (i) $\beta_0 = 0$, *bias* > 0; (ii) $\beta_0 > 0$, *bias* > 0; (iii) $\beta_0 < 0$, *bias* > 0.

proof (iii)

$$T(\beta_1) = (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\ \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\ - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \}$$

when $\beta_0 < 0$ and $\beta_2 < 0 \Rightarrow \exp(\beta_0) - 1 < 0 \quad \exp(\beta_2) - 1 < 0$

According to $(a-1)(b-1) = ab - a - b + 1$, when $ab > 0 \Rightarrow ab + 1 > a + b$

$$1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\ > 1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\ \Rightarrow \exp(\beta_0 + \beta_2) + 1 > \exp(\beta_0) + \exp(\beta_2)$$

when

$$\beta_1 > \log \left\{ \frac{\exp(\beta_0 + \beta_2) + 1}{\exp(\beta_0) + \exp(\beta_2)} \right\} > 0$$

$$\beta_1 > \log \left\{ \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} \right\} > 0$$

$$\Rightarrow \exp(\beta_1) > \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} > 1$$

$$\Rightarrow T(\beta_1) = (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\ \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\ - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \} \\ < 0$$

Therefore, when $\beta_2 < 0$, $\beta_1 > 0$, $\beta_0 < 0$, then $bias > 0$.

In conclusion:

1: $\beta_2 = 0$, $bias = 0$.

2: $\beta_2 \neq 0$, $\beta_1 = 0$: (i) $\beta_0 = 0$, $bias = 0$; (ii) $\beta_0 > 0$, $bias > 0$; (iii) $\beta_0 < 0$, $bias < 0$.

3: (i) $\beta_1\beta_2 > 0$, $bias < 0$. (ii) $\beta_1\beta_2 < 0$, $bias > 0$.

Supplementary A

The theoretical results of others causal diagrams (Figure 1b-Figure 1f) have been shown in the supplementary of manuscript.

(1) Figure 1(b) is a depiction through two series mediators, decomposing total effects into direct ($E \rightarrow D$) and indirect ($E \rightarrow M_1 \rightarrow M_2 \rightarrow D$) components.

On the odds ratio ($OR_{E \rightarrow D}^{TE}$) scale, the total effect ($\beta_{E \rightarrow D}^{TE} = \log(OR_{E \rightarrow D}^{TE})$), comparing exposure level e with e^* , we could obtain the total effect:

$$\begin{aligned}\beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\ &= \log \left\{ \frac{P(D_e = 1) / \{1 - P(D_e = 1)\}}{P(D_{e^*} = 1) / \{1 - P(D_{e^*} = 1)\}} \right\} \\ &= \log \left\{ \frac{P(D_e = 1) \times \{1 - P(D_{e^*} = 1)\}}{\{1 - P(D_e = 1)\} \times P(D_{e^*} = 1)} \right\} \\ &= \log \left\{ \frac{P(D = 1 | e = 1) \times P(D = 0 | e^* = 0)}{P(D = 0 | e = 1) \times P(D = 1 | e^* = 0)} \right\} \\ &= \log \left\{ \frac{\xi_1}{\xi_2} \right\}\end{aligned}$$

$$\xi_1 = \left[\sum_{m_1 m_2} P(D = 1 | e = 1, m_2) P(m_2 | m_1) P(m_1 | e = 1) \right] \times \left[\sum_{m_1 m_2} P(D = 0 | e^* = 0, m_2) P(m_2 | m_1) P(m_1 | e^* = 0) \right]$$

$$\xi_2 = \left[\sum_{m_1 m_2} P(D = 0 | e = 1, m_2) P(m_2 | m_1) P(m_1 | e = 1) \right] \times \left[\sum_{m_1 m_2} P(D = 1 | e^* = 0, m_2) P(m_2 | m_1) P(m_1 | e^* = 0) \right]$$

The effect ($\beta_{ED|M_1}(m_1)$) of adjusting for mediator M_1 by logistic regression model can

be given

$$\begin{aligned}\beta_{ED|M_1}(m_1) &= \text{logit}\{P(D = 1 | e = 1, m_1)\} - \text{logit}\{P(D = 1 | e^* = 0, m_1)\} \\ &= \log \left\{ \frac{P(D = 1 | e = 1, m_1) P(D = 0 | e^* = 0, m_1)}{P(D = 0 | e = 1, m_1) P(D = 1 | e^* = 0, m_1)} \right\} \\ &= \log \left\{ \frac{\left[\sum_{m_2} P(D = 1 | e = 1, m_2) P(m_2 | m_1) \right] \times \left[\sum_{m_2} P(D = 0 | e^* = 0, m_2) P(m_2 | m_1) \right]}{\left[\sum_{m_2} P(D = 0 | e = 1, m_2) P(m_2 | m_1) \right] \times \left[\sum_{m_2} P(D = 1 | e^* = 0, m_2) P(m_2 | m_1) \right]} \right\}\end{aligned}$$

The effect ($\beta_{ED|M_2}(m_2)$) of adjusting for mediator M_2 by logistic regression model can

be given

$$\begin{aligned}\beta_{ED|M_2}(m_2) &= \text{logit}\{P(D = 1 | e = 1, m_2)\} - \text{logit}\{P(D = 1 | e^* = 0, m_2)\} \\ &= \log \left\{ \frac{P(D = 1 | e = 1, m_2) P(D = 0 | e^* = 0, m_2)}{P(D = 0 | e = 1, m_2) P(D = 1 | e^* = 0, m_2)} \right\}\end{aligned}$$

The effect ($\beta_{ED|M_1, M_2}(m_1, m_2)$) of adjusting for mediator M_1 M_2 by logistic regression model can be given

$$\begin{aligned}\beta_{ED|M_1, M_2}(m_1, m_2) &= \text{logit}\{P(D=1|e=1, m_1, m_2)\} - \text{logit}\{P(D=1|e^*=0, m_1, m_2)\} \\ &= \log\left\{\frac{P(D=1|e=1, m_1, m_2)P(D=0|e^*=0, m_1, m_2)}{P(D=0|e=1, m_1, m_2)P(D=1|e^*=0, m_1, m_2)}\right\} \\ &= \log\left\{\frac{P(D=1|e=1, m_2)P(D=0|e^*=0, m_2)}{P(D=0|e=1, m_2)P(D=1|e^*=0, m_2)}\right\}\end{aligned}$$

Therefore, we could evaluate the biases that contains three adjustment models: A) adjustment for M_1 , $\text{bias}(m_1) = \beta_{ED|M_1}(m_1) - \beta_{E \rightarrow D}^{TE}$; B) adjustment for M_2 , $\text{bias}(m_2) = \beta_{ED|M_2}(m_2) - \beta_{E \rightarrow D}^{TE}$ and C) adjustment for M_1 and M_2 , $\text{bias}(m_1, m_2) = \beta_{ED|M_1, M_2}(m_1, m_2) - \beta_{E \rightarrow D}^{TE}$.

(2) Figure 1c shows that the exposure E independently causes M_1 and M_2 and indirectly influences the outcome D through M_1 and M_2 , forming three causal paths $E \rightarrow D$, $E \rightarrow M_1 \rightarrow D$ and $E \rightarrow M_2 \rightarrow D$.

On the odds ratio ($OR_{E \rightarrow D}^{TE}$) scale, the total effect ($\beta_{E \rightarrow D}^{TE} = \log(OR_{E \rightarrow D}^{TE})$), comparing exposure level e with e^* , we could obtain the total effect:

$$\begin{aligned}\beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\ &= \log\left\{\frac{P(D_e=1)/\{1-P(D_e=1)\}}{P(D_{e^*}=1)/\{1-P(D_{e^*}=1)\}}\right\} \\ &= \log\left\{\frac{P(D_e=1) \times \{1-P(D_{e^*}=1)\}}{\{1-P(D_e=1)\} \times P(D_{e^*}=1)}\right\} \\ &= \log\left\{\frac{P(D=1|e=1) \times P(D=0|e^*=0)}{P(D=0|e=1) \times P(D=1|e^*=0)}\right\} \\ &= \log\left\{\frac{\xi_1}{\xi_2}\right\}\end{aligned}$$

$$\xi_1 = \left[\sum_{m_1, m_2} P(D=1|e=1, m_1, m_2)P(m_2|e=1)P(m_1|e=1) \right] \times \left[\sum_{m_1, m_2} P(D=0|e^*=0, m_1, m_2)P(m_2|e^*=0)P(m_1|e^*=0) \right]$$

$$\xi_2 = \left[\sum_{m_1, m_2} P(D=0|e=1, m_1, m_2)P(m_2|e=1)P(m_1|e=1) \right] \times \left[\sum_{m_1, m_2} P(D=1|e^*=0, m_1, m_2)P(m_2|e^*=0)P(m_1|e^*=0) \right]$$

The effect ($\beta_{ED|M_1}(m_1)$) of adjusting for mediator M_1 by logistic regression model can be given

$$\begin{aligned} \beta_{ED|M_1}(m_1) &= \text{logit}\{P(D=1|e=1,m_1)\} - \text{logit}\{P(D=1|e^*=0,m_1)\} \\ &= \log\left\{\frac{P(D=1|e=1,m_1)P(D=0|e^*=0,m_1)}{P(D=0|e=1,m_1)P(D=1|e^*=0,m_1)}\right\} \\ &= \log\left\{\frac{\left[\sum_{m_2} P(D=1|e=1,m_1,m_2)P(m_2|e=1)\right] \times \left[\sum_{m_2} P(D=0|e^*=0,m_1,m_2)P(m_2|e^*=0)\right]}{\left[\sum_{m_2} P(D=0|e=1,m_1,m_2)P(m_2|e=1)\right] \times \left[\sum_{m_2} P(D=1|e^*=0,m_1,m_2)P(m_2|e^*=0)\right]}\right\} \end{aligned}$$

The effect ($\beta_{ED|M_2}(m_2)$) of adjusting for mediator M_2 by logistic regression model can

be given

$$\begin{aligned} \beta_{ED|M_2}(m_2) &= \text{logit}\{P(D=1|e=1,m_2)\} - \text{logit}\{P(D=1|e^*=0,m_2)\} \\ &= \log\left\{\frac{P(D=1|e=1,m_2)P(D=0|e^*=0,m_2)}{P(D=0|e=1,m_2)P(D=1|e^*=0,m_2)}\right\} \\ &= \log\left\{\frac{\left[\sum_{m_1} P(D=1|e=1,m_1,m_2)P(m_1|e=1)\right] \times \left[\sum_{m_1} P(D=0|e^*=0,m_1,m_2)P(m_1|e^*=0)\right]}{\left[\sum_{m_1} P(D=0|e=1,m_1,m_2)P(m_1|e=1)\right] \times \left[\sum_{m_1} P(D=1|e^*=0,m_1,m_2)P(m_1|e^*=0)\right]}\right\} \end{aligned}$$

The effect ($\beta_{ED|M_1,M_2}(m_1,m_2)$) of adjusting for mediator $M_1 M_2$ by logistic regression

model can be given

$$\begin{aligned} \beta_{ED|M_1,M_2}(m_1,m_2) &= \text{logit}\{P(D=1|e=1,m_1,m_2)\} - \text{logit}\{P(D=1|e^*=0,m_1,m_2)\} \\ &= \log\left\{\frac{P(D=1|e=1,m_1,m_2)P(D=0|e^*=0,m_1,m_2)}{P(D=0|e=1,m_1,m_2)P(D=1|e^*=0,m_1,m_2)}\right\} \end{aligned}$$

Therefore, we could evaluate the biases that contains three adjustment models: A)

adjustment for M_1 , $bias(m_1) = \beta_{ED|M_1}(m_1) - \beta_{E \rightarrow D}^{TE}$; B) adjustment for M_2 ,

$bias(m_2) = \beta_{ED|M_2}(m_2) - \beta_{E \rightarrow D}^{TE}$ and C) adjustment for M_1 and M_2 ,

$bias(m_1,m_2) = \beta_{ED|M_1,M_2}(m_1,m_2) - \beta_{E \rightarrow D}^{TE}$.

(3) In Figure 1d, there exists five paths from E to D : $E \rightarrow D$, $E \rightarrow M_1 \rightarrow D$, $E \rightarrow M_2 \rightarrow D$, $E \rightarrow M_1 \leftarrow M_2 \rightarrow D$ and $E \rightarrow M_2 \rightarrow M_1 \rightarrow D$. In particular, the path $E \rightarrow M_1 \leftarrow M_2 \rightarrow D$ is a blocked path, due to the M_1 being a collider node.

On the odds ratio ($OR_{E \rightarrow D}^{TE}$) scale, the total effect ($\beta_{E \rightarrow D}^{TE} = \log(OR_{E \rightarrow D}^{TE})$), comparing exposure level e with e^* , we could obtain the total effect:

$$\begin{aligned}
\beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\
&= \log \left\{ \frac{P(D_e = 1) / \{1 - P(D_e = 1)\}}{P(D_{e^*} = 1) / \{1 - P(D_{e^*} = 1)\}} \right\} \\
&= \log \left\{ \frac{P(D_e = 1) \times \{1 - P(D_{e^*} = 1)\}}{\{1 - P(D_e = 1)\} \times P(D_{e^*} = 1)} \right\} \\
&= \log \left\{ \frac{P(D = 1 | e = 1) \times P(D = 0 | e^* = 0)}{P(D = 0 | e = 1) \times P(D = 1 | e^* = 0)} \right\} \\
&= \log \left\{ \frac{\xi_1}{\xi_2} \right\}
\end{aligned}$$

$$\begin{aligned}
\xi_1 &= \left[\sum_{m_1, m_2} P(D = 1 | e = 1, m_1, m_2) P(m_2 | e = 1) P(m_1 | e = 1, m_2) \right] \\
&\quad \times \left[\sum_{m_1, m_2} P(D = 0 | e^* = 0, m_1, m_2) P(m_2 | e^* = 0) P(m_1 | e^* = 0, m_2) \right] \\
\xi_2 &= \left[\sum_{m_1, m_2} P(D = 0 | e = 1, m_1, m_2) P(m_2 | e = 1) P(m_1 | e = 1, m_2) \right] \\
&\quad \times \left[\sum_{m_1, m_2} P(D = 1 | e^* = 0, m_1, m_2) P(m_2 | e^* = 0) P(m_1 | e^* = 0, m_2) \right]
\end{aligned}$$

The effect ($\beta_{ED|M_1}(m_1)$) of adjusting for mediator M_1 by logistic regression model can

be given

$$\begin{aligned}
\beta_{ED|M_1}(m_1) &= \text{logit}\{P(D = 1 | e = 1, m_1)\} - \text{logit}\{P(D = 1 | e^* = 0, m_1)\} \\
&= \log \left\{ \frac{P(D = 1 | e = 1, m_1) P(D = 0 | e^* = 0, m_1)}{P(D = 0 | e = 1, m_1) P(D = 1 | e^* = 0, m_1)} \right\} \\
&= \log \left\{ \frac{\left[\sum_{m_2} P(D = 1 | e = 1, m_1, m_2) P(m_2 | e = 1, m_1) \right] \times \left[\sum_{m_2} P(D = 0 | e^* = 0, m_1, m_2) P(m_2 | e^* = 0, m_1) \right]}{\left[\sum_{m_2} P(D = 0 | e = 1, m_1, m_2) P(m_2 | e = 1, m_1) \right] \times \left[\sum_{m_2} P(D = 1 | e^* = 0, m_1, m_2) P(m_2 | e^* = 0, m_1) \right]} \right\} \\
&= \log \left\{ \frac{\xi_1}{\xi_2} \right\}
\end{aligned}$$

$$\begin{aligned}
\xi_1 &= \left[\sum_{m_2} P(D = 1 | e = 1, m_1, m_2) \frac{P(m_1 | e = 1, m_2) P(m_2 | e = 1)}{\sum_{m_2} P(m_1 | e = 1, m_2) P(m_2 | e = 1)} \right] \\
&\quad \times \left[\sum_{m_2} P(D = 0 | e^* = 0, m_1, m_2) \frac{P(m_1 | e^* = 0, m_2) P(m_2 | e^* = 0)}{\sum_{m_2} P(m_1 | e^* = 0, m_2) P(m_2 | e^* = 0)} \right] \\
\xi_2 &= \left[\sum_{m_2} P(D = 0 | e = 1, m_1, m_2) \frac{P(m_1 | e = 1, m_2) P(m_2 | e = 1)}{\sum_{m_2} P(m_1 | e = 1, m_2) P(m_2 | e = 1)} \right] \\
&\quad \times \left[\sum_{m_2} P(D = 1 | e^* = 0, m_1, m_2) \frac{P(m_1 | e^* = 0, m_2) P(m_2 | e^* = 0)}{\sum_{m_2} P(m_1 | e^* = 0, m_2) P(m_2 | e^* = 0)} \right]
\end{aligned}$$

The effect ($\beta_{ED|M_2}(m_2)$) of adjusting for mediator M_2 by logistic regression model can

be given

$$\begin{aligned}
\beta_{ED|M_2}(m_2) &= \text{logit}\{P(D=1|e=1,m_2)\} - \text{logit}\{P(D=1|e^*=0,m_2)\} \\
&= \log \left\{ \frac{P(D=1|e=1,m_2)P(D=0|e^*=0,m_2)}{P(D=0|e=1,m_2)P(D=1|e^*=0,m_2)} \right\} \\
&= \log \left\{ \frac{\left[\sum_{m_1} P(D=1|e=1,m_1,m_2)P(m_1|e=1,m_2) \right] \times \left[\sum_{m_1} P(D=0|e^*=0,m_1,m_2)P(m_1|e^*=0,m_2) \right]}{\left[\sum_{m_1} P(D=0|e=1,m_1,m_2)P(m_1|e=1,m_2) \right] \times \left[\sum_{m_1} P(D=1|e^*=0,m_1,m_2)P(m_1|e^*=0,m_2) \right]} \right\}
\end{aligned}$$

The effect ($\beta_{ED|M_1,M_2}(m_1,m_2)$) of adjusting for mediator M_1 M_2 by logistic regression model can be given

$$\begin{aligned}
\beta_{ED|M_1,M_2}(m_1,m_2) &= \text{logit}\{P(D=1|e=1,m_1,m_2)\} - \text{logit}\{P(D=1|e^*=0,m_1,m_2)\} \\
&= \log \left\{ \frac{P(D=1|e=1,m_1,m_2)P(D=0|e^*=0,m_1,m_2)}{P(D=0|e=1,m_1,m_2)P(D=1|e^*=0,m_1,m_2)} \right\}
\end{aligned}$$

Therefore, we could evaluate the biases that contains three adjustment models: A) adjustment for M_1 , $\text{bias}(m_1) = \beta_{ED|M_1}(m_1) - \beta_{E \rightarrow D}^{TE}$; B) adjustment for M_2 , $\text{bias}(m_2) = \beta_{ED|M_2}(m_2) - \beta_{E \rightarrow D}^{TE}$ and C) adjustment for M_1 and M_2 , $\text{bias}(m_1,m_2) = \beta_{ED|M_1,M_2}(m_1,m_2) - \beta_{E \rightarrow D}^{TE}$.

(4) In Figure 1e, the causal diagrams contained a confounder of exposure-outcome relationship. On the odds ratio ($OR_{E \rightarrow D}^{TE}$) scale, the total effect ($\beta_{E \rightarrow D}^{TE} = \log(OR_{E \rightarrow D}^{TE})$), comparing exposure level e with e^* , we could obtain the total effect:

$$\begin{aligned}
\beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\
&= \log \left\{ \frac{P(D_e=1)/\{1-P(D_e=1)\}}{P(D_{e^*}=1)/\{1-P(D_{e^*}=1)\}} \right\} \\
&= \log \left\{ \frac{P(D_e=1) \times \{1-P(D_{e^*}=1)\}}{\{1-P(D_e=1)\} \times P(D_{e^*}=1)} \right\} \\
&= \log \left\{ \frac{P(D=1|e=1) \times P(D=0|e^*=0)}{P(D=0|e=1) \times P(D=1|e^*=0)} \right\} \\
&= \log \left\{ \frac{\left[\sum_{mu} P(D=1|e=1,m,u)P(m|e=1)P(u) \right] \times \left[\sum_{mu} P(D=0|e^*=0,m,u)P(m|e^*=0)P(u) \right]}{\left[\sum_{mu} P(D=0|e=1,m,u)P(m|e=1)P(u) \right] \times \left[\sum_{mu} P(D=1|e^*=0,m,u)P(m|e^*=0)P(u) \right]} \right\}
\end{aligned}$$

The effect ($\beta_{ED|M}(m)$) of adjusting for mediator M by logistic regression model can be given

$$\begin{aligned}
\beta_{ED|M}(m) &= \log it(P(D=1|e=1,m)) - \log it(P(D=1|e^*=0,m)) \\
&= \log \left\{ \frac{P(D=1|e=1,m) \times P(D=0|e^*=0,m)}{P(D=0|e=1,m) \times P(D=1|e^*=0,m)} \right\} \\
&= \log \left\{ \frac{[\sum_u P(D=1|e=1,m,u)p(u|e=1,m)] \times [\sum_u P(D=0|e^*=0,m,u)p(u|e^*=0,m)]}{[\sum_u P(D=0|e=1,m,u)p(u|e=1,m)] \times [\sum_u P(D=1|e^*=0,m,u)p(u|e^*=0,m)]} \right\} \\
&= \log \left\{ \frac{[\sum_u P(D=1|e=1,m,u) \frac{p(e=1|u)p(u)}{\sum_u p(e=1|u)p(u)}] \times [\sum_u P(D=0|e^*=0,m,u) \frac{p(e^*=0|u)p(u)}{\sum_u p(e^*=0|u)p(u)}]}{[\sum_u P(D=0|e=1,m,u) \frac{p(e=1|u)p(u)}{\sum_u p(e=1|u)p(u)}] \times [\sum_u P(D=1|e^*=0,m,u) \frac{p(e^*=0|u)p(u)}{\sum_u p(e^*=0|u)p(u)}]} \right\}
\end{aligned}$$

Therefore, we could evaluate the biases of adjustment models:

$$bias(m) = \beta_{ED|M}(m) - \beta_{E \rightarrow D}^{TE}$$

(5) Figure 1f is a depiction of two parallel mediators M_1 and M_2 with confounder.

On the odds ratio ($OR_{E \rightarrow D}^{TE}$) scale, the total effect ($\beta_{E \rightarrow D}^{TE} = \log(OR_{E \rightarrow D}^{TE})$), comparing exposure level e with e^* , we could obtain the total effect:

$$\begin{aligned}
\beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\
&= \log \left\{ \frac{P(D_e=1) / \{1 - P(D_e=1)\}}{P(D_{e^*}=1) / \{1 - P(D_{e^*}=1)\}} \right\} \\
&= \log \left\{ \frac{P(D_e=1) \times \{1 - P(D_{e^*}=1)\}}{\{1 - P(D_e=1)\} \times P(D_{e^*}=1)} \right\} \\
&= \log \left\{ \frac{P(D=1|e=1) \times P(D=0|e^*=0)}{P(D=0|e=1) \times P(D=1|e^*=0)} \right\} \\
&= \log \left\{ \frac{\xi_1}{\xi_2} \right\}
\end{aligned}$$

$$\begin{aligned}
\xi_1 &= [\sum_{m_1 m_2 u} P(D=1|e=1, m_1, m_2, u) P(m_2|e=1) P(m_1|e=1) P(u)] \\
&\quad \times [\sum_{m_1 m_2 u} P(D=0|e^*=0, m_1, m_2, u) P(m_2|e^*=0) P(m_1|e^*=0) P(u)] \\
\xi_2 &= [\sum_{m_1 m_2 u} P(D=0|e=1, m_1, m_2, u) P(m_2|e=1) P(m_1|e=1) P(u)] \\
&\quad \times [\sum_{m_1 m_2 u} P(D=1|e^*=0, m_1, m_2, u) P(m_2|e^*=0) P(m_1|e^*=0) P(u)]
\end{aligned}$$

The effect ($\beta_{ED|M_1}(m_1)$) of adjusting for mediator M_1 by logistic regression model can be given

$$\begin{aligned}\beta_{ED|M_1}(m_1) &= \text{logit}\{P(D=1|e=1,m_1)\} - \text{logit}\{P(D=1|e^*=0,m_1)\} \\ &= \log \left\{ \frac{P(D=1|e=1,m_1)P(D=0|e^*=0,m_1)}{P(D=0|e=1,m_1)P(D=1|e^*=0,m_1)} \right\} \\ &= \log \left\{ \frac{\xi_1}{\xi_2} \right\}\end{aligned}$$

$$\begin{aligned}\xi_1 &= \left[\sum_{m_2 u} P(D=1|e=1,m_1,m_2,u) \frac{P(m_2|e=1)P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right] \\ &\quad \times \left[\sum_{m_2 u} P(D=0|e^*=0,m_1,m_2,u) \frac{P(m_2|e^*=0)P(e^*=0|u)P(u)}{\sum_u P(e^*=0|u)P(u)} \right] \\ \xi_2 &= \left[\sum_{m_2 u} P(D=0|e=1,m_1,m_2,u) \frac{P(m_2|e=1)P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right] \\ &\quad \times \left[\sum_{m_2 u} P(D=1|e^*=0,m_1,m_2,u) \frac{P(m_2|e^*=0)P(e^*=0|u)P(u)}{\sum_u P(e^*=0|u)P(u)} \right]\end{aligned}$$

The effect ($\beta_{ED|M_2}(m_2)$) of adjusting for mediator M_2 by logistic regression model can be given

$$\begin{aligned}\beta_{ED|M_2}(m_2) &= \text{logit}\{P(D=1|e=1,m_2)\} - \text{logit}\{P(D=1|e^*=0,m_2)\} \\ &= \log \left\{ \frac{P(D=1|e=1,m_2)P(D=0|e^*=0,m_2)}{P(D=0|e=1,m_2)P(D=1|e^*=0,m_2)} \right\} \\ &= \log \left\{ \frac{\xi_1}{\xi_2} \right\}\end{aligned}$$

$$\begin{aligned}\xi_1 &= \left[\sum_{m_1 u} P(D=1|e=1,m_1,m_2,u) \frac{P(m_1|e=1)P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right] \\ &\quad \times \left[\sum_{m_1 u} P(D=0|e^*=0,m_1,m_2,u) \frac{P(m_1|e^*=0)P(e^*=0|u)P(u)}{\sum_u P(e^*=0|u)P(u)} \right] \\ \xi_2 &= \left[\sum_{m_1 u} P(D=0|e=1,m_1,m_2,u) \frac{P(m_1|e=1)P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right] \\ &\quad \times \left[\sum_{m_1 u} P(D=1|e^*=0,m_1,m_2,u) \frac{P(m_1|e^*=0)P(e^*=0|u)P(u)}{\sum_u P(e^*=0|u)P(u)} \right]\end{aligned}$$

The effect ($\beta_{ED|M_1,M_2}(m_1,m_2)$) of adjusting for mediator $M_1 M_2$ by logistic regression model can be given

$$\begin{aligned}
& \beta_{ED|M_1, M_2}(m_1, m_2) \\
&= \text{logit}\{P(D=1|e=1, m_1, m_2)\} - \text{logit}\{P(D=1|e^*=0, m_1, m_2)\} \\
&= \log \left\{ \frac{P(D=1|e=1, m_1, m_2)P(D=0|e^*=0, m_1, m_2)}{P(D=0|e=1, m_1, m_2)P(D=1|e^*=0, m_1, m_2)} \right\} \\
&= \log \left\{ \frac{\left[\sum_u P(D=1|e=1, m_1, m_2, u) \frac{P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right] \times \left[\sum_u P(D=0|e^*=0, m_1, m_2, u) \frac{P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right]}{\left[\sum_u P(D=0|e=1, m_1, m_2, u) \frac{P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right] \times \left[\sum_u P(D=1|e^*=0, m_1, m_2, u) \frac{P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right]} \right\}
\end{aligned}$$

Therefore, we could evaluate the biases that contains three adjustment models: A) adjustment for M_1 , $\text{bias}(m_1) = \beta_{ED|M_1}(m_1) - \beta_{E \rightarrow D}^{TE}$; B) adjustment for M_2 , $\text{bias}(m_2) = \beta_{ED|M_2}(m_2) - \beta_{E \rightarrow D}^{TE}$ and C) adjustment for M_1 and M_2 , $\text{bias}(m_1, m_2) = \beta_{ED|M_1, M_2}(m_1, m_2) - \beta_{E \rightarrow D}^{TE}$.

Supplementary B

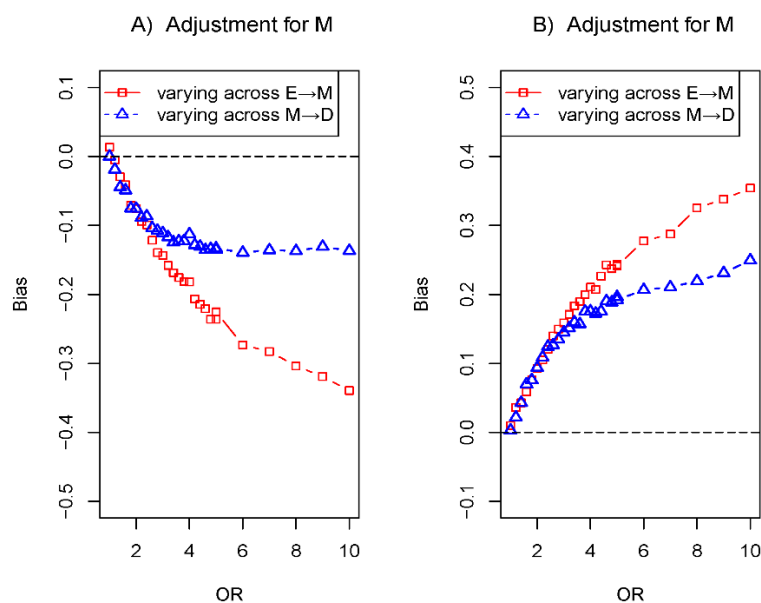


Figure S1: The biases with the effects $E \rightarrow M$ (red) and $M \rightarrow D$ (blue) increasing, respectively. Comparison of the bias of different effects in adjustment mediator.

The Figure S1-A obtained the result $\text{bias} < 0$ in Figure 3a with the effects $E \rightarrow M$, $M \rightarrow D$ and $E \rightarrow D$ fixing to $\ln 2$. The Figure S1-B gained the result $\text{bias} > 0$ in Figure 3c with the effects $E \rightarrow M$ and $E \rightarrow D$ fixing to $\ln 2$, effect $M \rightarrow D$ fixing to $-\ln 2$. We could obtain the bias performances of varying across the effects of exposure-mediator and mediator-outcome. The effect $E \rightarrow M$ of varying across was more sensitive than the effect $M \rightarrow D$ of varying across in Figure S1.

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*
Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	1
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any pre-specified hypotheses	3-4
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5-6
Participants	6	(a) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> —Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants	5-6
		(b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5-6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-6
Bias	9	Describe any efforts to address potential sources of bias	5-6
Study size	10	Explain how the study size was arrived at	5-6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Not applicable
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	4-6
		(b) Describe any methods used to examine subgroups and interactions	Not applicable
		(c) Explain how missing data were addressed	Not applicable
		(d) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed	Not applicable

		<i>Cross-sectional study</i> —If applicable, describe analytical methods taking account of sampling strategy	
		(e) Describe any sensitivity analyses	6
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Not applicable
		(b) Give reasons for non-participation at each stage	Not applicable
		(c) Consider use of a flow diagram	Not applicable
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7-13
		(b) Indicate number of participants with missing data for each variable of interest	Not applicable
		(c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	Not applicable
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time	Not applicable
		<i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure	Not applicable
		<i>Cross-sectional study</i> —Report numbers of outcome events or summary measures	7-13
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	7-13
		(b) Report category boundaries when continuous variables were categorized	Not applicable
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not applicable
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	7-13
Discussion			
Key results	18	Summarise key results with reference to study objectives	13-15
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13-15
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	13-15
Generalisability	21	Discuss the generalisability (external validity) of the study results	14
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	16

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Sensitivity analysis for mistakenly adjusting for mediators in estimating total effect in observational studies

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2016-015640.R3
Article Type:	Research
Date Submitted by the Author:	04-Jul-2017
Complete List of Authors:	Wang, Tingting; Shandong University, Department of Epidemiology and Biostatistics Li, Hongkai; Shandong University, Department of Epidemiology and Biostatistics Su, Ping; Shandong University, Department of Epidemiology and Biostatistics Yu, Yuanyuan; Shandong University, Department of Epidemiology and Biostatistics Sun, Xiaoru; Shandong University, Department of Epidemiology and Biostatistics Liu, Yi; Shandong University, Department of Epidemiology and Biostatistics Yuan, Zhongshang; Shandong University, Department of Epidemiology and Biostatistics Xue, Fuzhong; Shandong University, Department of Epidemiology and Biostatistics
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Public health
Keywords:	EPIDEMIOLOGY, PUBLIC HEALTH, STATISTICS & RESEARCH METHODS

SCHOLARONE™
Manuscripts

1 Sensitivity analysis for mistakenly adjusting for mediators in 2 estimating total effect in observational studies

3 Tingting Wang[†], Hongkai Li[†], Ping Su, Yuanyuan Yu, Xiaoru Sun, Yi Liu, Zhongshang Yuan,
4 Fuzhong Xue*

5 Department of Biostatistics, School of Public Health, Shandong University, Jinan City, Shandong
6 Province, P.R.China

7 Email addresses:

8 TTW: wangtingting161780@163.com

9 HKL: lihongkaiyouxiang@163.com

10 PS: sdsuping@163.com

11 YYY: yu_yy_1993@163.com

12 XRS: sunny150139@126.com

13 YL: liuyi238@sdu.edu.cn

14 ZSY: yuanzhongshang@sdu.edu.cn

15
16 *Corresponding Author

17 xuefzh@sdu.edu.cn (FZX)

18
19 **Word Count: 3726**

20 **Abstract**

21 **Objectives:** In observational studies, epidemiologists often attempt to estimate the
22 total effect of an exposure on an outcome of interest. However, when the underlying
23 diagram is unknown and only limited knowledge is available, dissecting bias
24 performances is essential to estimating the total effect of an exposure on an outcome
25 when mistakenly adjusting for mediators under logistic regression. Through
26 simulation, we focus on six causal diagrams concerning different roles of mediators.
27 Sensitivity analysis was conducted to assess the bias performances of varying across
28 exposure-mediator effects and mediator-outcome effects when adjusting for the
29 mediator.

30 **Setting:** Based on the causal relationships in the real world, we compare the biases of
31 varying across the effects of exposure-mediator with those of varying across the
32 effects of mediator-outcome when adjusting for the mediator. The magnitude of the

1 bias was defined by the difference between the estimated effect (using logistic
2 regression) and the total effect of the exposure on the outcome.

3 **Results:** In four scenarios (a single mediator, two series mediators, two independent
4 parallel mediators or two correlated parallel mediators), the biases of varying across
5 the effects of exposure-mediator were greater than those of varying across the effects
6 of mediator-outcome when adjusting for the mediator. In contrast, in two other
7 scenarios (a single mediator or two independent parallel mediators in the presence of
8 unobserved confounders), the biases of varying across the effects of
9 exposure-mediator were less than those of varying across the effects of
10 mediator-outcome when adjusting for the mediator.

11 **Conclusions:** The biases were more sensitive to the variation of effects of
12 exposure-mediator than the effects of mediator-outcome when adjusting for the
13 mediator in the absence of unobserved confounders, while the biases were more
14 sensitive to the variation of effects of mediator-outcome than those of
15 exposure-mediator in the presence of an unobserved confounder.

16 **Strengths and limitations of this study**

17 1) For six different causal diagrams, we compared biases of distinct adjustment
18 strategies with and without adjusting for mediators by conducting simulation studies.

19 2) Sensitivity analysis was conducted to assess the performances of varying across the
20 effects of exposure-mediator and mediator-outcome.

21 3) The simulation schemes and parameters were conducted mainly based on real
22 observational studies.

23 4) The combination of theoretical derivation and simulation studies make the results
24 more credible.

1
2
3 5) The limitation of these simulation studies was that they operated under the
4
5 framework of logistic regression and therefore focused on only binary variables.
6

7 **Introduction**

8
9
10 Estimating the total effect of the exposure (E) on the outcome (D) is a great challenge
11
12 in epidemiology studies because confounders are commonly confused with
13
14 mediators.¹⁻³ If confounders and mediators are misclassified, the ability to control
15
16 confounders in the estimation of the total effect of the exposure on the outcome is
17
18 hampered. In fact, various strategies are used to eliminate confounding bias in
19
20 observational studies. The conventional approaches include multivariate regression,
21
22 stratification, standardization and inverse-probability weighting.⁴⁻⁵ Furthermore,
23
24 causal diagrams provide a formal conceptual framework for identifying and selecting
25
26 confounders,⁶⁻⁷ so that analysis can avoid falling into analytic pitfalls.⁸ In practice,
27
28 even the underlying causal diagrams and the role of covariates (mediator, confounder,
29
30 collider and instrumental variable) are not completely understood, as investigators
31
32 usually adjust for the covariates that are associated with the outcome and exposure.⁹⁻¹²
33
34
35 Therefore, our paper focuses on the biases of varying across the effects of
36
37 exposure-mediator ($E \rightarrow M$) and mediator-outcome ($M \rightarrow D$) when mistakenly adjusting
38
39 for mediators under the logistic regression model.
40
41
42

43
44 Several causal inference studies have made considerable contributions to mediation
45
46 analysis by providing definitions for direct and indirect effects that allow for the
47
48 decomposition of a total effect into a direct and an indirect effect.¹³⁻²¹ Arbitrarily
49
50 adjusting for a mediator would generally bias the estimate of the total effect of the
51
52 exposure on the outcome.^{8,22-23} Practically, it can mistakenly identify a
53
54 non-confounding risk factor as a confounder. In the perspective of causal diagrams,
55
56
57
58
59
60

1 little attention has been paid to the biases when adjusting for mediators under the
2 logistic regression model in estimating the total effect of E on D . Hence, we focused
3 on the sensitivity analysis technique to assess the biases of varying across the effects
4 of $E \rightarrow M$ and $M \rightarrow D$ when adjusting for the mediator.

5 In this paper, six typical causal diagrams corresponding to causal correlation are
6 given in Figure 1: a single mediator (Figure 1a); two series mediators (Figure 1b); two
7 independent parallel mediators (Figure 1c); two correlated parallel mediators (Figure
8 1d); a single mediator with an unobserved confounder (Figure 1e); and two parallel
9 mediators with an unobserved confounder (Figure 1f). The paper aims to explore the
10 sensitivity of biases to the variation of the effects of $E \rightarrow D$ and $M \rightarrow D$ when adjusting
11 for the mediator. Hence, both theoretical proofs and quantitative simulations were
12 performed to dissect the bias of varying across the effect of $E \rightarrow M$ and varying across
13 the effect of $M \rightarrow D$ when adjusting for mediators under the logistic model.

14 **Methods**

15 A directed acyclic graph (*DAG*) is composed of variables (nodes) and arrows (directed
16 edges) between nodes such that the graph is acyclic. The causal diagrams are
17 formalized as directed acyclic graphs (*DAGs*), providing investigators with powerful
18 tools for bias assessment.²⁴ It provides a device for deducing the statistical
19 associations implied by causal relations. Furthermore, given a set of observed
20 statistical associations, a researcher knowledgeable about causal diagrams theory can
21 systematically characterize all causal structures compatible with the observations.²⁵⁻²⁶

22 The total effect of the exposure on the outcome can be calculated based on the
23 *do-calculus* and *back-door* criterion proposed by Judea Pearl.²⁷⁻²⁸ For exposure X and

1 outcome Y , a set of variables Z satisfies the backdoor path criterion with respect to (X ,
 2 Y) if no variable in Z is a descendant of X and Z blocks all back-door paths from X to
 3 Y . Then, the effect of X on Y is given by the following formula:

$$P(y|do(x)) = \sum_z P(y|x,z)P(z)$$

4 Note that the expression on the right hand side of the equation is simply a
 5 standardized mean. The difference $E(Y|do(x')) - E(Y|do(x''))$ is taken as the
 6 definition of “causal effect”, where x' and x'' are two distinct realizations of X .²³
 7 The interventional distribution, such as that corresponding to $Y(x)$, namely
 8 $P(y|do(x))$, is not necessarily equal to a conditional distribution $P(y|x)$. It stands for
 9 the probability of $Y = y$ when the exposure X is set to level x . The ignorability
 10 assumption $Y(x) \perp X$ states that if we happen to have information on the exposure
 11 variable, it does not give us any information about the outcome Y after the
 12 intervention $do(x)$ was performed. In addition, it can be shown that if ignorability
 13 holds for $Y(x)$ and X (alternatively if there are no back-door paths from X to Y in the
 14 corresponding causal DAGs), then $P(y|do(x)) = P(y|x)$.²⁹⁻³⁰

15 Let D_e and M_e denote the values of the outcome and mediator that would have
 16 been observed had the exposure E been set to level e , respectively. On the odds ratio
 17 ($OR_{E \rightarrow D}^{TE}$) scale, the total effect ($\beta_{E \rightarrow D}^{TE} = \log(OR_{E \rightarrow D}^{TE})$), comparing exposure level e
 18 with e^* , is given as the following: $OR_{E \rightarrow D}^{TE} = \frac{P(D_e = 1) / \{1 - P(D_e = 1)\}}{P(D_{e^*} = 1) / \{1 - P(D_{e^*} = 1)\}}$.²⁰⁻²¹ While the
 19 effect ($\beta_{ED|M}(m)$) of adjusting for mediator M by the logistic regression model can be
 20 given as the following:
 21

$$\begin{aligned}\beta_{ED|M}(m) &= \text{logit}\{P(D=1|e=1,m)\} - \text{logit}\{P(D=1|e^*=0,m)\} \\ &= \log\left\{\frac{P(D=1|e=1,m)P(D=0|e^*=0,m)}{P(D=0|e=1,m)P(D=1|e^*=0,m)}\right\}\end{aligned}$$

where $P(D=1|e,m)$ denotes the probability of $D=1$ when the exposure E and mediator M have been set to level e and m , respectively. Taking Figure 1a as an example, the logistic regression is as follows:

$$\text{logit}\{P(D=1|e,m)\} = \alpha_1 + \beta_0 e + \beta_2 m.$$

Therefore, the total effect ($\beta_{E \rightarrow D}^{TE}$) of exposure E on outcome D on the scale of logarithm odds ratio was equal to

$$\begin{aligned}\beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\ &= \log\left\{\frac{P(D_e=1)/\{1-P(D_e=1)\}}{P(D_{e^*}=1)/\{1-P(D_{e^*}=1)\}}\right\} \\ &= \text{logit}\{P(D_e=1)\} - \text{logit}\{P(D_{e^*}=1)\} \\ &= \text{logit}\{P(D=1|e=1)\} - \text{logit}\{P(D=1|e^*=0)\} \\ &= \text{logit}\left\{\sum_m P(D=1|e=1,m)P(m|e=1)\right\} - \text{logit}\left\{\sum_m P(D=1|e^*=0,m)P(m|e^*=0)\right\}\end{aligned}$$

The effect estimation ($\hat{\beta}_{ED|M}(m)$) of adjusting for mediator M by the logistic regression model was equal to

$$\hat{\beta}_{ED|M}(m) = \text{logit}\{\hat{P}(D=1|e=1,m)\} - \text{logit}\{\hat{P}(D=1|e^*=0,m)\}$$

where $\hat{P}(D=1|e=1,m)$ denotes the probability of $D=1$ when the exposure E and mediator M have been set to level $e=1$ and m , respectively. Additionally, $\hat{P}(D=1|e^*=0,m)$ denotes the probability of $D=1$ when the exposure E and mediator M have been set to level $e^*=0$ and m , respectively. The theoretical results of other causal diagrams in Figure 1 have been shown in the supplementary A.

Note that the bias was defined by taking a difference between effect estimation by adjusting for the mediator using logistic regression and the total effect of exposure E on outcome D i.e., $\text{bias} = E[\hat{\beta}_{ED|M}(m)] - \beta_{E \rightarrow D}^{TE}$. We dissected the behavior of the biases by varying across the effects of $E \rightarrow M$ and $M \rightarrow D$ when mistakenly adjusting

1 for the mediator under the framework of the logistic regression model.

2 **Simulation**

3 Six scenarios are designed to dissect the sensitivity of bias to the variation of the
4 effects of exposure-mediator and mediator-outcome when adjusting for mediators
5 under the framework of the logistic regression model; these DAGs are shown in
6 Figure 1. We made the following assumptions for the simulation: 1) all variables were
7 binary, following a Bernoulli distribution; and 2) the effects from parent nodes to their
8 child node were positive and log-linearly additive. Taking Figure 1a as an example,
9 we randomly generated the exposure following a Bernoulli distribution (i.e. let
10 $P(e = 1) = \pi$). Then, we used $P_M = \exp(\alpha_0 + \beta_1 e) / \{1 + \exp(\alpha_0 + \beta_1 e)\}$ to calculate the
11 distribution probability of child node M from its parent node E . Similarly,
12 $P_D = \exp(\alpha_1 + \beta_0 e + \beta_2 m) / \{1 + \exp(\alpha_1 + \beta_0 e + \beta_2 m)\}$ generated the distribution
13 probability of D , where the parameters α_0 and α_1 denoted the intercept of M and
14 D respectively, and effect parameters $\beta_0, \beta_1, \beta_2$ referred to the effects of the parent
15 node on their corresponding child node using a log odds ratio scale.

16 After generating data, we dissected the behavior of the biases between the effects of
17 $E \rightarrow M$ and $M \rightarrow D$ when mistakenly adjusting for mediators under the logistic
18 regression model. In scenario 1 (Figure 1a), we compared performances by varying
19 across the effects of $E \rightarrow M$ and $M \rightarrow D$. Similarly, in scenario 2 (Figure 1b), the effects
20 of $E \rightarrow M_1$, $M_1 \rightarrow M_2$ and $M_2 \rightarrow D$ were explored. In scenario 3 (Figure 1c), we dissected
21 the effects of $E \rightarrow M_1$ ($E \rightarrow M_2$) and $M_1 \rightarrow D$ ($M_2 \rightarrow D$). The comparison of scenario 4
22 (Figure 1d) was the same as scenario 3 (Figure 1c). In scenario 5 (Figure 1e), the

1 effects of $E \rightarrow M$ and $M \rightarrow D$ were excavated. Scenario 6 (Figure 1f) was identical to
2 the scenario 3. We explored the biases when adjusting for mediators under the logistic
3 regression model and thus identified the sensitivity of biases to the variation of the
4 effects of exposure-mediator and mediator-outcome.

5 For each of the 6 simulation scenarios, we observed the biases of varying across
6 distinct effects when adjusting for mediators using the logistic regression model with
7 1000 simulation repetitions. All simulations were conducted using software R from
8 CRAN (<http://cran.r-project.org/>).

9 Results

10 *Scenario 1: one single mediator (Figure 1a)*

11 In Figure 1a, E has a direct ($E \rightarrow D$) effect and an indirect ($E \rightarrow M \rightarrow D$) effect on D . In
12 Figure 2, Figure 2A depicted that the bias of varying across the effect of $E \rightarrow M$ was
13 clearly greater than the bias of varying across the effect of $M \rightarrow D$. That is, the
14 sensitivity of bias to the variation of the effect $E \rightarrow M$ was greater than that of the
15 effect of $M \rightarrow D$ when adjusting for the mediator M using the logistic regression model.
16 In particular, if the effect of $E \rightarrow M$ was specified to zero in Figure 2B, M would be
17 associated with D conditional on E and unconditionally independent with E , and M
18 would become an independent risk factor of the outcome, as adjusting for M would
19 obtain a positive “bias”. Such bias was a consequence of the non-collapsibility of the
20 odds ratio, and the M-conditional ORs must be farther from 1 than the unconditional
21 ORs.³¹⁻³² In fact, both adjustment and non-adjustment for M should yield unbiased
22 causal effect estimates. Certainly, in this case, both the marginal OR and conditional
23 OR obtained from standardization and inverse-probability weighting were equal to the
24 total effect.³³ Moreover, Figure 2A indicated that adjusting for mediator M was indeed
25 biased to the total effect of the exposure on the outcome.

1 The total effect ($\beta_{E \rightarrow D}^{TE}$) of exposure E on outcome D on the log odds ratio scale
 2 was equal to

$$\begin{aligned} \beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) = \log \left\{ \frac{P(D_e = 1) / \{1 - P(D_e = 1)\}}{P(D_{e^*} = 1) / \{1 - P(D_{e^*} = 1)\}} \right\} \\ &= \log \left\{ \frac{P(D_e = 1) \times \{1 - P(D_{e^*} = 1)\}}{\{1 - P(D_e = 1)\} \times P(D_{e^*} = 1)} \right\} \\ &= \log \left\{ \frac{P(D = 1 | e = 1) \times P(D = 0 | e^* = 0)}{P(D = 0 | e = 1) \times P(D = 1 | e^* = 0)} \right\} \\ &= \log \left\{ \frac{[\sum_m P(D = 1 | e = 1, m) P(m | e = 1)] \times [\sum_m P(D = 0 | e^* = 0, m) P(m | e^* = 0)]}{[\sum_m P(D = 0 | e = 1, m) P(m | e = 1)] \times [\sum_m P(D = 1 | e^* = 0, m) P(m | e^* = 0)]} \right\} \end{aligned}$$

4 The effect ($\beta_{ED|M}(m)$) of adjusting for mediator M by the logistic regression model
 5 can be given as follows:

$$\begin{aligned} \beta_{ED|M}(m) &= \text{logit} \{P(D = 1 | e = 1, m)\} - \text{logit} \{P(D = 1 | e^* = 0, m)\} \\ &= \log \left\{ \frac{P(D = 1 | e = 1, m) \times \{1 - P(D = 1 | e^* = 0, m)\}}{\{1 - P(D = 1 | e = 1, m)\} \times P(D = 1 | e^* = 0, m)} \right\} \\ &= \beta_0 \end{aligned}$$

7 β_0 denotes coefficient of E adjusting for M using the logistic regression model.

8 Furthermore, the effect of adjusting for M was equal to the controlled direct effect.¹⁹

9 Therefore, the bias of adjusting for the mediator using the logistic regression model
 10 could be obtained i.e., $\text{bias} = \beta_{ED|M}(m) - \beta_{E \rightarrow D}^{TE}$. We added signs to the edges of the
 11 directed acyclic graph to indicate the presence of a particular positive or negative
 12 effect in Figure 3. Therefore, we gained $\text{bias} < 0$ under the condition of $\beta_1 * \beta_2 > 0$
 13 (the effect $E \rightarrow M$ β_1 and the effect $M \rightarrow D$ β_2), indicating that the total effect of E on D
 14 was biased when adjusting for M using the logistic regression model in Figure 3a,
 15 Figure 3b, Figure 3e & Figure 3f. In addition, the bias was less than zero when the
 16 effect $E \rightarrow M$ (β_1) and the effect $M \rightarrow D$ (β_2) shared same signs. (i.e., both the effects

1 $E \rightarrow M$ ($\beta_1 > 0$) and $M \rightarrow D$ ($\beta_2 > 0$) were a positive sign or both the effects $E \rightarrow M$
2 ($\beta_1 < 0$) and $M \rightarrow D$ ($\beta_2 < 0$) were a negative sign). Furthermore, we obtained $bias > 0$,
3 if $\beta_1 * \beta_2 < 0$, suggesting that the total effect of E on D was biased when adjusting for
4 M in Figure 3c, Figure 3d, Figure 3g & Figure 3h. In addition, the bias was greater
5 than zero when the signs of the effects $E \rightarrow M$ (β_1) and $M \rightarrow D$ (β_2) were the opposite.
6 The results illustrated that the bias was less than zero in the case in which the effects
7 of exposure-mediator and mediator-outcome shared the same sign; the bias was
8 greater than zero under the circumstance in which the effects of exposure-mediator
9 and mediator-outcome had opposite signs. We also illustrated the case of Figure 3c
10 with the effects $E \rightarrow M$ and $E \rightarrow D$ as greater than zero, and the effect $M \rightarrow D$ as less
11 than zero in supplementary B. More details of theoretical derivation can be found in
12 Appendix.

13 **Scenario 2: two series mediators (Figure 1b)**

14 Figure 1b is a depiction through two series mediators, decomposing total effects into
15 direct effect ($E \rightarrow D$) and indirect effect ($E \rightarrow M_1 \rightarrow M_2 \rightarrow D$). The bias of varying across
16 the effect of $E \rightarrow M_1$ was greater than that of varying across the effect of $M_2 \rightarrow D$ under
17 adjustment for M_1 , M_2 and $M_1 M_2$ together in Figure 4, respectively. In this situation,
18 the correlation of series mediators was strong enough to prevent M_2 from becoming
19 an independent cause of the outcome.

20 **Scenario 3: two independent parallel mediators (Figure 1c)**

21 Figure 1c shows that the exposure E independently causes M_1 and M_2 and indirectly
22 influences the outcome D through M_1 and M_2 , forming three causal paths $E \rightarrow D$,

1
2
3
4 1 $E \rightarrow M_1 \rightarrow D$ and $E \rightarrow M_2 \rightarrow D$. For Figure 5, the results indicated that the bias of varying
5
6 2 across the effect of $E \rightarrow M_1$ was considerably greater than that of varying across the
7
8 3 effect of $M_1 \rightarrow D$ under adjustment for M_1 in Figure 5A. However, the bias of varying
9
10 4 across the effect of $E \rightarrow M_2$ was nearly equal to that of varying across the effect of
11
12 5 $M_2 \rightarrow D$ under the identical model of adjustment for M_1 in Figure 5A. Then, a result
13
14 6 similar to the one above can be obtained in Figure 5B. In addition, Figure 5C
15
16 7 indicated that biases of varying across the effects of $E \rightarrow M_1$ and $E \rightarrow M_2$ were
17
18 8 obviously greater than those of varying across the effects of $M_1 \rightarrow D$ and $M_2 \rightarrow D$ while
19
20 9 simultaneously adjusting for M_1 and M_2 .

21
22
23
24
25
26
27 10 ***Scenario 4: two correlated parallel mediators (Figure 1d)***

28 11 In Figure 1d, there exist five paths from E to D : $E \rightarrow D$, $E \rightarrow M_1 \rightarrow D$, $E \rightarrow M_2 \rightarrow D$,
29
30 12 $E \rightarrow M_1 \leftarrow M_2 \rightarrow D$ and $E \rightarrow M_2 \rightarrow M_1 \rightarrow D$. In particular, the path $E \rightarrow M_1 \leftarrow M_2 \rightarrow D$ is a
31
32 13 blocked path, due to M_1 being a collider node. In Figure 6, Figure 6A indicated that
33
34 14 the bias of varying across the effect of $E \rightarrow M_1$ was clearly greater than that of varying
35
36 15 across the effect of $M_1 \rightarrow D$ under adjustment for M_1 . However, the bias of varying
37
38 16 across the effect of $E \rightarrow M_2$ was almost equal to that of varying across the effect of
39
40 17 $M_2 \rightarrow D$ under the identical adjustment model. Similarly, an analogous result of the
41
42 18 behavior of the biases is shown in Figure 6B. In addition, the biases of varying across
43
44 19 the effects of $E \rightarrow M_1$ and $E \rightarrow M_2$ were greater than those of varying across the effects
45
46 20 of $M_1 \rightarrow D$ and $M_2 \rightarrow D$ when adjusting for M_1 and M_2 in Figure 6C. Simultaneously,
47
48 21 the bias was more sensitive to the variation of the effect of $E \rightarrow M_2$ than the effect of
49
50 22 $E \rightarrow M_1$ under adjustment for M_1 and M_2 , while adjusting for the collider node M_1
51
52 23 would partially open the path $E \rightarrow M_1 \leftarrow M_2 \rightarrow D$.

53
54
55
56
57
58 24 ***Scenario 5: a single mediator with an unobserved confounder (Figure 1e)***

1
2
3
4 1 Figure 1e provides a causal diagram representing the relationship among exposure E ,
5
6 2 outcome D , mediator M and unobserved confounder U . It revealed that the bias of
7
8 3 varying across the effect of $E \rightarrow M$ was lower than that of varying across the effect of
9
10 4 $M \rightarrow D$. An unobserved confounder distorts the association between the exposure and
11
12 5 outcome ($E \leftarrow U \rightarrow D$) in Figure 7.

13
14
15
16
17 6 **Scenario 6: two parallel mediators with an unobserved confounder (Figure 1f)**

18
19 7 As described above, Figure 1f is a depiction of two parallel mediators M_1 and M_2 with
20
21 8 an unobserved confounder U . For Figure 8, the bias of varying across the effect of
22
23 9 $E \rightarrow M_1$ was clearly less than that of varying across the effect of $M_1 \rightarrow D$ under the
24
25 10 adjustment for M_1 in Figure 8A. However, the bias of varying across the effect of
26
27 11 $E \rightarrow M_2$ was greater than that of varying across the effect of $M_2 \rightarrow D$ under the identical
28
29 12 model adjusting for M_1 . A similar result can also be obtained in Figure 8B. In addition,
30
31 13 biases of varying across the effects of $E \rightarrow M_1$ and $E \rightarrow M_2$ were distinctly less than
32
33 14 those of varying across the effects of $M_1 \rightarrow D$ and $M_2 \rightarrow D$ under the common model of
34
35 15 adjusting for M_1 and M_2 in Figure 8C.

36
37
38
39
40
41 16 **Application**

42
43
44 17 In this analysis, we evaluated two statistical models (unadjusted and M-adjusted) to
45
46 18 assess the effect of diabetes on cardiovascular diseases under scenario 1. Information
47
48 19 from 22,900 individuals were collected from the Health Management Center of
49
50 20 Shandong Provincial Hospital (HMCSPH). All individuals were Urban Han Chinese
51
52 21 and more than 20 years of age and they underwent a physical examination in 2013.
53
54
55 22 Many studies focused on the associations between diabetes and metabolic
56
57
58
59
60

1
2
3 syndrome,³⁴ and between metabolic syndrome and cardiovascular disease.³⁵

4
5
6 The exposure indicator E takes a value of 1 if individuals suffer from diabetes and
7
8 takes a value of zero otherwise. The outcome D (cardiovascular diseases) takes a
9
10 value of 1 if individuals are diagnosed with cardiovascular diseases and takes a value
11
12 of 0 otherwise. The mediator M (metabolic syndrome) takes a value the value of 1 if
13
14 individuals diagnosed with metabolic syndrome and takes a value of 0 otherwise.
15
16 After adjusting for age and gender, using the logistic regression model obtained the
17
18 total effect of diabetes E on cardiovascular diseases D equal to $\beta = 0.598$ (95%
19
20 confidence interval (CI), 0.307~0.877). Then, the effect of adjusting for metabolic
21
22 syndrome M was equal to $\beta_M = 0.429$ (95% confidence interval (CI), 0.113~0.736).
23
24 Therefore, the bias was $\beta_M - \beta = -0.169 < 0$, suggesting that the effect of E on D was
25
26 underestimated when adjusting for the mediator M . This bias can have negative
27
28 implications on the interpretation of the effects of diabetes on cardiovascular diseases.
29
30 The adjustment for the mediator produced biased estimates, and adjustment was thus
31
32 inappropriate and should have been avoided. A specific example was the adjustment
33
34 for time-varying confounders that are also mediators using methods including
35
36 standardization, inverse-probability weighting, and G-estimation.³⁶ That is,
37
38 investigators should remember to consider biological and clinical information when
39
40 specifying a statistical model.
41
42
43
44
45
46
47
48
49
50

51 Discussion

52
53 In the paper, we dissected the sensitivity of bias to the variation of the effects of
54
55 exposure-mediator and mediator-outcome when adjusting for mediators under the
56
57 framework of the logistic regression model. In four scenarios (a single mediator in
58
59
60

1
2
3 1 Figure 1a of scenario 1, two series mediators in Figure 1b of scenario 2, two
4 independent parallel mediators in Figure 1c of scenario 3 or two correlated parallel
5 mediators in Figure 1d of scenario 4), the bias of varying across the effect of
6 exposure-mediator was greater than that of varying across the effect of
7 mediator-outcome when adjusting for the mediator (Figure 2, Figure 4, Figure 5 &
8 Figure 6). However, in two other scenarios (a single mediator or two independent
9 parallel mediators in the presence of unobserved confounders in Figure 1e of scenario
10 5 & Figure 1f of scenario 6), the biases were more sensitive to the variation of the
11 effect of mediator-outcome than the effect of exposure-mediator when adjusting for
12 the mediator (Figure 7 & Figure 8).

13
14 Conditioning on a mediator is of concern in all areas of epidemiologic
15 studies,^{13,19,37} it indeed lead to bias in estimating the total effect of the exposure on the
16 outcome.^{8,22-23} Mediators and confounders are indistinguishable in terms of statistical
17 association and conceptual grounds.³ Most of the studies focus on the mediation effect
18 analysis such as the calculation of direct effect and indirect effect.^{20-21,38-41} Recently,
19 some authors have used causal diagrams to describe how to appropriately handle
20 matching variables. In addition, they have proven that matching on mediator M
21 renders M and D independent (by design) in the matched study. Matching on variables
22 that are affected by the exposure and the outcome, i.e., mediators between the
23 exposure and the outcome, would ordinary produce irremediable bias. Furthermore,
24 matching on mediator M blocks the causal path $E \rightarrow M \rightarrow D$ and thus produces
25 unfaithfulness in estimating the total effect E on D .^{31,42} Little effort has been made to
learn the performances of biases when adjusting for a mediator in estimating the total
effect of an exposure on an outcome. Our study results revealed that the biases were
more sensitive to the variation of the effects of exposure-mediator than effects of

1
2
3 1 mediator-outcome when adjusting for the mediator in the absence of the unobserved
4
5 2 confounder in causal diagrams (Figure 1a, Figure 1b, Figure 1c & Figure 1d).
6
7 3 Nevertheless, for causal diagrams (Figure 1e & Figure 1f), the biases were more
8
9 4 sensitive to the variation of effects of mediator-outcome than the effects of
10
11 5 exposure-mediator when adjusting for a mediator in the presence of the unobserved
12
13 6 confounder. Therefore, the biases of varying across different effects depended on the
14
15 7 causal diagrams framework and whether an unobserved confounder existed.

18
19 8 The causal diagrams depicted in Figure 1 are indeed very simplistic and concise, as
20
21 9 they exclude the confounding factors of E and M as well as M and D . In practical
22
23 10 applications, there exist some confounders in each pair of relationships among E , M ,
24
25 11 and D . In addition, our simulation study was not comprehensive enough to evaluate
26
27 12 the bias performances when adjusting for the mediator under logistic regression
28
29 13 because it considered only binary variables, certain scenarios of effect size and
30
31 14 common types of models. In medical research, regression modeling is commonly used
32
33 15 to adjust for covariates associated with both the outcome and exposure. In this paper,
34
35 16 the biases are defined by the difference between M-adjusted and unadjusted ORs,
36
37 17 some of which is attributable to the non-collapsibility of the OR. In the field of causal
38
39 18 inference, standardization and inverse-probability weighting may obtain a different
40
41 19 bias from that of regression modeling, and they may be better alternatives to calculate
42
43 20 bias⁴⁻⁵. Therefore, in future research, the methods of standardization and
44
45 21 inverse-probability weighting could be used to calculate the biases of this paper
46
47 22 definition. Future research should further reinforce the mechanisms and conceptual
48
49 23 frameworks of confounders and mediators from causal diagrams to avoid falling into

1 analytic pitfalls.

2 **Conclusion**

3 In conclusion, the sensitivity of biases to the variation of the effects of
4 exposure-mediator and mediator-outcome were related to whether there was an
5 unobserved confounder in causal diagrams. The biases were more sensitive to the
6 variation of the effects of exposure-mediator than the effects of mediator-outcome
7 when adjusting for the mediator in the absence of unobserved confounders, while the
8 biases were more sensitive to the variation of the effects of mediator-outcome than the
9 effects of exposure-mediator in the presence of unobserved confounders.

10 **Statements**

11 **Acknowledgments**

12 We would like to thank the anonymous reviewers and academic editors for providing
13 us with constructive comments and suggestions and also wish to acknowledge our
14 colleagues for their invaluable work.

15 **Authors' contributions**

16 TTW and HKL jointly conceived the idea behind the article and designed the study.
17 TTW helped conduct the literature review, performed the simulation and prepared the
18 draft of the manuscript. PS, YYY, XRS, YL and ZSY participated in the design of the
19 study and the revision of the manuscript. FZX advised on critical revision of the
20 manuscript for important intellectual content. All authors read and approved the final
21 manuscript.

22 **Funding**

23 This work was supported by grants from the National Natural Science Foundation of
24 China (grant number 81573259)

25 **Competing interests**

26 The authors declare that they have no competing interests.

1 Ethics approval and materials

2 Ethics Committee of the School of Public Health (20140322), Shandong University.

3 Written informed consent was obtained from all participants.

4 Provenance and peer review

5 Not commissioned; externally peer reviewed.

6 Data sharing statement

7 No additional data are available.

8 Open Access

9 This is an Open Access article distributed in accordance with the Creative Commons
10 Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to
11 distribute, remix, adapt, build upon this work noncommercially, and license their
12 derivative works on different terms, provided the original work is properly cited and
13 the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

14 References

- 15 1 Bulterys M, Morgenstern H. Confounding or intermediate effect? An appraisal of
16 iatrogenic bias in perinatal AIDS research. *Paediatr Perinat Epidemiol*
17 1993;7:387–394.
- 18 2 Herná'n MA, Hernandez-Diaz S, Werler MM, et al. Causal knowledge as a
19 prerequisite for confounding evaluation: an application to birth defects
20 epidemiology. *Am J Epidemiol* 2002;155:176–184.
- 21 3 MacKinnon DP, Krull JL, and Lockwood CM. Equivalence of the mediation,
22 confounding and suppression effect. *Prevention Science* 2000;1:173–181.
- 23 4 Mansournia MA, Altman DG. Inverse probability weighting[J]. *Bmj*, 2016,
24 352:i189.
- 25 5 Gharibzadeh S, Mohammad K, Rahimiforoushani A, Amouzegar A, Mansournia
26 MA. Standardization as a tool for causal inference in medical research. *Arch Iran*
27 *Med.* 2016;19:666-70.
- 28 6 Howards PP, Schisterman EF, Poole C, et al. “Toward a clearer definition of
29 confounding” revisited with directed acyclic graphs, *Am J Epidemiol*
30 2012;176:506–511.
- 31 7 Williamson EJ, Aitken Z, Lawrie J, et al. Introduction to causal diagrams for
32 confounder selection. *Respirology*. 2014;19:303-311.
- 33 8 Wilcox AJ, Weinberg CR, Basso O. On the pitfalls of adjusting for gestational
34 age at birth. *Am J Epidemiol.* 2011;174:1062–1068.

- 1
2
3 1 9 Psaty BM, Koepsell TD, Lin D, et al. Assessment and control for confounding by
4 2 indication in observational studies. *J Am Geriatr Soc* 1999; 47:749-54.
5 3 10 Tilaki K H. Methodological issues of confounding in analytical epidemiology
6 4 studies. *Caspian J intern Med* 2012;3:488-495.
7 5 11 Kamangar F MD. Confounding variable in epidemiology studies: Basics and
8 6 beyond. *Arch Tran Med.* 2012;15:508-516.
9 7 12 Gaskins AJ, Schisterman EF. The effect of lipid adjustment on the analysis of
10 8 environmental contaminants and the outcome of human health risks. *Methods*
11 9 *Mol Biol* 2009;580:371-381.
12 10 13 Robins JM, Greenland S. Identifiability and exchangeability for direct and
13 11 indirect effects. *Epidemiology* 1992;3:143-155.
14 12 14 Pearl J. Direct and indirect effects. Seventeenth conference of uncertainty in
15 13 artificial intelligence, 2001. San Francisco, CA: Morgan Kaufmann, 2001.
16 14 15 Huang B, Sivaganesan S, Succop P, et al. Statistical assessment of mediational
17 15 effects for logistic mediational models. *Statist. Med* 2004;23:2713-2728.
18 16 16 Valeri L, Coull B A. Estimating causal contrasts involving intermediate variables
19 17 in the presence of selection bias [J]. *Statistics in Medicine*, 2016, 35(26):
20 18 4779-4793.
21 19 17 MacKinnon D P, Pirlott A G. Statistical approaches for enhancing causal
22 20 interpretation of the M to Y relation in mediation analysis [J]. *Personality and*
23 21 *Social Psychology Review*, 2015, 19(1): 30-43.
24 22 18 Imai K, Keele L, Tingley D. A general approach to causal mediation analysis.
25 23 *Psychol Methods.* 2010;15(4):309-334.
26 24 19 VanderWeele TJ. Bias formulas for sensitivity analysis for direct and indirect
27 25 effects. *Epidemiology.* 2010;21(4):540-551.
28 26 20 VanderWeele TJ, Vansteelandt S. Odds ratios for mediation analysis for a
29 27 dichotomous outcome. *Am J Epidemiol.* 2010;172(12):1339-1348.
30 28 21 Valeri L, VanderWeele TJ. Mediation analysis allowing for exposure-mediator
31 29 interactions and causal interpretation: theoretical assumptions and
32 30 implementation with SAS and SPSS macros [J]. *Psychological methods*, 2013,
33 31 18(2): 137.
34 32 22 Robins J. The control of confounding by intermediate variable. *Statistics in*
35 33 *Medicine.* 1989;8:679-701.
36 34 23 Schistermana EF, Coleb SR, Platt RW. Overadjustment Bias and Unnecessary
37 35 Adjustment in Epidemiologic Studies. *Epidemiology.* 2009;20:488-495.
38 36 24 Pearl J. Causal diagrams for empirical research. *Biometrika.* 1995;82:669-688.
39 37 25 Pearl J. Causality: models, reasoning, and inference. *Econometric Theory.*
40 38 2003;19: 675-685.
41 39 26 Joffe M, Gambhir M, Chadeau-Hyam M, et al. Causal diagrams in systems
42 40 epidemiology. *Emerg Themes Epidemiol.* 2012; 9:1-18.
43 41 27 Pearl, J. Causality: models, reasoning, and inference. Cambridge: Cambridge
44 42 University Press; 2000.
45 43 28 Pearl J. Causal inference in statistics: an overview. *Statistics Surveys.* 2009;
46 44 3:96-146.

- 1
2
3 1 29 Rubin DB. Should observational studies be designed to allow lack of balance in
4 2 covariate distributions across treatment groups? *Statistics in Medicine*.
5 3 2009;28:1420-1423.
6 4 30 Shpitser I. Counterfactual graphical models for longitudinal mediation analysis
7 5 with unobserved confounding. *Cognitive Science* 2013;37:1011-1035.
8 6 31 Mansournia MA, Hernán MA, Greenland S. Matched designs and causal
9 7 diagrams. *Int J Epidemiol*. 2013;42:860-869.
10 8 32 Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal
11 9 inference. *Stat Sci* 1999;14:29-46.
12 10 33 Mansournia MA, Greenland S. The relation of collapsibility and confounding to
13 11 faithfulness and stability. *Epidemiology*. 2015;26:466-72.
14 12 34 Clausen TD, Mathiesen ER, Hansen T, et al. Overweight and the metabolic
15 13 syndrome in adult offspring of women with diet-treated gestational diabetes
16 14 mellitus or type 1 diabetes.[J]. *J Clin Endocrinol Metab*, 2013, 94:2464-70.
17 15 35 Cheng JY, Wong VW, Tse YK, et al. Metabolic syndrome increases
18 16 cardiovascular events but not hepatic events and death in patients with chronic
19 17 hepatitis B[J]. *Hepatology*, 2016, 64:1507.
20 18 36 Mansournia MA, Danaei G, Forouzanfar MH, Mahmoudi M, Jamali M,
21 19 Mansournia N, Mohammad K. Effect of physical activity on functional
22 20 performance and knee pain in patients with osteoarthritis: analysis with marginal
23 21 structural models. *Epidemiology*. 2012;23:631-40.
24 22 37 VanderWeele TJ, Mumford SL, Schisterman EF. Conditioning on intermediates
25 23 in perinatal epidemiology. *Epidemiology*. 2012;23(1):1-9.
26 24 38 Lange T, Rasmussen M, Thygesen LC. Assessing natural direct and indirect
27 25 effects through multiple pathways. *American Journal of Epidemiology*. 2013;179,
28 26 513-518.
29 27 39 Shpitser, Ilya and VanderWeele, Tyler J. "A complete graphical criterion for the
30 28 adjustment formula in mediation analysis," *The International Journal of*
31 29 *Biostatistics*.2011;7,1-24.
32 30 40 Albert J M, Wang W. Sensitivity analyses for parametric causal mediation effect
33 31 estimation [J]. *Biostatistics*, 2015, 16(2): 339-351.
34 32 41 Wang W, Zhang B. Assessing natural direct and indirect effects for a continuous
35 33 exposure and a dichotomous outcome[J]. *Journal of Statistical Theory and*
36 34 *Practice*, 2016, 10(3): 574-587.
37 35 42 Li H, Yuan Z, Su P, et al. A simulation study on matched case-control designs in
38 36 the perspective of causal diagrams[J]. *BMC Med Res Methodol*, 2016,16,102 .

38 **Figure 1:** Six causal diagrams were designed for estimating the causal effect of E on
39 D . a) a single mediator M ; b) two series mediators M_1 and M_2 ; c) two independent
40 parallel mediators M_1 and M_2 ; d) two correlated parallel mediators M_1 and M_2 ; e) a
41 single mediator with an unobserved confounder U ; f) two independent parallel
42 mediators M_1 and M_2 with an unobserved confounder U .

43
44 **Figure 2:** The biases with the effects of $E \rightarrow M$ (red) and $M \rightarrow D$ (blue) increasing,

1 respectively. Comparison of the bias of different effects in adjustment mediator. The
2 OR of target effect (e.g. $E \rightarrow M$) from 1 to 10 given other effects fixed ln2 in Figure
3 2A. The OR of the effect of $M \rightarrow D$ from 1 to 10 with the effect of $E \rightarrow M$ being equal
4 to zero in Figure 2B (Color figure online).

5
6 **Figure 3:** Illustrating the use of positive and negative signs on edges $E \rightarrow M$, $M \rightarrow D$
7 and $E \rightarrow D$.

8
9 **Figure 4:** The biases with the effects of $E \rightarrow M_1$ (red), $M_1 \rightarrow M_2$ (blue) and $M_2 \rightarrow D$
10 (black) increasing, respectively. Comparison of the bias of different effects in three
11 adjustment models: A) adjustment for M_1 , B) adjustment for M_2 and C) adjustment for
12 M_1 and M_2 . The OR of target effect (e.g. $E \rightarrow M_1$) from 1 to 10 given the effect of M_1
13 $\rightarrow M_2$ fixed ln8 and other effects fixed ln2 in Figure 4 (Color figure online).

14
15 **Figure 5:** The biases with the effects of $E \rightarrow M_1$ (red), $E \rightarrow M_2$ (blue), $M_1 \rightarrow D$ (black)
16 and $M_2 \rightarrow D$ (green) increasing, respectively. Comparison of the bias of different
17 effects in three adjustment models: A) adjustment for M_1 , B) adjustment for M_2 and C)
18 adjustment for M_1 and M_2 . The OR of target effects (e.g. $E \rightarrow M_1$) from 1 to 10 given
19 other edges effects fixed ln2 in Figure 5 (Color figure online).

20
21 **Figure 6:** The biases with the effects of $E \rightarrow M_1$ (red), $E \rightarrow M_2$ (blue), $M_1 \rightarrow D$ (black),
22 $M_2 \rightarrow D$ (green) and the effect of $M_2 \rightarrow M_1$ (purple) increasing, respectively.
23 Comparison of the bias of different effects in three adjustment models: A) adjustment
24 for M_1 , B) adjustment for M_2 and C) adjustment for M_1 and M_2 . The OR of target
25 effects (e.g. $E \rightarrow M_1$) from 1 to 10 given other effects fixed ln2 in Figure 6 (Color
26 figure online).

27
28 **Figure 7:** The biases with the effects of $E \rightarrow M$ (red) and $M \rightarrow D$ (blue) respectively.
29 Comparison of the bias of different effects in adjustment mediator M . The OR of
30 target effects (e.g. $E \rightarrow M$) from 1 to 10 given the effects of causal edges fixed ln2 and
31 the effect of confounder edges fixed ln5 in Figure ln8 (Color figure online).

32
33 **Figure 8:** The biases with the effects of $E \rightarrow M_1$ (red), $E \rightarrow M_2$ (blue), $M_1 \rightarrow D$ (black)
34 and $M_2 \rightarrow D$ (green) respectively. Comparison of the bias of different effects in three
35 adjustment models: A) adjustment for M_1 , B) adjustment for M_2 , and C) adjustment
36 for M_1 and M_2 . The OR of target effects (e.g. $E \rightarrow M_1$) from 1 to 10 given the effects of
37 causal edges fixed ln2 and the effect of confounder edges fixed ln5 in Figure 8 (Color
38 figure online).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

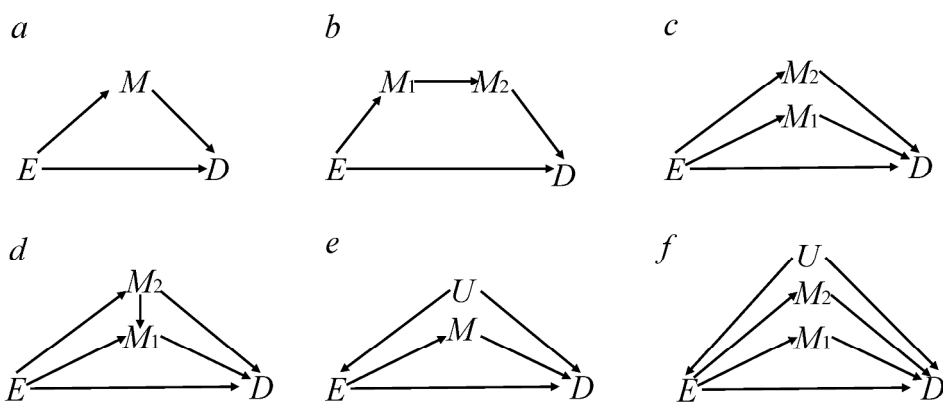


Figure 1: Six causal diagrams were designed for estimating the causal effect of E on D.

252x110mm (300 x 300 DPI)

er review only

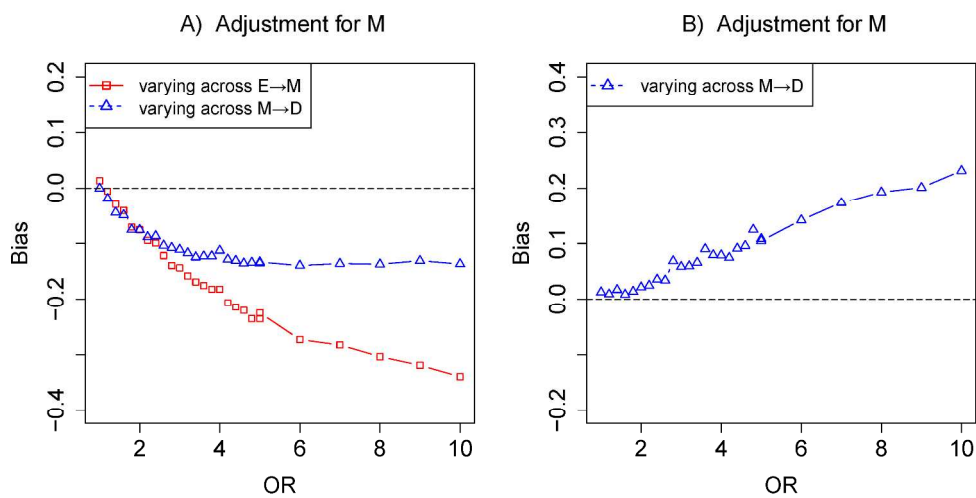


Figure 2 : The biases with the effects E→M (red) and M→D (blue) increasing, respectively.

281x148mm (300 x 300 DPI)

review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

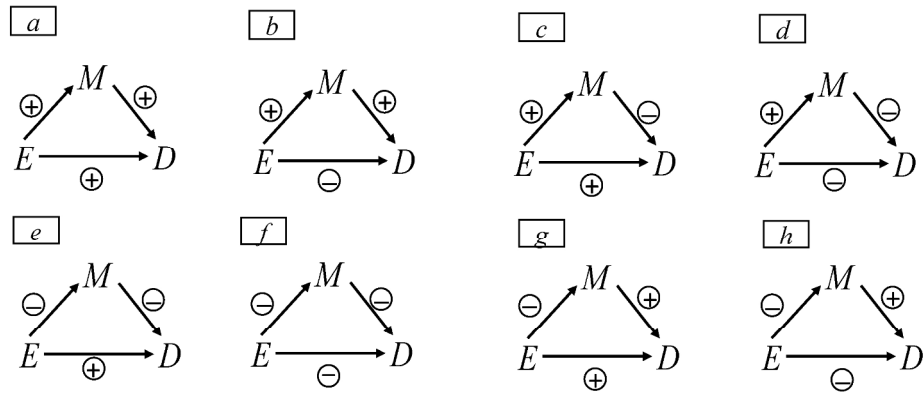


Figure 3: Illustrating the use of positive and negative signs on edges $E \rightarrow M$, $M \rightarrow D$ and $E \rightarrow D$.

237x106mm (300 x 300 DPI)

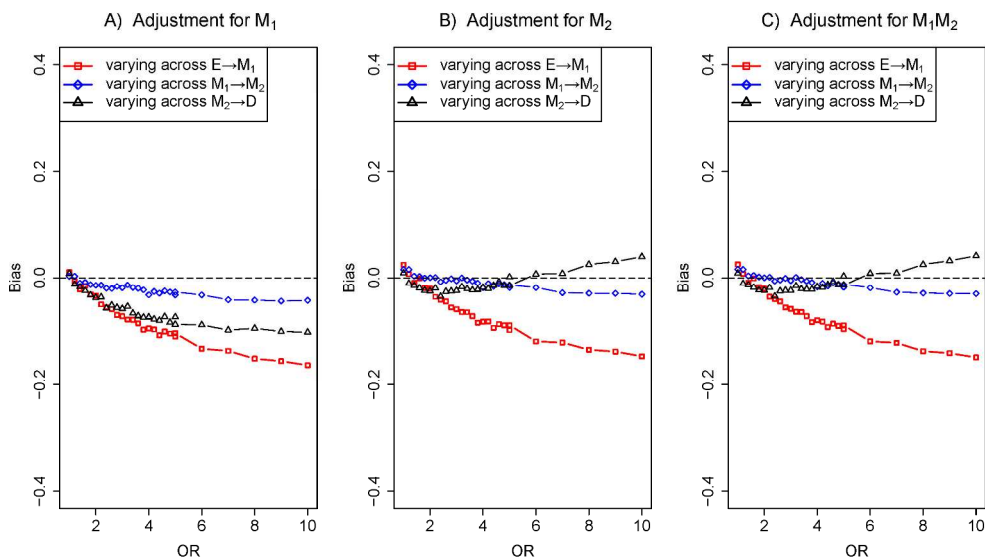


Figure 4: The biases with the effects E→M₁ (red), M₁→M₂ (blue) and M₂→D (black) increasing, respectively.

270x155mm (300 x 300 DPI)

Review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

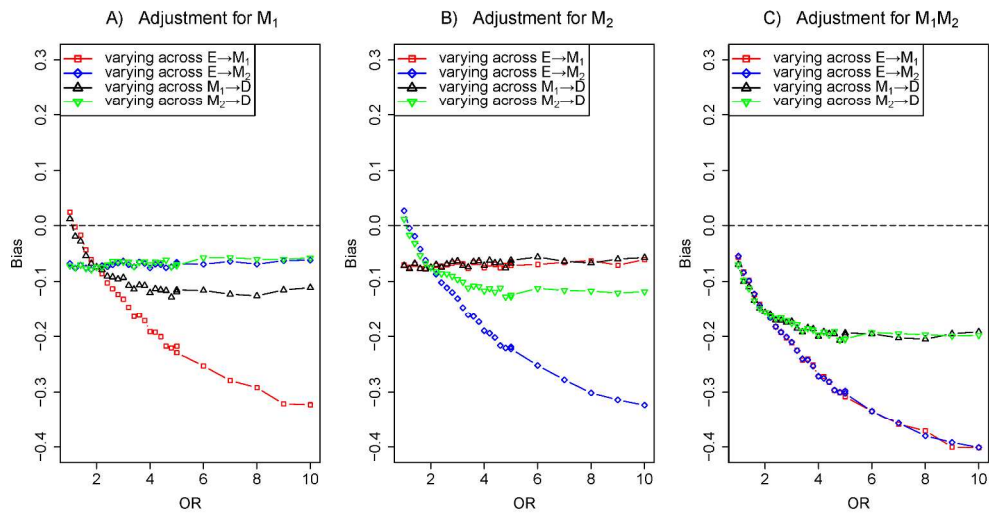


Figure 5 : The biases with the effects $E \rightarrow M_1$ (red), $E \rightarrow M_2$ (blue), $M_1 \rightarrow D$ (black) and $M_2 \rightarrow D$ (green) increasing, respectively.

279x147mm (300 x 300 DPI)

review only

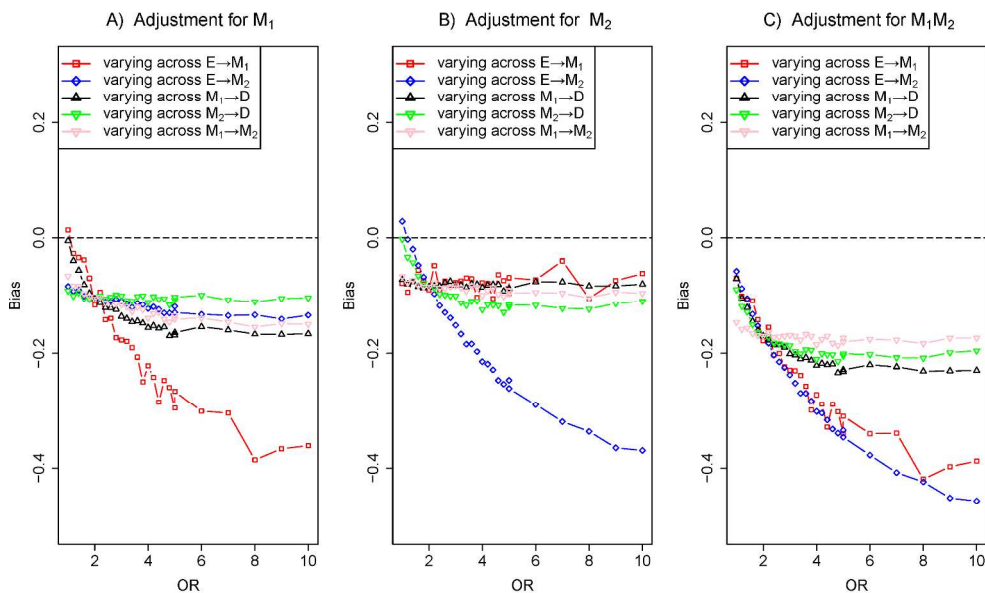


Figure 6: The biases with the effects $E \rightarrow M_1$ (red), $E \rightarrow M_2$ (blue), $M_1 \rightarrow D$ (black), $M_2 \rightarrow D$ (green) and the effect $M_2 \rightarrow M_1$ (purple) increasing, respectively.

278x169mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

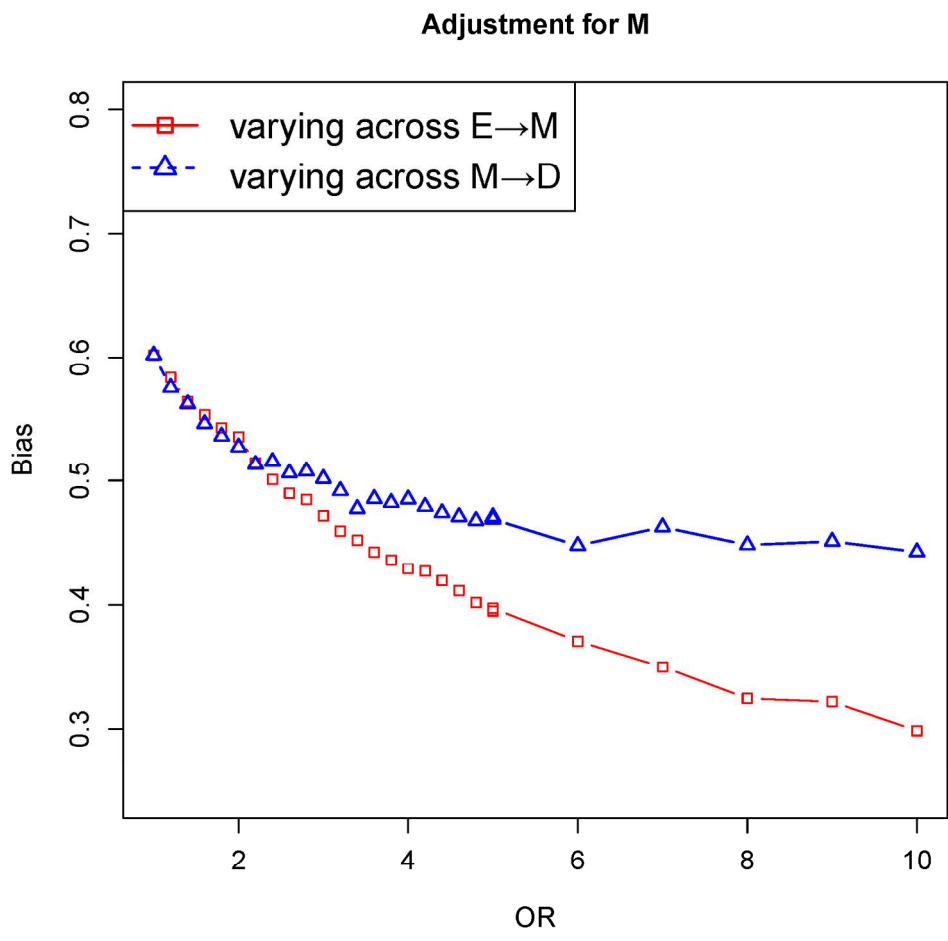


Figure 7: The biases with the effects E→M (red) and M→D (blue) respectively.

177x177mm (300 x 300 DPI)

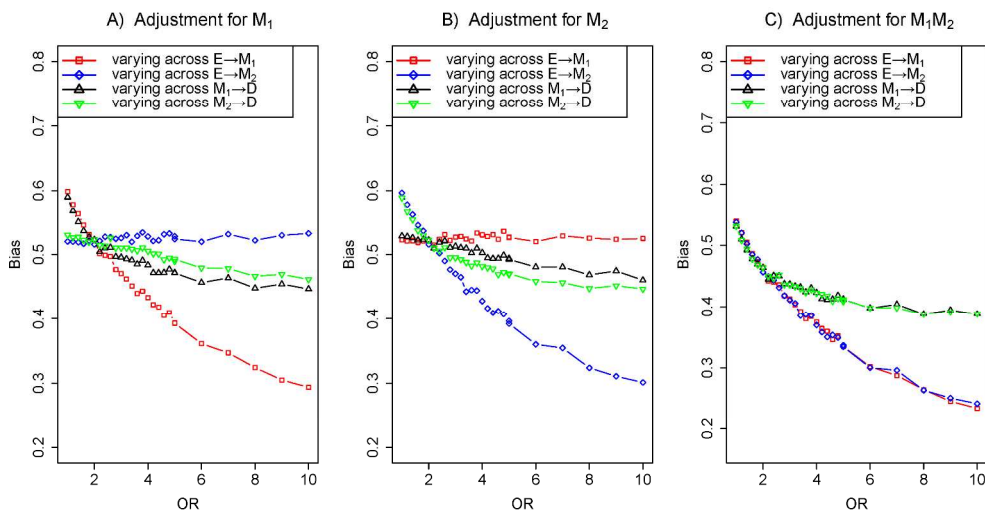


Figure 8 : The biases with the effects E→M₁ (red), E→M₂ (blue), M₁→D (black) and M₂→D (green) respectively.

281x148mm (300 x 300 DPI)

review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Appendix:

The effect of adjusting for mediator was biased for estimating the total effect of exposure on outcome using logistic regression model. Theoretical derivation of Figure 1a as follow:

Suppose the logistic models among E , M and D are:

$$\text{logit}\{P(D=1|e,m)\} = \alpha_1 + \beta_0 e + \beta_2 m,$$

$$\text{logit}\{P(M=1|e)\} = \alpha_0 + \beta_1 e.$$

The total effect ($\beta_{E \rightarrow D}^{TE}$) of exposure E on outcome D on the odds ratio ($OR_{E \rightarrow D}^{TE}$) scale was equal to

$$\begin{aligned} \beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\ &= \log \left\{ \frac{P(D_e = 1) / \{1 - P(D_e = 1)\}}{P(D_{e^*} = 1) / \{1 - P(D_{e^*} = 1)\}} \right\} \\ &= \log \left\{ \frac{P(D_e = 1) \times \{1 - P(D_{e^*} = 1)\}}{\{1 - P(D_e = 1)\} \times P(D_{e^*} = 1)} \right\} \\ &= \log \left\{ \frac{P(D=1|e=1) \times P(D=0|e^*=0)}{P(D=0|e=1) \times P(D=1|e^*=0)} \right\} \\ &= \log \left\{ \frac{\left[\sum_m P(D=1|e=1,m)P(m|e=1) \right] \times \left[\sum_m P(D=0|e^*=0,m)P(m|e^*=0) \right]}{\left[\sum_m P(D=0|e=1,m)P(m|e=1) \right] \times \left[\sum_m P(D=1|e^*=0,m)P(m|e^*=0) \right]} \right\} \end{aligned}$$

The effect ($\beta_{ED|M}(m)$) of adjusting for mediator M by logistic regression model is given

$$\begin{aligned} \beta_{ED|M}(m) &= \text{logit}\{P(D=1|e=1,m)\} - \text{logit}\{P(D=1|e^*=0,m)\} \\ &= \log \left\{ \frac{P(D=1|e=1,m) \times P(D=0|e^*=0,m)}{P(D=0|e=1,m) \times P(D=1|e^*=0,m)} \right\} \\ &= \beta_0 \end{aligned}$$

Therefore,

$$\begin{aligned}
 \text{bias} &= \beta_0 - \log(OR_{E \rightarrow D}^{TE}) \\
 &= \log \left\{ \frac{\exp(\beta_0)}{\exp(\beta_0) \frac{\exp(\beta_2) \times A_1 + \exp(\beta_2) \times B_1 + C_1 + D_1}{\exp(\beta_2) \times A_1 + B_1 + \exp(\beta_2) \times C_1 + D_1}} \right\} \\
 &= \log \left\{ \frac{\exp(\beta_2) \times A_1 + B_1 + \exp(\beta_2) \times C_1 + D_1}{\exp(\beta_2) \times A_1 + \exp(\beta_2) \times B_1 + C_1 + D_1} \right\}
 \end{aligned}$$

where

$$A_1 = \exp(\beta_1 + \alpha_0) \times (1 + \exp(\beta_0 + \alpha_1)) \times \exp(\alpha_0) \times (1 + \exp(\alpha_1))$$

$$B_1 = \exp(\beta_1 + \alpha_0) \times (1 + \exp(\beta_0 + \alpha_1)) \times (1 + \exp(\beta_2 + \alpha_1))$$

$$C_1 = (1 + \exp(\beta_0 + \beta_2 + \alpha_1)) \times \exp(\alpha_0) \times (1 + \exp(\alpha_1))$$

$$D_1 = (1 + \exp(\beta_0 + \beta_2 + \alpha_1)) \times (1 + \exp(\beta_2 + \alpha_1))$$

Focusing on the difference of between $\exp(\beta_2) \times B_1 + C_1$ and $B_1 + \exp(\beta_2) \times C_1$.

$$\begin{aligned}
 T(\beta_1) &= \exp(\beta_2) \times B_1 + C_1 - (B_1 + \exp(\beta_2) \times C_1) \\
 &= \exp(\beta_2) \times (B_1 - C_1) - (B_1 - C_1) \\
 &= (\exp(\beta_2) - 1) \times (B_1 - C_1) \\
 &= (\exp(\beta_2) - 1) \times (\exp(\beta_1 + \alpha_0) \times (1 + \exp(\beta_0 + \alpha_1)) \times (1 + \exp(\beta_2 + \alpha_1)) \\
 &\quad - (1 + \exp(\beta_0 + \beta_2 + \alpha_1)) \times \exp(\alpha_0) \times (1 + \exp(\alpha_1))) \\
 &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \times [\exp(\beta_1) \times (1 + \exp(\beta_0 + \alpha_1)) \times (1 + \exp(\beta_2 + \alpha_1)) \\
 &\quad - (1 + \exp(\beta_0 + \beta_2 + \alpha_1)) \times (1 + \exp(\alpha_1))]
 \end{aligned}$$

Then, detailed dissection:

1: $\beta_2 = 0$, $\text{bias} = 0$.

2: $\beta_2 > 0$,

① $\beta_1 = 0$: (i) $\beta_0 = 0$, $\text{bias} = 0$; (ii) $\beta_0 > 0$, $\text{bias} > 0$; (iii) $\beta_0 < 0$, $\text{bias} < 0$.

② $\beta_1 < 0$: (i) $\beta_0 = 0$, $\text{bias} > 0$; (ii) $\beta_0 > 0$, $\text{bias} > 0$; (iii) $\beta_0 < 0$, $\text{bias} > 0$.

proof (iii)

$$\begin{aligned}
 T(\beta_1) &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\
 &\quad \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\
 &\quad - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \}
 \end{aligned}$$

when $\beta_0 < 0$ and $\beta_2 > 0 \Rightarrow \exp(\beta_0) - 1 < 0 \quad \exp(\beta_2) - 1 > 0$

According to $(a-1)(b-1) = ab - a - b + 1$, when $(a-1)(b-1) < 0 \Rightarrow ab + 1 < a + b$

$$\begin{aligned}
 & 1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\
 & < 1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\
 & \Rightarrow \exp(\beta_0 + \beta_2) + 1 < \exp(\beta_0) + \exp(\beta_2)
 \end{aligned}$$

when

$$\beta_1 < \log \left\{ \frac{\exp(\beta_0 + \beta_2) + 1}{\exp(\beta_0) + \exp(\beta_2)} \right\} < 0$$

$$\beta_1 < \log \left\{ \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} \right\} < 0$$

$$\Rightarrow \exp(\beta_1) < \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} < 1$$

$$\begin{aligned}
 \Rightarrow T(\beta_1) &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\
 &\quad \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\
 &\quad - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \} \\
 &< 0
 \end{aligned}$$

Therefore, when $\beta_2 > 0$, $\beta_1 < 0$, $\beta_0 < 0$, then $bias > 0$.

③ $\beta_1 > 0$: (i) $\beta_0 = 0$, $bias < 0$; (ii) $\beta_0 < 0$, $bias < 0$; (iii) $\beta_0 > 0$, $bias < 0$.

proof (iii)

$$\begin{aligned}
 T(\beta_1) &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\
 &\quad \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\
 &\quad - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \}
 \end{aligned}$$

when $\beta_0 > 0$ and $\beta_2 > 0 \Rightarrow \exp(\beta_0) - 1 > 0 \quad \exp(\beta_2) - 1 > 0$

According to $(a-1)(b-1) = ab - a - b + 1$, when $ab > 0 \Rightarrow ab + 1 > a + b$

$$\begin{aligned}
 & 1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\
 & > 1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\
 & \Rightarrow \exp(\beta_0 + \beta_2) + 1 > \exp(\beta_0) + \exp(\beta_2)
 \end{aligned}$$

when

$$\beta_1 > \log \left\{ \frac{\exp(\beta_0 + \beta_2) + 1}{\exp(\beta_0) + \exp(\beta_2)} \right\} > 0$$

$$\beta_1 > \log \left\{ \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} \right\} > 0$$

$$\Rightarrow \exp(\beta_1) > \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} > 1$$

$$\begin{aligned} \Rightarrow T(\beta_1) &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\ &\quad \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\ &\quad - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \} \\ &> 0 \end{aligned}$$

Therefore, when $\beta_2 > 0$, $\beta_1 > 0$, $\beta_0 > 0$, then *bias* < 0.

3: $\beta_2 < 0$,

① $\beta_1 = 0$: (i) $\beta_0 = 0$, *bias* = 0; (ii) $\beta_0 > 0$, *bias* > 0; (iii) $\beta_0 < 0$, *bias* < 0.

② $\beta_1 < 0$: (i) $\beta_0 = 0$, *bias* < 0; (ii) $\beta_0 < 0$, *bias* < 0; (iii) $\beta_0 > 0$, *bias* < 0.

proof (iii)

$$\begin{aligned} T(\beta_1) &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\ &\quad \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\ &\quad - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \} \end{aligned}$$

when $\beta_0 > 0$ and $\beta_2 < 0 \Rightarrow \exp(\beta_0) - 1 > 0 \quad \exp(\beta_2) - 1 < 0$

According to $(a-1)(b-1) = ab - a - b + 1$, when $ab < 0 \Rightarrow ab + 1 < a + b$

$$\begin{aligned} 1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\ < 1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\ \Rightarrow \exp(\beta_0 + \beta_2) + 1 < \exp(\beta_0) + \exp(\beta_2) \end{aligned}$$

when

$$\beta_1 < \log \left\{ \frac{\exp(\beta_0 + \beta_2) + 1}{\exp(\beta_0) + \exp(\beta_2)} \right\} < 0$$

$$\beta_1 < \log \left\{ \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} \right\} < 0$$

$$\Rightarrow \exp(\beta_1) < \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} < 1$$

$$\begin{aligned} \Rightarrow T(\beta_1) &= (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\ &\quad \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\ &\quad - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \} \\ &> 0 \end{aligned}$$

Therefore, when $\beta_2 < 0$, $\beta_1 < 0$, $\beta_0 > 0$, then *bias* < 0.

③ $\beta_1 > 0$: (i) $\beta_0 = 0$, *bias* > 0; (ii) $\beta_0 > 0$, *bias* > 0; (iii) $\beta_0 < 0$, *bias* > 0.

proof (iii)

$$T(\beta_1) = (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\ \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\ - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \}$$

when $\beta_0 < 0$ and $\beta_2 < 0 \Rightarrow \exp(\beta_0) - 1 < 0 \quad \exp(\beta_2) - 1 < 0$

According to $(a-1)(b-1) = ab - a - b + 1$, when $ab > 0 \Rightarrow ab + 1 > a + b$

$$1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\ > 1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1) \\ \Rightarrow \exp(\beta_0 + \beta_2) + 1 > \exp(\beta_0) + \exp(\beta_2)$$

when

$$\beta_1 > \log \left\{ \frac{\exp(\beta_0 + \beta_2) + 1}{\exp(\beta_0) + \exp(\beta_2)} \right\} > 0$$

$$\beta_1 > \log \left\{ \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} \right\} > 0$$

$$\Rightarrow \exp(\beta_1) > \frac{1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)}{1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)} > 1$$

$$\Rightarrow T(\beta_1) = (\exp(\beta_2) - 1) \times \exp(\alpha_0) \\ \times \{ \exp(\beta_1) \times [1 + \exp(\beta_0 + \alpha_1) + \exp(\beta_2 + \alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \\ - [1 + \exp(\beta_0 + \beta_2 + \alpha_1) + \exp(\alpha_1) + \exp(\beta_0 + \beta_2 + 2\alpha_1)] \} \\ < 0$$

Therefore, when $\beta_2 < 0$, $\beta_1 > 0$, $\beta_0 < 0$, then $bias > 0$.

In conclusion:

1: $\beta_2 = 0$, $bias = 0$.

2: $\beta_2 \neq 0$, $\beta_1 = 0$: (i) $\beta_0 = 0$, $bias = 0$; (ii) $\beta_0 > 0$, $bias > 0$; (iii) $\beta_0 < 0$, $bias < 0$.

3: (i) $\beta_1\beta_2 > 0$, $bias < 0$. (ii) $\beta_1\beta_2 < 0$, $bias > 0$.

Supplementary A

The theoretical results of others causal diagrams (Figure 1b-Figure 1f) have been shown in the supplementary of manuscript.

(1) Figure 1(b) is a depiction through two series mediators, decomposing total effects into direct ($E \rightarrow D$) and indirect ($E \rightarrow M_1 \rightarrow M_2 \rightarrow D$) components.

On the odds ratio ($OR_{E \rightarrow D}^{TE}$) scale, the total effect ($\beta_{E \rightarrow D}^{TE} = \log(OR_{E \rightarrow D}^{TE})$), comparing exposure level e with e^* , we could obtain the total effect:

$$\begin{aligned} \beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\ &= \log \left\{ \frac{P(D_e = 1) / \{1 - P(D_e = 1)\}}{P(D_{e^*} = 1) / \{1 - P(D_{e^*} = 1)\}} \right\} \\ &= \log \left\{ \frac{P(D_e = 1) \times \{1 - P(D_{e^*} = 1)\}}{\{1 - P(D_e = 1)\} \times P(D_{e^*} = 1)} \right\} \\ &= \log \left\{ \frac{P(D = 1 | e = 1) \times P(D = 0 | e^* = 0)}{P(D = 0 | e = 1) \times P(D = 1 | e^* = 0)} \right\} \\ &= \log \left\{ \frac{\xi_1}{\xi_2} \right\} \end{aligned}$$

$$\xi_1 = \left[\sum_{m_1 m_2} P(D = 1 | e = 1, m_2) P(m_2 | m_1) P(m_1 | e = 1) \right] \times \left[\sum_{m_1 m_2} P(D = 0 | e^* = 0, m_2) P(m_2 | m_1) P(m_1 | e^* = 0) \right]$$

$$\xi_2 = \left[\sum_{m_1 m_2} P(D = 0 | e = 1, m_2) P(m_2 | m_1) P(m_1 | e = 1) \right] \times \left[\sum_{m_1 m_2} P(D = 1 | e^* = 0, m_2) P(m_2 | m_1) P(m_1 | e^* = 0) \right]$$

The effect ($\beta_{ED|M_1}(m_1)$) of adjusting for mediator M_1 by logistic regression model can

be given

$$\begin{aligned} \beta_{ED|M_1}(m_1) &= \text{logit}\{P(D = 1 | e = 1, m_1)\} - \text{logit}\{P(D = 1 | e^* = 0, m_1)\} \\ &= \log \left\{ \frac{P(D = 1 | e = 1, m_1) P(D = 0 | e^* = 0, m_1)}{P(D = 0 | e = 1, m_1) P(D = 1 | e^* = 0, m_1)} \right\} \\ &= \log \left\{ \frac{\left[\sum_{m_2} P(D = 1 | e = 1, m_2) P(m_2 | m_1) \right] \times \left[\sum_{m_2} P(D = 0 | e^* = 0, m_2) P(m_2 | m_1) \right]}{\left[\sum_{m_2} P(D = 0 | e = 1, m_2) P(m_2 | m_1) \right] \times \left[\sum_{m_2} P(D = 1 | e^* = 0, m_2) P(m_2 | m_1) \right]} \right\} \end{aligned}$$

The effect ($\beta_{ED|M_2}(m_2)$) of adjusting for mediator M_2 by logistic regression model can

be given

$$\begin{aligned} \beta_{ED|M_2}(m_2) &= \text{logit}\{P(D = 1 | e = 1, m_2)\} - \text{logit}\{P(D = 1 | e^* = 0, m_2)\} \\ &= \log \left\{ \frac{P(D = 1 | e = 1, m_2) P(D = 0 | e^* = 0, m_2)}{P(D = 0 | e = 1, m_2) P(D = 1 | e^* = 0, m_2)} \right\} \end{aligned}$$

The effect ($\beta_{ED|M_1, M_2}(m_1, m_2)$) of adjusting for mediator M_1 M_2 by logistic regression model can be given

$$\begin{aligned}\beta_{ED|M_1, M_2}(m_1, m_2) &= \text{logit}\{P(D=1|e=1, m_1, m_2)\} - \text{logit}\{P(D=1|e^*=0, m_1, m_2)\} \\ &= \log\left\{\frac{P(D=1|e=1, m_1, m_2)P(D=0|e^*=0, m_1, m_2)}{P(D=0|e=1, m_1, m_2)P(D=1|e^*=0, m_1, m_2)}\right\} \\ &= \log\left\{\frac{P(D=1|e=1, m_2)P(D=0|e^*=0, m_2)}{P(D=0|e=1, m_2)P(D=1|e^*=0, m_2)}\right\}\end{aligned}$$

Therefore, we could evaluate the biases that contains three adjustment models: A) adjustment for M_1 , $\text{bias}(m_1) = \beta_{ED|M_1}(m_1) - \beta_{E \rightarrow D}^{TE}$; B) adjustment for M_2 , $\text{bias}(m_2) = \beta_{ED|M_2}(m_2) - \beta_{E \rightarrow D}^{TE}$ and C) adjustment for M_1 and M_2 , $\text{bias}(m_1, m_2) = \beta_{ED|M_1, M_2}(m_1, m_2) - \beta_{E \rightarrow D}^{TE}$.

(2) Figure 1c shows that the exposure E independently causes M_1 and M_2 and indirectly influences the outcome D through M_1 and M_2 , forming three causal paths $E \rightarrow D$, $E \rightarrow M_1 \rightarrow D$ and $E \rightarrow M_2 \rightarrow D$.

On the odds ratio ($OR_{E \rightarrow D}^{TE}$) scale, the total effect ($\beta_{E \rightarrow D}^{TE} = \log(OR_{E \rightarrow D}^{TE})$), comparing exposure level e with e^* , we could obtain the total effect:

$$\begin{aligned}\beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\ &= \log\left\{\frac{P(D_e=1)/\{1-P(D_e=1)\}}{P(D_{e^*}=1)/\{1-P(D_{e^*}=1)\}}\right\} \\ &= \log\left\{\frac{P(D_e=1) \times \{1-P(D_{e^*}=1)\}}{\{1-P(D_e=1)\} \times P(D_{e^*}=1)}\right\} \\ &= \log\left\{\frac{P(D=1|e=1) \times P(D=0|e^*=0)}{P(D=0|e=1) \times P(D=1|e^*=0)}\right\} \\ &= \log\left\{\frac{\xi_1}{\xi_2}\right\}\end{aligned}$$

$$\xi_1 = \left[\sum_{m_1, m_2} P(D=1|e=1, m_1, m_2)P(m_2|e=1)P(m_1|e=1)\right] \times \left[\sum_{m_1, m_2} P(D=0|e^*=0, m_1, m_2)P(m_2|e^*=0)P(m_1|e^*=0)\right]$$

$$\xi_2 = \left[\sum_{m_1, m_2} P(D=0|e=1, m_1, m_2)P(m_2|e=1)P(m_1|e=1)\right] \times \left[\sum_{m_1, m_2} P(D=1|e^*=0, m_1, m_2)P(m_2|e^*=0)P(m_1|e^*=0)\right]$$

The effect ($\beta_{ED|M_1}(m_1)$) of adjusting for mediator M_1 by logistic regression model can be given

$$\begin{aligned} \beta_{ED|M_1}(m_1) &= \text{logit}\{P(D=1|e=1,m_1)\} - \text{logit}\{P(D=1|e^*=0,m_1)\} \\ &= \log\left\{\frac{P(D=1|e=1,m_1)P(D=0|e^*=0,m_1)}{P(D=0|e=1,m_1)P(D=1|e^*=0,m_1)}\right\} \\ &= \log\left\{\frac{\left[\sum_{m_2} P(D=1|e=1,m_1,m_2)P(m_2|e=1)\right] \times \left[\sum_{m_2} P(D=0|e^*=0,m_1,m_2)P(m_2|e^*=0)\right]}{\left[\sum_{m_2} P(D=0|e=1,m_1,m_2)P(m_2|e=1)\right] \times \left[\sum_{m_2} P(D=1|e^*=0,m_1,m_2)P(m_2|e^*=0)\right]}\right\} \end{aligned}$$

The effect ($\beta_{ED|M_2}(m_2)$) of adjusting for mediator M_2 by logistic regression model can

be given

$$\begin{aligned} \beta_{ED|M_2}(m_2) &= \text{logit}\{P(D=1|e=1,m_2)\} - \text{logit}\{P(D=1|e^*=0,m_2)\} \\ &= \log\left\{\frac{P(D=1|e=1,m_2)P(D=0|e^*=0,m_2)}{P(D=0|e=1,m_2)P(D=1|e^*=0,m_2)}\right\} \\ &= \log\left\{\frac{\left[\sum_{m_1} P(D=1|e=1,m_1,m_2)P(m_1|e=1)\right] \times \left[\sum_{m_1} P(D=0|e^*=0,m_1,m_2)P(m_1|e^*=0)\right]}{\left[\sum_{m_1} P(D=0|e=1,m_1,m_2)P(m_1|e=1)\right] \times \left[\sum_{m_1} P(D=1|e^*=0,m_1,m_2)P(m_1|e^*=0)\right]}\right\} \end{aligned}$$

The effect ($\beta_{ED|M_1,M_2}(m_1,m_2)$) of adjusting for mediator $M_1 M_2$ by logistic regression

model can be given

$$\begin{aligned} \beta_{ED|M_1,M_2}(m_1,m_2) &= \text{logit}\{P(D=1|e=1,m_1,m_2)\} - \text{logit}\{P(D=1|e^*=0,m_1,m_2)\} \\ &= \log\left\{\frac{P(D=1|e=1,m_1,m_2)P(D=0|e^*=0,m_1,m_2)}{P(D=0|e=1,m_1,m_2)P(D=1|e^*=0,m_1,m_2)}\right\} \end{aligned}$$

Therefore, we could evaluate the biases that contains three adjustment models: A)

adjustment for M_1 , $bias(m_1) = \beta_{ED|M_1}(m_1) - \beta_{E \rightarrow D}^{TE}$; B) adjustment for M_2 ,

$bias(m_2) = \beta_{ED|M_2}(m_2) - \beta_{E \rightarrow D}^{TE}$ and C) adjustment for M_1 and M_2 ,

$bias(m_1,m_2) = \beta_{ED|M_1,M_2}(m_1,m_2) - \beta_{E \rightarrow D}^{TE}$.

(3) In Figure 1d, there exists five paths from E to D : $E \rightarrow D$, $E \rightarrow M_1 \rightarrow D$, $E \rightarrow M_2 \rightarrow D$, $E \rightarrow M_1 \leftarrow M_2 \rightarrow D$ and $E \rightarrow M_2 \rightarrow M_1 \rightarrow D$. In particular, the path $E \rightarrow M_1 \leftarrow M_2 \rightarrow D$ is a blocked path, due to the M_1 being a collider node.

On the odds ratio ($OR_{E \rightarrow D}^{TE}$) scale, the total effect ($\beta_{E \rightarrow D}^{TE} = \log(OR_{E \rightarrow D}^{TE})$), comparing exposure level e with e^* , we could obtain the total effect:

$$\begin{aligned}
\beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\
&= \log \left\{ \frac{P(D_e = 1) / \{1 - P(D_e = 1)\}}{P(D_{e^*} = 1) / \{1 - P(D_{e^*} = 1)\}} \right\} \\
&= \log \left\{ \frac{P(D_e = 1) \times \{1 - P(D_{e^*} = 1)\}}{\{1 - P(D_e = 1)\} \times P(D_{e^*} = 1)} \right\} \\
&= \log \left\{ \frac{P(D = 1 | e = 1) \times P(D = 0 | e^* = 0)}{P(D = 0 | e = 1) \times P(D = 1 | e^* = 0)} \right\} \\
&= \log \left\{ \frac{\xi_1}{\xi_2} \right\}
\end{aligned}$$

$$\begin{aligned}
\xi_1 &= \left[\sum_{m_1, m_2} P(D = 1 | e = 1, m_1, m_2) P(m_2 | e = 1) P(m_1 | e = 1, m_2) \right] \\
&\quad \times \left[\sum_{m_1, m_2} P(D = 0 | e^* = 0, m_1, m_2) P(m_2 | e^* = 0) P(m_1 | e^* = 0, m_2) \right] \\
\xi_2 &= \left[\sum_{m_1, m_2} P(D = 0 | e = 1, m_1, m_2) P(m_2 | e = 1) P(m_1 | e = 1, m_2) \right] \\
&\quad \times \left[\sum_{m_1, m_2} P(D = 1 | e^* = 0, m_1, m_2) P(m_2 | e^* = 0) P(m_1 | e^* = 0, m_2) \right]
\end{aligned}$$

The effect ($\beta_{ED|M_1}(m_1)$) of adjusting for mediator M_1 by logistic regression model can

be given

$$\begin{aligned}
\beta_{ED|M_1}(m_1) &= \text{logit}\{P(D = 1 | e = 1, m_1)\} - \text{logit}\{P(D = 1 | e^* = 0, m_1)\} \\
&= \log \left\{ \frac{P(D = 1 | e = 1, m_1) P(D = 0 | e^* = 0, m_1)}{P(D = 0 | e = 1, m_1) P(D = 1 | e^* = 0, m_1)} \right\} \\
&= \log \left\{ \frac{\left[\sum_{m_2} P(D = 1 | e = 1, m_1, m_2) P(m_2 | e = 1, m_1) \right] \times \left[\sum_{m_2} P(D = 0 | e^* = 0, m_1, m_2) P(m_2 | e^* = 0, m_1) \right]}{\left[\sum_{m_2} P(D = 0 | e = 1, m_1, m_2) P(m_2 | e = 1, m_1) \right] \times \left[\sum_{m_2} P(D = 1 | e^* = 0, m_1, m_2) P(m_2 | e^* = 0, m_1) \right]} \right\} \\
&= \log \left\{ \frac{\xi_1}{\xi_2} \right\}
\end{aligned}$$

$$\begin{aligned}
\xi_1 &= \left[\sum_{m_2} P(D = 1 | e = 1, m_1, m_2) \frac{P(m_1 | e = 1, m_2) P(m_2 | e = 1)}{\sum_{m_2} P(m_1 | e = 1, m_2) P(m_2 | e = 1)} \right] \\
&\quad \times \left[\sum_{m_2} P(D = 0 | e^* = 0, m_1, m_2) \frac{P(m_1 | e^* = 0, m_2) P(m_2 | e^* = 0)}{\sum_{m_2} P(m_1 | e^* = 0, m_2) P(m_2 | e^* = 0)} \right] \\
\xi_2 &= \left[\sum_{m_2} P(D = 0 | e = 1, m_1, m_2) \frac{P(m_1 | e = 1, m_2) P(m_2 | e = 1)}{\sum_{m_2} P(m_1 | e = 1, m_2) P(m_2 | e = 1)} \right] \\
&\quad \times \left[\sum_{m_2} P(D = 1 | e^* = 0, m_1, m_2) \frac{P(m_1 | e^* = 0, m_2) P(m_2 | e^* = 0)}{\sum_{m_2} P(m_1 | e^* = 0, m_2) P(m_2 | e^* = 0)} \right]
\end{aligned}$$

The effect ($\beta_{ED|M_2}(m_2)$) of adjusting for mediator M_2 by logistic regression model can

be given

$$\begin{aligned}
\beta_{ED|M_2}(m_2) &= \text{logit}\{P(D=1|e=1,m_2)\} - \text{logit}\{P(D=1|e^*=0,m_2)\} \\
&= \log \left\{ \frac{P(D=1|e=1,m_2)P(D=0|e^*=0,m_2)}{P(D=0|e=1,m_2)P(D=1|e^*=0,m_2)} \right\} \\
&= \log \left\{ \frac{\left[\sum_{m_1} P(D=1|e=1,m_1,m_2)P(m_1|e=1,m_2) \right] \times \left[\sum_{m_1} P(D=0|e^*=0,m_1,m_2)P(m_1|e^*=0,m_2) \right]}{\left[\sum_{m_1} P(D=0|e=1,m_1,m_2)P(m_1|e=1,m_2) \right] \times \left[\sum_{m_1} P(D=1|e^*=0,m_1,m_2)P(m_1|e^*=0,m_2) \right]} \right\}
\end{aligned}$$

The effect ($\beta_{ED|M_1,M_2}(m_1,m_2)$) of adjusting for mediator M_1 M_2 by logistic regression model can be given

$$\begin{aligned}
\beta_{ED|M_1,M_2}(m_1,m_2) &= \text{logit}\{P(D=1|e=1,m_1,m_2)\} - \text{logit}\{P(D=1|e^*=0,m_1,m_2)\} \\
&= \log \left\{ \frac{P(D=1|e=1,m_1,m_2)P(D=0|e^*=0,m_1,m_2)}{P(D=0|e=1,m_1,m_2)P(D=1|e^*=0,m_1,m_2)} \right\}
\end{aligned}$$

Therefore, we could evaluate the biases that contains three adjustment models: A) adjustment for M_1 , $\text{bias}(m_1) = \beta_{ED|M_1}(m_1) - \beta_{E \rightarrow D}^{TE}$; B) adjustment for M_2 , $\text{bias}(m_2) = \beta_{ED|M_2}(m_2) - \beta_{E \rightarrow D}^{TE}$ and C) adjustment for M_1 and M_2 , $\text{bias}(m_1,m_2) = \beta_{ED|M_1,M_2}(m_1,m_2) - \beta_{E \rightarrow D}^{TE}$.

(4) In Figure 1e, the causal diagrams contained a confounder of exposure-outcome relationship. On the odds ratio ($OR_{E \rightarrow D}^{TE}$) scale, the total effect ($\beta_{E \rightarrow D}^{TE} = \log(OR_{E \rightarrow D}^{TE})$), comparing exposure level e with e^* , we could obtain the total effect:

$$\begin{aligned}
\beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\
&= \log \left\{ \frac{P(D_e=1)/\{1-P(D_e=1)\}}{P(D_{e^*}=1)/\{1-P(D_{e^*}=1)\}} \right\} \\
&= \log \left\{ \frac{P(D_e=1) \times \{1-P(D_{e^*}=1)\}}{\{1-P(D_e=1)\} \times P(D_{e^*}=1)} \right\} \\
&= \log \left\{ \frac{P(D=1|e=1) \times P(D=0|e^*=0)}{P(D=0|e=1) \times P(D=1|e^*=0)} \right\} \\
&= \log \left\{ \frac{\left[\sum_{mu} P(D=1|e=1,m,u)P(m|e=1)P(u) \right] \times \left[\sum_{mu} P(D=0|e^*=0,m,u)P(m|e^*=0)P(u) \right]}{\left[\sum_{mu} P(D=0|e=1,m,u)P(m|e=1)P(u) \right] \times \left[\sum_{mu} P(D=1|e^*=0,m,u)P(m|e^*=0)P(u) \right]} \right\}
\end{aligned}$$

The effect ($\beta_{ED|M}(m)$) of adjusting for mediator M by logistic regression model can be given

$$\begin{aligned}
\beta_{ED|M}(m) &= \log it(P(D=1|e=1,m)) - \log it(P(D=1|e^*=0,m)) \\
&= \log \left\{ \frac{P(D=1|e=1,m) \times P(D=0|e^*=0,m)}{P(D=0|e=1,m) \times P(D=1|e^*=0,m)} \right\} \\
&= \log \left\{ \frac{\left[\sum_u P(D=1|e=1,m,u)p(u|e=1,m) \right] \times \left[\sum_u P(D=0|e^*=0,m,u)p(u|e^*=0,m) \right]}{\left[\sum_u P(D=0|e=1,m,u)p(u|e=1,m) \right] \times \left[\sum_u P(D=1|e^*=0,m,u)p(u|e^*=0,m) \right]} \right\} \\
&= \log \left\{ \frac{\left[\sum_u P(D=1|e=1,m,u) \frac{p(e=1|u)p(u)}{\sum_u p(e=1|u)p(u)} \right] \times \left[\sum_u P(D=0|e^*=0,m,u) \frac{p(e^*=0|u)p(u)}{\sum_u p(e^*=0|u)p(u)} \right]}{\left[\sum_u P(D=0|e=1,m,u) \frac{p(e=1|u)p(u)}{\sum_u p(e=1|u)p(u)} \right] \times \left[\sum_u P(D=1|e^*=0,m,u) \frac{p(e^*=0|u)p(u)}{\sum_u p(e^*=0|u)p(u)} \right]} \right\}
\end{aligned}$$

Therefore, we could evaluate the biases of adjustment models:

$$bias(m) = \beta_{ED|M}(m) - \beta_{E \rightarrow D}^{TE}$$

(5) Figure 1f is a depiction of two parallel mediators M_1 and M_2 with confounder.

On the odds ratio ($OR_{E \rightarrow D}^{TE}$) scale, the total effect ($\beta_{E \rightarrow D}^{TE} = \log(OR_{E \rightarrow D}^{TE})$), comparing exposure level e with e^* , we could obtain the total effect:

$$\begin{aligned}
\beta_{E \rightarrow D}^{TE} &= \log(OR_{E \rightarrow D}^{TE}) \\
&= \log \left\{ \frac{P(D_e=1) / \{1 - P(D_e=1)\}}{P(D_{e^*}=1) / \{1 - P(D_{e^*}=1)\}} \right\} \\
&= \log \left\{ \frac{P(D_e=1) \times \{1 - P(D_{e^*}=1)\}}{\{1 - P(D_e=1)\} \times P(D_{e^*}=1)} \right\} \\
&= \log \left\{ \frac{P(D=1|e=1) \times P(D=0|e^*=0)}{P(D=0|e=1) \times P(D=1|e^*=0)} \right\} \\
&= \log \left\{ \frac{\xi_1}{\xi_2} \right\}
\end{aligned}$$

$$\begin{aligned}
\xi_1 &= \left[\sum_{m_1 m_2 u} P(D=1|e=1, m_1, m_2, u) P(m_2|e=1) P(m_1|e=1) P(u) \right] \\
&\quad \times \left[\sum_{m_1 m_2 u} P(D=0|e^*=0, m_1, m_2, u) P(m_2|e^*=0) P(m_1|e^*=0) P(u) \right] \\
\xi_2 &= \left[\sum_{m_1 m_2 u} P(D=0|e=1, m_1, m_2, u) P(m_2|e=1) P(m_1|e=1) P(u) \right] \\
&\quad \times \left[\sum_{m_1 m_2 u} P(D=1|e^*=0, m_1, m_2, u) P(m_2|e^*=0) P(m_1|e^*=0) P(u) \right]
\end{aligned}$$

The effect ($\beta_{ED|M_1}(m_1)$) of adjusting for mediator M_1 by logistic regression model can be given

$$\begin{aligned}\beta_{ED|M_1}(m_1) &= \text{logit}\{P(D=1|e=1,m_1)\} - \text{logit}\{P(D=1|e^*=0,m_1)\} \\ &= \log \left\{ \frac{P(D=1|e=1,m_1)P(D=0|e^*=0,m_1)}{P(D=0|e=1,m_1)P(D=1|e^*=0,m_1)} \right\} \\ &= \log \left\{ \frac{\xi_1}{\xi_2} \right\}\end{aligned}$$

$$\begin{aligned}\xi_1 &= \left[\sum_{m_2 u} P(D=1|e=1,m_1,m_2,u) \frac{P(m_2|e=1)P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right] \\ &\quad \times \left[\sum_{m_2 u} P(D=0|e^*=0,m_1,m_2,u) \frac{P(m_2|e^*=0)P(e^*=0|u)P(u)}{\sum_u P(e^*=0|u)P(u)} \right] \\ \xi_2 &= \left[\sum_{m_2 u} P(D=0|e=1,m_1,m_2,u) \frac{P(m_2|e=1)P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right] \\ &\quad \times \left[\sum_{m_2 u} P(D=1|e^*=0,m_1,m_2,u) \frac{P(m_2|e^*=0)P(e^*=0|u)P(u)}{\sum_u P(e^*=0|u)P(u)} \right]\end{aligned}$$

The effect ($\beta_{ED|M_2}(m_2)$) of adjusting for mediator M_2 by logistic regression model can be given

$$\begin{aligned}\beta_{ED|M_2}(m_2) &= \text{logit}\{P(D=1|e=1,m_2)\} - \text{logit}\{P(D=1|e^*=0,m_2)\} \\ &= \log \left\{ \frac{P(D=1|e=1,m_2)P(D=0|e^*=0,m_2)}{P(D=0|e=1,m_2)P(D=1|e^*=0,m_2)} \right\} \\ &= \log \left\{ \frac{\xi_1}{\xi_2} \right\}\end{aligned}$$

$$\begin{aligned}\xi_1 &= \left[\sum_{m_1 u} P(D=1|e=1,m_1,m_2,u) \frac{P(m_1|e=1)P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right] \\ &\quad \times \left[\sum_{m_1 u} P(D=0|e^*=0,m_1,m_2,u) \frac{P(m_1|e^*=0)P(e^*=0|u)P(u)}{\sum_u P(e^*=0|u)P(u)} \right] \\ \xi_2 &= \left[\sum_{m_1 u} P(D=0|e=1,m_1,m_2,u) \frac{P(m_1|e=1)P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right] \\ &\quad \times \left[\sum_{m_1 u} P(D=1|e^*=0,m_1,m_2,u) \frac{P(m_1|e^*=0)P(e^*=0|u)P(u)}{\sum_u P(e^*=0|u)P(u)} \right]\end{aligned}$$

The effect ($\beta_{ED|M_1,M_2}(m_1,m_2)$) of adjusting for mediator $M_1 M_2$ by logistic regression model can be given

$$\begin{aligned}
& \beta_{ED|M_1, M_2}(m_1, m_2) \\
&= \text{logit}\{P(D=1|e=1, m_1, m_2)\} - \text{logit}\{P(D=1|e^*=0, m_1, m_2)\} \\
&= \log \left\{ \frac{P(D=1|e=1, m_1, m_2)P(D=0|e^*=0, m_1, m_2)}{P(D=0|e=1, m_1, m_2)P(D=1|e^*=0, m_1, m_2)} \right\} \\
&= \log \left\{ \frac{\left[\sum_u P(D=1|e=1, m_1, m_2, u) \frac{P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right] \times \left[\sum_u P(D=0|e^*=0, m_1, m_2, u) \frac{P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right]}{\left[\sum_u P(D=0|e=1, m_1, m_2, u) \frac{P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right] \times \left[\sum_u P(D=1|e^*=0, m_1, m_2, u) \frac{P(e=1|u)P(u)}{\sum_u P(e=1|u)P(u)} \right]} \right\}
\end{aligned}$$

Therefore, we could evaluate the biases that contains three adjustment models: A) adjustment for M_1 , $\text{bias}(m_1) = \beta_{ED|M_1}(m_1) - \beta_{E \rightarrow D}^{TE}$; B) adjustment for M_2 , $\text{bias}(m_2) = \beta_{ED|M_2}(m_2) - \beta_{E \rightarrow D}^{TE}$ and C) adjustment for M_1 and M_2 , $\text{bias}(m_1, m_2) = \beta_{ED|M_1, M_2}(m_1, m_2) - \beta_{E \rightarrow D}^{TE}$.

Supplementary B

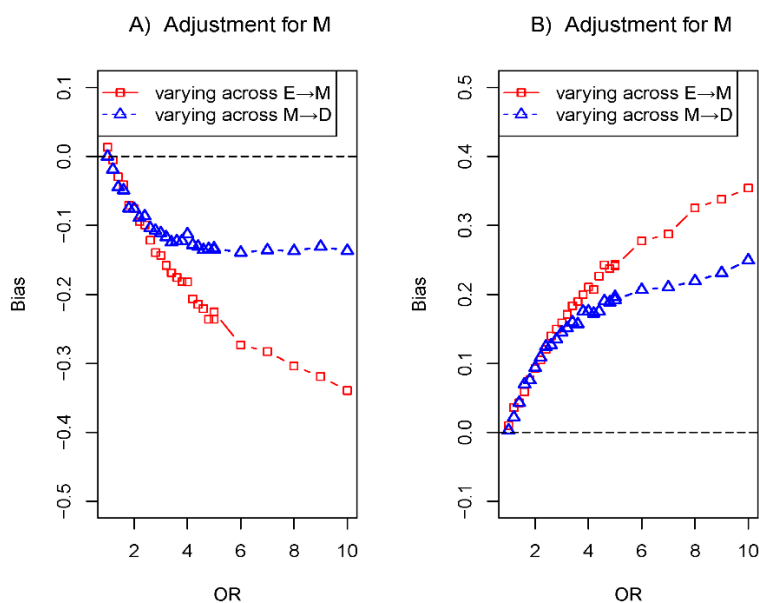


Figure S1: The biases with the effects $E \rightarrow M$ (red) and $M \rightarrow D$ (blue) increasing, respectively. Comparison of the bias of different effects in adjustment mediator.

The Figure S1-A obtained the result $\text{bias} < 0$ in Figure 3a with the effects $E \rightarrow M$, $M \rightarrow D$ and $E \rightarrow D$ fixing to $\ln 2$. The Figure S1-B gained the result $\text{bias} > 0$ in Figure 3c with the effects $E \rightarrow M$ and $E \rightarrow D$ fixing to $\ln 2$, effect $M \rightarrow D$ fixing to $-\ln 2$. We could obtain the bias performances of varying across the effects of exposure-mediator and mediator-outcome. The effect $E \rightarrow M$ of varying across was more sensitive than the effect $M \rightarrow D$ of varying across in Figure S1.

STROBE 2007 (v4) checklist of items to be included in reports of observational studies in epidemiology*
Checklist for cohort, case-control, and cross-sectional studies (combined)

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	1
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any pre-specified hypotheses	3-4
Methods			
Study design	4	Present key elements of study design early in the paper	4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5-6
Participants	6	(a) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> —Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants	5-6
		(b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5-6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-6
Bias	9	Describe any efforts to address potential sources of bias	5-6
Study size	10	Explain how the study size was arrived at	5-6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Not applicable
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding	4-6
		(b) Describe any methods used to examine subgroups and interactions	Not applicable
		(c) Explain how missing data were addressed	Not applicable
		(d) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed	Not applicable

		<i>Cross-sectional study</i> —If applicable, describe analytical methods taking account of sampling strategy	
		(e) Describe any sensitivity analyses	6
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Not applicable
		(b) Give reasons for non-participation at each stage	Not applicable
		(c) Consider use of a flow diagram	Not applicable
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	7-13
		(b) Indicate number of participants with missing data for each variable of interest	Not applicable
		(c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	Not applicable
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time	Not applicable
		<i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure	Not applicable
		<i>Cross-sectional study</i> —Report numbers of outcome events or summary measures	7-13
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	7-13
		(b) Report category boundaries when continuous variables were categorized	Not applicable
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not applicable
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	7-13
Discussion			
Key results	18	Summarise key results with reference to study objectives	13-15
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13-15
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	13-15
Generalisability	21	Discuss the generalisability (external validity) of the study results	14
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	16

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.