

RESEARCH

Open Access



Noncollapsibility and its role in quantifying confounding bias in logistic regression

Noah A. Schuster^{1*}, Jos W. R. Twisk¹, Gerben ter Riet^{2,3}, Martijn W. Heymans¹ and Judith J. M. Rijnhart¹

Abstract

Background: Confounding bias is a common concern in epidemiological research. Its presence is often determined by comparing exposure effects between univariable- and multivariable regression models, using an arbitrary threshold of a 10% difference to indicate confounding bias. However, many clinical researchers are not aware that the use of this change-in-estimate criterion may lead to wrong conclusions when applied to logistic regression coefficients. This is due to a statistical phenomenon called noncollapsibility, which manifests itself in logistic regression models. This paper aims to clarify the role of noncollapsibility in logistic regression and to provide guidance in determining the presence of confounding bias.

Methods: A Monte Carlo simulation study was designed to uncover patterns of confounding bias and noncollapsibility effects in logistic regression. An empirical data example was used to illustrate the inability of the change-in-estimate criterion to distinguish confounding bias from noncollapsibility effects.

Results: The simulation study showed that, depending on the sign and magnitude of the confounding bias and the noncollapsibility effect, the difference between the effect estimates from univariable- and multivariable regression models may underestimate or overestimate the magnitude of the confounding bias. Because of the noncollapsibility effect, multivariable regression analysis and inverse probability weighting provided different but valid estimates of the confounder-adjusted exposure effect. In our data example, confounding bias was underestimated by the change in estimate due to the presence of a noncollapsibility effect.

Conclusion: In logistic regression, the difference between the univariable- and multivariable effect estimate might not only reflect confounding bias but also a noncollapsibility effect. Ideally, the set of confounders is determined at the study design phase and based on subject matter knowledge. To quantify confounding bias, one could compare the unadjusted exposure effect estimate and the estimate from an inverse probability weighted model.

Keywords: Logistic regression, Confounding, Noncollapsibility, Confounder-adjustment, Univariable regression analysis, Multivariable regression analysis, Inverse probability weighting, Conditional effect, Marginal effect

*Correspondence: n.schuster@amsterdamumc.nl

¹ Department of Epidemiology and Data Science, Amsterdam Public Health Research Institute, Amsterdam UMC - Location VU University Medical Center, De Boelelaan 1117, Amsterdam, The Netherlands
Full list of author information is available at the end of the article



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

In observational studies, the exposure levels are often influenced by characteristics of the study subjects. As a result, differences in background characteristics between exposed and unexposed individuals may exist. If these characteristics are also associated with the outcome, crude comparison of the average outcomes in both exposure groups does not yield an unbiased estimate of the exposure effect [1–5]. Therefore, to obtain unbiased effects, adjustment for this imbalance in background characteristics is necessary. This is also called adjustment for confounding.

When selecting confounders for adjustment, researchers often use statistical methods to quantify the confounding bias. That is, oftentimes the confounding bias is quantified by comparing the exposure effect between a univariable- and a multivariable regression model, also called the change-in-estimate criterion [4, 6, 7]. However, this method may lead to wrong conclusions about the presence and magnitude of confounding bias, as in logistic regression covariates may affect the effect estimate through two separate mechanisms: through confounding when covariates are associated with both the exposure and the outcome, and through noncollapsibility which is present when covariates are associated with the outcome [8]. The total difference between the effect estimate from a univariable- and multivariable regression model may therefore be decomposed into an estimate of confounding bias and an estimate of the noncollapsibility effect [7, 9]. Furthermore, even in the absence of confounding the exposure effect coefficients from both models might still differ. Thus, the change-in-estimate may misrepresent the true confounding bias [4].

Various rescaling methods have been proposed in the social sciences literature, which aim to equalize the scales of the effect estimates from a univariable and a multivariable regression model [10–13]. However, when applied to effect estimates from a logistic regression, these rescaling measures are approximate rather than exact [10, 11, 14]. Janes et al. [9] and Pang et al. [7] proposed an exact measure of confounding bias for logistic regression models. This measure is based on the comparison of the effect estimates from a univariable regression model and an inverse probability weighted (IPW) model. The latter is another popular method to adjust for confounding.

Noncollapsibility may not only affect the differences between the effect estimates from a univariable- and multivariable regression model, it also causes differences between the effect estimates from a multivariable regression model and an IPW model. Whereas multivariable regression and IPW provide the same effect estimates in linear regression, this does not necessarily hold for logistic regression [7, 9, 15]. That is, when a noncollapsibility

effect is present, multivariable regression adjustment and IPW both yield valid estimates of the confounder-adjusted exposure effect, but their magnitude and interpretation differ [7, 16, 17]. Therefore, the difference between the effect estimates from a multivariable regression model and IPW can be used to quantify the magnitude of noncollapsibility.

Because noncollapsibility is a relatively unknown mechanism among clinical researchers, many are unaware that the change-in-estimate criterion may lead to wrong conclusions about the presence and magnitude of confounding bias. Therefore, this paper aims to clarify the role of noncollapsibility in logistic regression and to provide guidance in determining the presence of confounding bias. First, we review the different confounder-adjustment methods and provide a detailed explanation of the noncollapsibility effect. Then, we use a Monte Carlo simulation study to uncover patterns of confounding bias and noncollapsibility effects in logistic regression. Subsequently, using an empirical data example, we demonstrate that the change-in-estimate criterion to determine confounding bias may be misleading. Finally, we provide guidance in determining the set of confounders and quantifying confounding bias.

Confounder adjustment and noncollapsibility

The presence and magnitude of confounding bias for models with a binary outcome is commonly determined by comparing the exposure effect estimates from a univariable- (Eq. 1) and multivariable (Eq. 2) logistic regression model:

$$\text{logit}(\Pr(Y = 1|X)) = i_1 + \beta_1 X \quad (1)$$

$$\begin{aligned} \text{logit}(\Pr(Y = 1|X, C_1, \dots, C_n)) = & i_2 + \beta'_1 X + \beta'_2 C_1 \\ & + \dots + \beta'_{n+1} C_n \end{aligned} \quad (2)$$

where in both equations, Y and X represent the outcome and exposure variables and i_1 and i_2 represent the intercept terms, respectively. In Eq. 1, β_1 represents the *unadjusted* exposure effect estimate. In Eq. 2, β'_1 represents the multivariable confounder-adjusted exposure effect estimate and β'_2 to β'_{n+1} are the coefficients corresponding to observed background covariates C_1 to C_n . When C_1 to C_n are truly confounders, then β_1 will be a biased estimate of the causal exposure-outcome effect. Assuming that Eq. (2) contains all confounders of the exposure-outcome effect, β'_1 will have a causal interpretation. In practice researchers often determine the magnitude of confounding as the change in estimate, which is computed as the difference between β'_1 and β_1 . When using the change-in-estimate criterion to determine the

presence of confounding bias typically a 10% difference between β'_1 and β_1 is used in practice as an arbitrary threshold indicating confounding due to covariates C_1 to C_n in the association between X and Y [6, 18, 19].

When based on logistic regression, $\beta'_1 - \beta_1$ may not only represent confounding bias but also a noncollapsibility effect. This noncollapsibility effect is sometimes also referred to as a form of the Simpson's paradox [16]. The noncollapsibility effect is caused by a difference in the scale on which β_1 and β'_1 are estimated. In linear regression, the total variance is the same for nested models: when the explained variance increases through adding a covariate to the model, the unexplained variance decreases by the same amount. As a result, effect estimates from nested linear models are on the same scale and thus *collapsible*. In logistic regression, however, the unexplained variance has a fixed value of 3.29 [8]. Adding covariates that are associated with the outcome (e.g., confounders) increases the explained variance and forces the total variance of Y to increase. When the total variance of Y increases, the scale of the estimated coefficients changes, causing negative exposure effects to become more negative and positive exposure effects more positive. This change in scales is called the *noncollapsibility* effect [5, 7, 8]. Thus, to determine confounding bias, exposure effect estimates cannot be simply compared between nested logistic regression models as the difference might not only reflect confounding bias but also a noncollapsibility effect [8]. The noncollapsibility effect also occurs when a covariate is associated with outcome Y but not with exposure X (i.e., when the covariate is not a confounder). The change in estimate then represents the noncollapsibility effect only, falsely indicating the presence of confounding bias. To preserve space in the main text, a hypothetical example illustrating how the change in estimate might be affected by the noncollapsibility effect in the absence of confounding is given in additional file A. An explanation of noncollapsibility based on a contingency table is provided by for example Pang et al. [7].

Recent studies by Janes et al. and Pang et al. presented an exact estimate of confounding bias unaffected by noncollapsibility based on logistic regression [7, 9], using the difference between the univariable exposure effect estimate and the effect estimate from an IPW model. With IPW, confounding bias is eliminated by creating a pseudo-population in which each covariate combination is balanced between both exposure groups [20–22]. When there is perfect covariate balance there is no longer an association between covariates C_1 to C_n and exposure status X . This pseudo-population can be created by weighting subjects so that for each combination of baseline covariates the sums of contributions for both

exposure groups are equal [1, 20]. These weights are the inverse of the probability that a subject was exposed, i.e. the inverse of a propensity score [23].

The propensity score is the predicted probability of endorsing the exposure, which can be estimated using Eq. 3:

$$PS = Pr(X = 1|C) = \frac{1}{1 + e^{-(i_3 + \lambda_1 C_1 + \dots + \lambda_n C_n)}} \quad (3)$$

where X represents exposure, i_3 is the model intercept and λ_1 to λ_n are regression coefficients corresponding to covariates C_1 to C_n . The propensity score methodology can also be extended to continuous exposure variables using the Generalized Propensity Score (GPS), which has a similar balancing property to the classic propensity score. For more information on how to perform propensity score analysis with a continuous exposure variable, see Hirano (2004) and Imai (2004) [24, 25].

For exposed subjects, the weight is calculated as $\frac{1}{PS}$ and for unexposed subjects as $\frac{1}{1-PS}$ [1, 20, 22]. Using these calculations, subjects with a propensity score close to 0 end up with large weights, and subjects with a propensity score close to 1 end up with small weights. Because in some situations these weights cause the IPW model to be unstable, stabilized weights have been proposed [26]. For exposed subjects, the stabilized weight is calculated as $\frac{p}{PS}$ and for unexposed subjects as $\frac{1-p}{1-PS}$, where p is the probability of exposure without considering covariates C_1 to C_n [2, 26]. Subsequently, a weighted regression analysis with exposure X as the only independent variable is carried out. We call the confounder-adjusted exposure effect estimate from the IPW model β_1^* .

Difference between IPW- and multivariable confounder-adjusted exposure effect estimates

Multivariable regression adjustment and IPW provide identical exposure effect estimates when based on linear regression, but not when based on logistic regression [15, 27]. The difference between the IPW confounder-adjusted exposure effect estimate β_1^* and the multivariable confounder-adjusted exposure effect estimate β'_1 is caused by noncollapsibility, and the difference between the unadjusted exposure effect estimate β_1 and β_1^* provides a measure of confounding bias [7, 9, 14]. This is because in an IPW model the total variance remains equal to the total variance of the unadjusted model, while in a multivariable regression model the addition of variables to the model leads to higher variance, changing the scale of the exposure effect estimate. This means that when there is confounding in a logistic regression model, multivariable regression analysis and IPW lead to different confounder-adjusted estimates of the exposure effect.

Although β'_1 and β_1^* are both valid estimates, they apply to different target populations and have their own respective interpretation [8, 27].

Simulation study

Simulation methods

A Monte Carlo simulation study was designed to investigate patterns of confounding bias and noncollapsibility effects in logistic regression. The R programming language version 4.0.2 [28] and STATA statistical software release 14 [29] were used to generate and analyze the data, respectively.

Three continuous covariates were generated from a standard normal distribution. The dichotomous exposure and outcome were generated from a binomial distribution conditional on the covariates and the covariates and exposure, respectively. Sample sizes were 250, 500, 750 and 1000. The parameter values for the exposure-outcome effect, confounder-exposure effect and the confounder-outcome effect were set to -1.42, -0.92, -0.38, 0, 0.38, 0.92 and 1.42. This way, the conditions reflected situations with combinations of zero effects, and positive and negative small (-0.38 and 0.38), medium (-0.92 and 0.92) and large (-1.42 and 1.42) effect sizes were mimicked [30]. The total number of conditions was 1,372 with 1,000 repetitions per condition, resulting in 1,372,000 observations. Subsequently, we estimated the unadjusted exposure effect estimate β_1 , the multivariable confounder-adjusted exposure effect estimate β'_1 and the IPW confounder-adjusted exposure effect estimate β_1^*

based on the simulated data. From these effect estimates we computed the change in estimate, the confounding bias and the noncollapsibility effect. The simulation code is available in additional file B.

Simulation scenarios

We expected to observe four scenarios based on the simulated data. In the first scenario (Fig. 1a), the covariates are associated with both the exposure and the outcome. In this scenario there will be both confounding bias ($\beta_1 - \beta_1^*$) and a noncollapsibility effect ($\beta_1^* - \beta'_1$). Because the exposure-outcome effect is simulated to be positive and negative, we also expect to see positive and negative noncollapsibility effect estimates. This means that $\beta'_1 - \beta_1$ might result in an under- or overestimation of the true confounding effect [8]. In the second scenario (Fig. 1b) the covariates are associated with both the exposure and outcome, but exposure and outcome are not associated with each other. In this scenario, any differences between β'_1 and β_1 are fully explained by the covariates, so there is confounding bias without a noncollapsibility effect [8, 15]. In the third scenario (Fig. 1c), the covariates are only associated with the outcome. In this scenario there is a noncollapsibility effect but no confounding bias. In real-life situations with this structure, using the change-in-estimate criterion may lead one to conclude that the covariates are confounders in the relation between the exposure and the outcome although the difference between β'_1 and β_1 is caused entirely by the noncollapsibility effect [7, 8]. In the fourth scenario (Fig. 1d), the

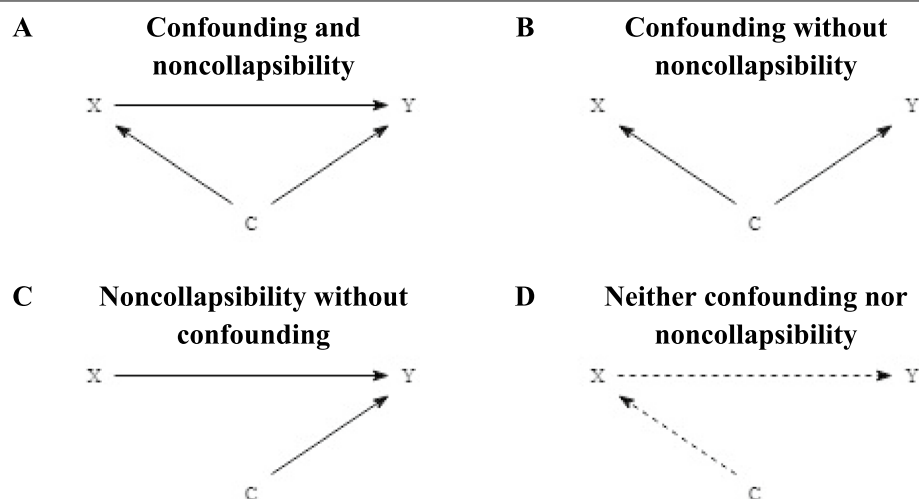


Fig. 1 Directed acyclic graphs of the four possible scenarios into which each simulated condition can be classified. Panel **A**: both confounding and noncollapsibility. Panel **B**: confounding without noncollapsibility. Panel **C**: noncollapsibility without confounding. Panel **D**: neither confounding nor noncollapsibility. C represents three continuous covariates, X represents the dichotomous exposure and Y represents the dichotomous outcome. The dotted line in panel D between the covariates and the exposure and between the exposure and the outcome indicate there may or may not be an association

covariates may be associated with the exposure, but not with the outcome. In this scenario, there is neither confounding bias nor a noncollapsibility effect and β_1, β'_1 and β_1^* are identical. This scenario is also called *strict* collapsibility [15, 31, 32].

Simulation results

The difference between β'_1 and β_1 can be negative, zero or positive, depending on the magnitude of the confounding bias and the noncollapsibility effect (Table 1). Only when there was no noncollapsibility effect (i.e., $\beta_1^* - \beta'_1 = 0$), the change in estimate equaled the estimate of confounding bias. The noncollapsibility effect was zero when the exposure-outcome effect was zero and the confounder-exposure and confounder-outcome effects were both non-zero. When the exposure-outcome effect was also non-zero, the difference between β'_1 and β_1 reflected both confounding bias and the noncollapsibility effect. In those situations, the change-in-estimate criterion could both under- and overestimate the true confounding bias. When the confounding bias and noncollapsibility effect had similar signs, i.e. both were positive or negative, $\beta'_1 - \beta_1$ overestimated the true confounding bias. When the confounding bias and noncollapsibility effect had opposite signs, i.e. one was positive while the other was negative, the true confounding bias could be under- or overestimated by $\beta'_1 - \beta_1$, depending on the magnitude of the confounding bias and noncollapsibility effect.

Thus, when the exposure-outcome effect is non-zero, the change-in-estimate criterion might falsely indicate the presence of confounding or it might under- or overestimate the true confounding bias. Patterns of confounding bias and the noncollapsibility effect were similar across sample sizes and will be described below.

Confounding bias

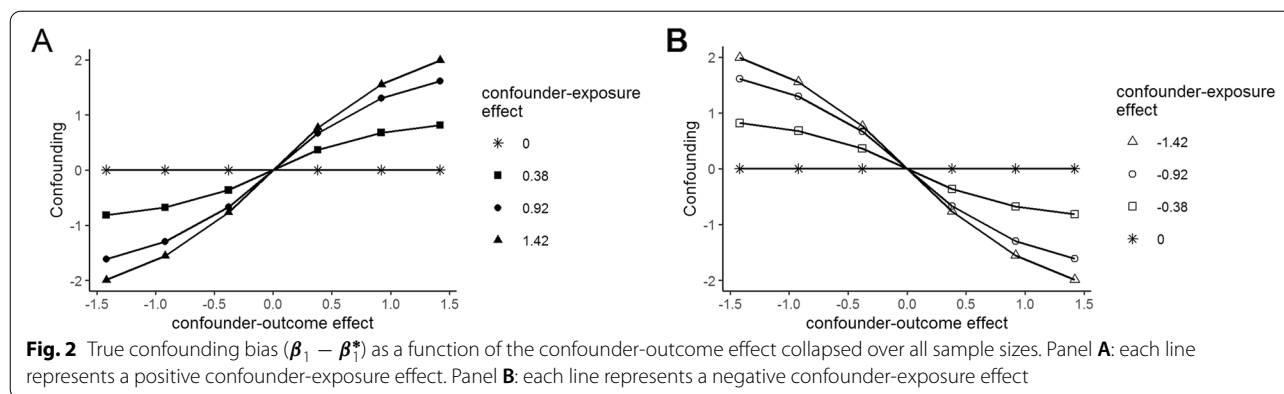
Figure 2 plots confounding bias ($\beta_1 - \beta_1^*$) as a function of the confounder-outcome effect with the lines in panel A representing positive confounder-outcome effects of various magnitudes and the lines in panel B representing negative confounder-outcome effects of various magnitudes. Confounding bias was positive when the confounder-exposure effect and the confounder-outcome effect were both positive (panel A, first quadrant) and when they were both negative (panel B, second quadrant). When the effects had opposite signs, confounding bias was negative. The magnitude of confounding bias increased as the confounder-exposure or confounder-outcome effect increased in magnitude. There was no confounding bias when one or both effects equaled zero.

The noncollapsibility effect

Figure 3 plots the noncollapsibility effect ($\beta_1^* - \beta'_1$) as a function of the confounder-outcome effect with the lines in panel A representing positive exposure-outcome

Table 1 Difference between univariable- and multivariable exposure effects as combination of confounding bias and the noncollapsibility effect

Difference between multivariable- and univariable effect estimate ($\beta'_1 - \beta_1$)	Confounding bias ($\beta_1 - \beta_1^*$)	Noncollapsibility effect ($\beta_1^* - \beta'_1$)
Negative	Negative value	Negative value
	Zero	Negative value
	Negative value	Zero
	Positive value	Greater negative value than the positive confounding bias value
	Greater negative value than the positive noncollapsibility effect value	Positive value
Zero	Zero	Zero
	Equal positive value as the negative noncollapsibility effect value	Equal negative value as the positive confounding bias value
	Equal negative value as the positive noncollapsibility effect value	Equal positive value as the negative confounding bias value
Positive	Positive value	Positive value
	Zero	Positive value
	Positive value	Zero
	Negative value	Greater positive value than the negative confounding bias value
	Greater positive value than the negative noncollapsibility effect value	Negative value



effects of various magnitudes and the lines in panel B representing negative exposure-outcome effects of various magnitudes. The noncollapsibility effect and the exposure-outcome effect were inversely related: when the latter effect was positive, the noncollapsibility effect was negative, and vice versa. The noncollapsibility effect increased in magnitude as both the exposure-outcome effect and the confounder-outcome effect increased in magnitude. When either effect was zero, there was no noncollapsibility effect, regardless of the magnitude of the other effect.

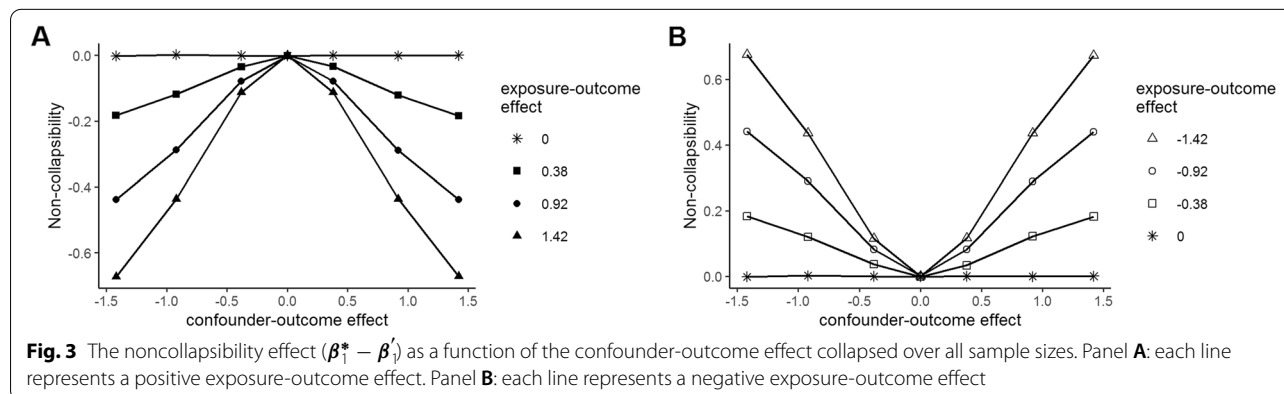
Empirical data example

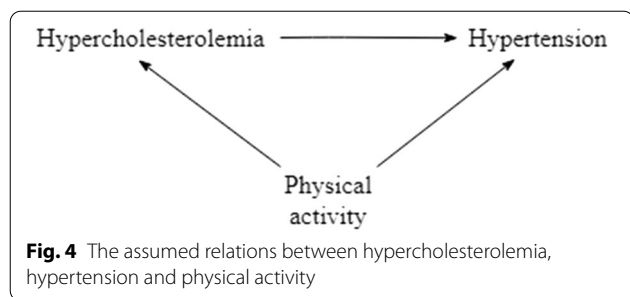
To illustrate how the noncollapsibility effect might affect conclusions about confounding bias in practice we use an example from the Amsterdam Growth and Health Longitudinal Study (AGHLS). The AGHLS started in 1976 with the aim to examine growth and health among teenagers. Over the years, health and lifestyle measures, determinants of chronic diseases and parameters for the investigation of deterioration in health with age have been measured [33]. The data in this example were collected in 2000, when the participants were in their late 30s. Using data from the AGHLS we investigated the association between hypercholesterolemia and hypertension, potentially confounded by physical activity. Using

multivariable regression analysis and IPW we estimated the confounder-adjusted effect of hypercholesterolemia on hypertension in our sample, β_1' and β_1^* , respectively. To quantify the magnitude of confounding bias and the noncollapsibility effect, we also estimated the unadjusted exposure effect β_1 using univariable regression analysis. Cut-offs for hypercholesterolemia and hypertension were based on guidelines from the U.S. National Institutes of Health (NIH) and NIH's National Heart, Lung and Blood Institute, respectively [34, 35]. Physical activity was defined as the total hours per week spent on light, moderate or vigorous activities. Only subjects with complete data on the variables were considered in the analysis ($n = 349$). Figure 4 provides a graphical representation of the assumed relations among the variables.

Table 2 shows the effect estimates from univariable- and multivariable regression analysis and IPW. The unadjusted effect estimate β_1 was 0.90, corresponding to an odds ratio (OR) of 2.46. The multivariable confounder-adjusted exposure effect estimate β_1' was 0.93, corresponding to an OR of 2.53. The IPW confounder-adjusted exposure effect estimate β_1^* was 0.99, corresponding to an OR of 2.69.

The difference between β_1' and β_1 was -0.03, or 3.3%. If one would use the change-in-estimate criterion with a cut-off of 10% to determine the presence of confounding,





then physical activity would not be considered a confounder. Using the difference between β_1 and β_1^* , the estimate of confounding bias was $0.90 - 0.99 = -0.09$. This corresponds to a 10% change in the exposure effect estimate. The noncollapsibility effect estimate was $0.99 - 0.93 = 0.06$. Because of this noncollapsibility effect, the estimate of the true confounding bias of physical activity was considerably larger than it seemed based on the difference between β_1' and β_1 . Thus, in our data example, the conventional method to determine the presence of confounding led to an underestimation of the true confounding bias of physical activity.

Discussion

This paper aimed to clarify the role of noncollapsibility in determining the magnitude of confounding bias in logistic regression. Because the difference between β_1' and β_1 reflects both confounding bias and a noncollapsibility effect, in logistic regression the change-in-estimate criterion should not be used to determine the presence of confounding. This was illustrated in our data example, in which confounding bias was underestimated because of the magnitude of the noncollapsibility effect. Our simulation study showed that confounding was mainly determined by the combination of the magnitude of the confounder-exposure and confounder-outcome effects, whereas noncollapsibility was mostly determined by the

magnitude of the combination of the exposure-outcome and confounder-outcome effects. In situations in which confounding approached zero and noncollapsibility was non-zero, the change-in-estimate criterion wrongly indicated the presence of confounding bias, when in reality the difference between β_1' and β_1 was caused solely by the noncollapsibility effect.

Recommendations for practice

Rather than using an arbitrary statistical rule such as the 10% cut-off based on the change-in-estimate criterion, it is generally recommended to determine the confounder set based on subject matter knowledge. Directed acyclic graphs (DAGs) are helpful to determine which set of confounders should be adjusted for to eliminate confounding bias [36, 37]. DAGs are causal diagrams in which the arrows represent the causal relations among variables. Therefore, DAGs contain information about the causal model that cannot be provided by statistical methods. For example, assuming the DAG is a correct representation of the causal relations among variables, it clarifies what the minimally sufficient set of confounders is to block any backdoor paths (i.e., confounding paths) from the exposure to the outcome. The amount of confounding bias could be quantified by looking at the difference between the unadjusted univariable exposure effect estimate β_1 and the IPW confounder-adjusted exposure effect estimate β_1^* as proposed by Pang et al. [7] and Janes et al. [9]. Bootstrap confidence intervals can be used to determine the statistical significance of the confounding bias.

Because of the noncollapsibility effect, multivariable regression analysis and IPW provide different estimates of the exposure effect. Multivariable regression analysis results in a *conditional* exposure effect estimate [16, 38], whereas IPW results in a *population-average* or *marginal* exposure effect estimate [16, 38–40]. Marginal exposure effects can also be estimated with standardization using G-computation. A step-by-step demonstration of G-computation can be found elsewhere [41]. It is often suggested that a population-average effect estimate should be reported when the target population is the entire study population, while the conditional exposure effect should be reported if the target population is a subset of the study population [7, 8, 16, 38, 39, 42, 43]. Although this distinction is known from the literature, when it comes to the practical application, the exact differences between the two exposure effect estimates and their respective interpretations remain unclear.

In this study, we assume correct specification of both the confounder-exposure and the confounder-outcome effect. When these are not correctly specified, bias might be introduced and the difference between the

Table 2 Relationship between hypercholesterolemia and hypertension estimated using univariable- and multivariable regression analysis and IPW

	β	SE(β)	95%CI	P
Univariable exposure effect				
Hypercholesterolemia	0.90	0.23	0.47; 1.35	<0.01
Multivariable confounder-adjusted exposure effect				
Hypercholesterolemia	0.93	0.23	0.48; 1.38	<0.01
Physical activity	0.01	0.01	-0.02; 0.03	0.60
IPW confounder-adjusted exposure effect				
Hypercholesterolemia	0.99	0.16	0.69; 1.30	<0.01

Abbreviations: SE: standard error; CI: confidence interval

unadjusted univariable exposure effect estimate and the IPW confounder-adjusted exposure effect estimate might not only reflect confounding bias but also the misspecification of the underlying models. Therefore, correct specification of all effects is necessary to estimate unbiased exposure effects and correctly quantify confounding bias.

Conclusion

To summarize, in this study we showed that in logistic regression the difference between univariable- and multivariable effect estimates may reflect both confounding bias and a noncollapsibility effect. To avoid wrong conclusions with respect to the magnitude and presence of confounding bias, confounders are ideally determined based on subject matter knowledge. To quantify confounding bias, one could look at the difference between the unadjusted univariable exposure effect estimate and the IPW confounder-adjusted exposure effect estimate.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12874-021-01316-8>.

Additional file 1. Hypothetical data example to illustrate the noncollapsibility effect.

Additional file 2. Code used to generate data for the simulation study.

Acknowledgements

Not applicable

Authors' contributions

JR designed the study. NS performed the statistical analyses and drafted the manuscript. JR, NS, GtR, MH and JT contributed to data interpretation, critically revised the manuscript and approved the final version of the manuscript.

Authors' information

Not applicable.

Funding

This work was supported by the Amsterdam Public Health Research Institute.

Availability of data and materials

1. The dataset used and analyzed in the empirical data example is available from the corresponding author on reasonable request.
2. The code for the simulated dataset is included in this study in additional file B.
3. The detailed results of the simulation study are available on request from the corresponding author.

Declarations

Ethics approval and consent to participate

Ethics approval for the Amsterdam Growth and Health Longitudinal Study was given by the VU Medical Center ethical committee at each follow-up round.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Epidemiology and Data Science, Amsterdam Public Health Research Institute, Amsterdam UMC - Location VU University Medical Center, De Boelelaan 1117, Amsterdam, The Netherlands. ²Center of Expertise Urban Vitality, Faculty of Health, Amsterdam University of Applied Sciences, Amsterdam, The Netherlands. ³Department of Cardiology, Amsterdam Public Health Research Institute, Amsterdam UMC - Location AMC, Meibergdreef 9, Amsterdam, The Netherlands.

Received: 5 March 2021 Accepted: 7 May 2021

Published online: 05 July 2021

References

1. Austin Peter C. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivar Behav Res.* 2011;46(3):399–424. <https://doi.org/10.1080/00273171.2011.568786>.
2. Hernan MA, Robins JM. *Causal inference: what If.* Boca Raton: Chapman & Hall/CRC; 2020.
3. Samuels ML. Matching and design efficiency in epidemiological studies. *Biometrika.* 1981;68(3):577–88. <https://doi.org/10.1093/biomet/68.3.577>.
4. Miettinen OS, Cook EF. Confounding: essence and detection. *Am J Epidemiol.* 1981;114(4):593–603. <https://doi.org/10.1093/oxfordjournals.aje.a113225>.
5. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol.* 1986;15(3):413–9. <https://doi.org/10.1093/ije/15.3.413>.
6. Kleinbaum DG, Sullivan KM, Barker ND. *A pocket guide to epidemiology.* Springer Science + Business Media, LLC; 2007.
7. Pang M, Kaufman JS, Platt RW. Studying noncollapsibility of the odds ratio with marginal structural and logistic regression models. *Stat Methods Med Res.* 2016;25(5):1925–37. <https://doi.org/10.1177/0962280213505804>.
8. Mood C. Logistic regression: why we cannot do what we think we can do, and what we can do about it. *Eur Sociol Rev.* 2009;26(1):67–82. <https://doi.org/10.1093/esr/jcp006>.
9. Janes H, Dominici F, Zeger S. On quantifying the magnitude of confounding. *Biostatistics.* 2010;11(3):572–82. <https://doi.org/10.1093/biostatistics/kxq007>.
10. Cramer JS. Robustness of logit analysis: unobserved heterogeneity and mis-specified disturbances. *Oxford Bull Econ Stat.* 2007;69(4):545–55. <https://doi.org/10.1111/j.1468-0084.2007.00445.x>.
11. MacKinnon DP, Luecken LJ. How and for whom? Mediation and moderation in health psychology. *American Psychological Association*; 2008. p. S99-S100.
12. Karlson KB, Holm A, Breen R. Comparing regression coefficients between same-sample nested models using logit and probit: a new method. *Sociol Methodol.* 2012;42(1):286–313. <https://doi.org/10.1177/0081175012444861>.
13. Breen R, Karlson KB, Holm A. Total, direct, and indirect effects in logit and probit models. *Sociol Methods Res.* 2013;42(2):164–91. <https://doi.org/10.1177/0049124113494572>.
14. Rijnhart JJM, Valente MJ, MacKinnon DP. Total effect decomposition in mediation analysis in the presence of non-collapsibility. Submitted for publication.
15. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Stat Sci.* 1999;14(1):29–46.
16. Hernan MA, Clayton D, Keiding N. The Simpson's paradox unraveled. *Int J Epidemiol.* 2011;40(3):780–5. <https://doi.org/10.1093/ije/dyr041>.
17. Neuhaus JM, Jewell NP. A geometric approach to assess bias due to omitted covariates in generalized linear models. *Biometrika.* 1993;80(4):807–15. <https://doi.org/10.1093/biomet/80.4.807>.
18. Lee PH. Is a cutoff of 10% appropriate for the change-in-estimate criterion of confounder identification? *J Epidemiol.* 2014;24(2):161–7. <https://doi.org/10.2188/jea.JE20130062>.
19. Budtz-Jorgensen E, Keiding N, Grandjean P, Weihe P. Confounder selection in environmental epidemiology: assessment of health effects of

- prenatal mercury exposure. *Ann Epidemiol*. 2007;17(1):27–35. <https://doi.org/10.1016/j.annepidem.2006.05.007>.
20. Heinze G, Juni P. An overview of the objectives of and the approaches to propensity score analyses. *Eur Heart J*. 2011;32(14):1704–8. <https://doi.org/10.1093/eurheartj/ehr031>.
 21. Brookhart MA, Wyss R, Layton JB, Sturmer T. Propensity score methods for confounding control in nonexperimental research. *Circ Cardiovasc Qual Outcomes*. 2013;6(5):604–11. <https://doi.org/10.1161/CIRCOUTCOMES.113.000359>.
 22. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550–60. <https://doi.org/10.1097/00001648-200009000-00011>.
 23. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41–55. <https://doi.org/10.1093/biomet/70.1.41>.
 24. Hirano K, Imbens GW. The propensity score with continuous treatments. In: Gelmanand A, Meng X-L, editors. *Applied Bayesian modeling and causal inference from incomplete-data perspectives*. John Wiley & Sons, Ltd; 2004. p. 73–84.
 25. Imai K, van Dyk DA. Causal inference with general treatment Regimes. *J Am Stat Assoc*. 2004;99(467):854–66. <https://doi.org/10.1198/01621450400001187>.
 26. Xu S, Ross C, Raebel MA, Shetterly S, Blanchette C, Smith D. Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. *Value Health*. 2010;13(2):273–7. <https://doi.org/10.1111/j.1524-4733.2009.00671.x>.
 27. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Systematic differences in treatment effect estimates between propensity score methods and logistic regression. *Int J Epidemiol*. 2008;37(5):1142–7. <https://doi.org/10.1093/ije/dyn079>.
 28. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.
 29. StataCorp. Stata statistical software: release 14. College Station, TX: Stata-Corp LP; 2015.
 30. Chen H, Cohen P, Chen S. How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Comm Stat Simul Comput*. 2010;39(4):860–4. <https://doi.org/10.1080/03610911003650383>.
 31. Mansournia MA, Greenland S. The relation of collapsibility and confounding to faithfulness and stability. *Epidemiology*. 2015;26:466–72. <https://doi.org/10.1097/EDE.0000000000000291>.
 32. Whittemore AS. Collapsibility of multidimensional contingency tables. *Royal Statistical Society*. 1978;40(3):328–40.
 33. Wijnstok NJ, Hoekstra T, van Mechelen W, Kemper HC, Twisk JW. Cohort profile: the Amsterdam Growth and Health Longitudinal Study. *Int J Epidemiol*. 2013;42(2):422–9. <https://doi.org/10.1093/ije/dys028>.
 34. National Institutes of Health's U.S. National Library of Medicine. Cholesterol levels: what do the results mean n.d. [Available from: <https://medlineplus.gov/lab-tests/cholesterol-levels/>].
 35. NIH: National Heart L, and Blood Institute. High blood pressure: confirming high blood pressure 2020 [Available from: <https://www.nhlbi.nih.gov/health-topics/high-blood-pressure>].
 36. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. *BMC Med Res Methodol*. 2008;8:70. <https://doi.org/10.1186/1471-2288-8-70>.
 37. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10(1):37–48. <https://doi.org/10.1097/00001648-199901000-00008>.
 38. Daniel Rhian, Zhang Jingjing, Farewell Daniel. Making apples from oranges: comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. *Biomet J*. 2020;63(3):528–57. <https://doi.org/10.1002/bimj.201900297>.
 39. Breen Richard, Karlson Kristian Bernt, Holm Anders. Interpreting and understanding logits, probits, and other nonlinear probability models. *Annu Rev Sociol*. 2018;44(1):39–54. <https://doi.org/10.1146/annurev-soc-073117-041429>.
 40. Burgess Stephen. Estimating and contextualizing the attenuation of odds ratios due to non collapsibility. *Communications in Statistics - Theory and Methods*. 2016;46(2):786–804. <https://doi.org/10.1080/03610926.2015.1006778>.
 41. Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. *Am J Epidemiol*. 2011;173(7):731–8. <https://doi.org/10.1093/aje/kwq472>.
 42. Zhang Z. Estimating a marginal causal odds ratio subject to confounding. *Communications in Statistics - Theory and Methods*. 2008;38(3):309–21. <https://doi.org/10.1080/03610920802200076>.
 43. Karlson KB, Popham F, Holm A. marginal and conditional confounding using logits. *Sociol Methods Re*. 2021. <https://doi.org/10.1177/0049124121995548>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

