Original Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–13

DOI: 10.1177/0962280213505804
smm.sagepub.com
SAGE

Studying noncollapsibility of the odds ratio with marginal structural and logistic regression models

Menglan Pang, Jay S Kaufman and Robert W Platt

Abstract

One approach to quantifying the magnitude of confounding in observational studies is to compare estimates with and without adjustment for a covariate, but this strategy is known to be defective for noncollapsible measures such as the odds ratio. Comparing estimates from marginal structural and standard logistic regression models, the total difference between crude and conditional effects can be decomposed into the sum of a noncollapsibility effect and confounding bias. We provide an analytic approach to assess the noncollapsibility effect in a point-exposure study and provide a general formula for expressing the noncollapsibility effect. Next, we provide a graphical approach that illustrates the relationship between the noncollapsibility effect and the baseline risk, and reveals the behavior of the noncollapsibility effect for a range of different exposure and covariate effects. Various observations about noncollapsibility can be made from the different scenarios with or without confounding; for example, the magnitude of effect of the covariate plays a more important role in the noncollapsibility effect than does that of the effect of the exposure. In order to explore the noncollapsibility effect of the odds ratio in the presence of time-varying confounding, we simulated an observational cohort study. The magnitude of noncollapsibility was generally comparable to the effect in the point-exposure study in our simulation settings. Finally, in an applied example we demonstrate that collapsibility can have an important impact on estimation in practice.

Keywords

noncollapsibility, odds ratio, marginal structural model, logistic regression model, confounding bias

Consider a $2 \times 2 \times K$ contingency table with binary exposure (A), binary outcome (Y), and a third categorical variable (L). A measure of the exposure–outcome association is said to be collapsible if the value obtained from the marginal table summed over the K strata of L can be expressed as a weighted average of the K stratum-specific values.[1–5] The measure is said to be strictly collapsible if the marginal value and the stratum-specific values are all equal. In regression contexts, the measure is strictly

Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada

Corresponding author:
Jay S Kaufman, Department of Epidemiology, Biostatistics and Occupational Health, McGill University, 1020 Pine Avenue West, Montreal, QC, H3A 1A2 Canada.
Email: jay.kaufman@mcgill.ca

collapsible in a generalized linear model if the estimate remains constant when L is included and omitted.[6]

Collapsibility is a property of the specific effect measure, and many examples have been used to show that in a specific dataset, the risk difference (RD), risk ratio (RR), and odds ratio (OR) may have different collapsibility properties.[4–8] The conditions that are sufficient to guarantee collapsibility have been described previously. The RD and RR are collapsible when (i) L does not vary, or (ii) $A \coprod Y|\mathrm{L}$ (i.e. A and Y are statistically independent conditional on L), or (iii) $L \coprod Y|A$, or (iv) $L \coprod Y$ (the distributions of L are the same in the exposed and unexposed population), where $\coprod$ denotes statistical independence. On the other hand, the OR is collapsible under (i), or (ii), or (iii), or (v) $L \coprod A|Y$ but not under (iv).[5,9–14]

Noncollapsibility has often been a source of confusion in the literature when confounding is defined based on examination of the observed data alone (e.g. using a change in estimate criterion).[15,16] For example, noncollapsibility has been treated as confounding bias incorrectly when comparing the OR with or without adjusting for a baseline covariate.[17,18] Likewise, a covariate has been defined to be a nonconfounder if the marginal OR equals the conditional OR.[19] On the other hand, one standard definition of confounding is given as: "…L is a confounder if and only if both of the following conditions are satisfied: (a) L is a cause of disease in the unexposed cohort, and (b) L is associated with exposure in the cohort at the start of follow-up."[20] Based on this definition, randomization is considered to be a gold standard for control of confounding. Successful randomization ensures exchangeability of the exposed and unexposed populations, so that confounding is absent in expectation (i.e. condition (iv) earlier). When the RR or RD is used as the effect measure, therefore, conditions for nonconfounding and collapsibility are equivalent. However when measures such as the OR are used, confounding and noncollapsibility become distinct phenomena (since collapsibility is a consequence of condition (v) earlier, rather than condition (iv)). For example, conditional and crude ORs can be different even in large randomized experiments and there could be confounding bias according to the above definition even when the conditional and crude ORs are equal.[4] The distinction between confounding and noncollapsibility is still a source of great misunderstanding in applied data analysis.

Noncollapsibility of the OR derives from the fact that when the expected probability of outcome is modeled as a nonlinear function of the exposure, the marginal effect cannot be expressed as a weighted average of the conditional effects.[7,21] In the absence of confounding or when confounding is adjusted appropriately, both the marginal OR and conditional OR are valid measures. They are unbiased estimators for two different parameters, and the choice of reporting the marginal or conditional OR should depend on the research question. One should report the marginal OR if the average effect at the population level is of interest, while one should report the conditional OR if the conditional effect at the individual or subgroup level is of interest. Jensen's inequality provides theoretical justification for this noncollapsibility in the absence of confounding, requiring that the marginal OR is always shifted toward the null compared to the conditional OR.[22] Yet an explicit measure of the magnitude of noncollapsibility has not been proposed and would be useful in practice.

In this paper, we propose an approach to quantify the noncollapsibility effect and the confounding bias separately in a point-exposure study and investigate noncollapsibility in different scenarios. We also extend our study to a scenario with time-varying confounding, using simulations. We compare our results from the time-varying setting to a point-exposure study with similar characteristics. Finally, we demonstrate that collapsibility and confounding may both be important under reasonable settings, using a simple applied example.

1 Notation and definitions

In order to distinguish noncollapsibility of the OR from confounding, we first introduce the concept of the noncollapsibility effect. We define the noncollapsibility effect as the discrepancy between the marginal OR and the conditional OR after accounting completely for any confounding bias. Although this is not a true causal effect in the sense of a counterfactual contrast, we will nonetheless use this terminology to refer to the disparity between two different effect estimators.

1.1 Point-exposure study

Figure 1 presents casual diagrams for point-exposure studies with exposure A, outcome Y, and covariate L. Note L is a baseline covariate in the left panel of Figure 1, while L is a confounding variable in the right panel.

Table 1 shows the corresponding contingency table in a point-exposure study with all individuals recruited from the entire population (i.e. without sampling variation). We focus on the setting where all the variables are dichotomous with levels 0 and 1.

The conditional OR (OR_C) can be estimated with a standard logistic regression model (SLRM), with OR_C equal to e^{α} estimated from $g[E(Y|A,L)] = \mu + \alpha A + \beta L$, where g is the logit link function. The marginal OR (OR_m) can be estimated with a marginal structural model (MSM), corresponding to $e^{\alpha'}$ estimated from $g[E(Y|A)] = \mu' + \alpha' A$, estimated by weighting each subject by the inverse of the probability of receiving his/her exposure conditional on L. Finally, the crude OR (OR_{crude}) can be estimated with a SLRM, corresponding to e^{α^*} estimated from $g[E(Y|A)] = \mu^* + \alpha^* A$.

Since OR_C and OR_m are both asymptotically unbiased estimates of their respective parameters, the confounding bias is therefore measured by comparing the two marginal effects OR_{crude} and OR_m. The total discrepancy between OR_{crude} and OR_C in log scale can be consequently decomposed into two components, namely the noncollapsibility effect and the confounding bias. Following Janes et al.,[23] the decomposition can be written as

$$\log(OR_{crude}) - \log(OR_C) = [\log(OR_m) - \log(OR_C)] + [\log(OR_{crude}) - \log(OR_m)] \quad (1)$$

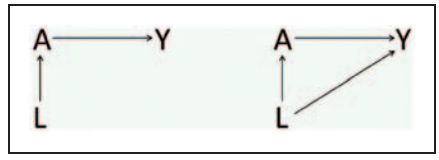

Figure 1. Causal diagrams for the point-exposure studies.

Table 1. Data from the point-exposure study.

L = 0			L = 1		
	A = 0	A = 1		A = 0	A = 1
Y = 0	a	b	Y = 0	e	f
Y = 1	c	d	Y = 1	g	h

so that the noncollapsibility effect can be measured by $\log(OR_m) - \log(OR_C)$, and the confounding bias is measured by $\log(OR_{crude}) - \log(OR_m)$. We propose two different approaches to study the noncollapsibility effect in the point-exposure study: an analytical approach and a graphical approach. The analytical approach is developed in order to derive the true value of the noncollapsibility effect as a function of related parameters, while the graphical approach is used to illustrate how the noncollapsibility effect is associated with these parameters.

2 An Analytical approach

We assume a deterministic process based on an infinite sample size. Probability therefore refers to proportions in the source population, and we use the notation and terminology of probability for convenience. The true OR and the estimated OR are therefore identical.

The outcome Y_i for each individual i is a random variable following the Bernoulli distribution with parameter $P = P(Y_i = 1|A = a_i, L = l_i)$. The probabilities of outcome conditional on A and L, the probability of A conditional on L, and the prevalence of L are the related parameters of interest. The effect measures OR_m, OR_{crude}, OR_C can all be computed analytically from these parameters.

Denote the conditional outcome probabilities by $P_{00} = P(Y = 1|A = 0, L = 0)$, $P_{10} = P(Y = 1|A = 1, L = 0)$, $P_{01} = P(Y = 1|A = 0, L = 1)$, and $P_{11} = P(Y = 1|A = 1, L = 1)$. Denote the conditional exposure probabilities by $P_{A0} = P(A = 1|L = 0)$ and $P_{A1} = P(A = 1|L = 1)$, and denote the prevalence of L by P_L. Finally, denote $q_{00} = 1 - P_{00}$, $q_{A0} = 1 - P_{A0}$, and the analogous notations for q_{10}, q_{01}, q_{11}, and q_{A1}.

In order to address confounding without additional considerations for effect measure modification (i.e. heterogeneity), constraints were imposed to guarantee homogeneous conditional ORs across strata of L. Denote the common effect of A on Y across L by OR_C, in which case P_{10} can be written as a function of P_{00} and OR_C

$$P_{10} = \frac{P_{00} \times OR_C}{1 - P_{00} + P_{00} \times OR_C} \tag{2}$$

A similar constraint can be made for P_{11}

$$P_{11} = \frac{P_{01} \times OR_C}{1 - P_{01} + P_{01} \times OR_C} \tag{3}$$

By fixing OR_C and constraining P_{10} and P_{11} as shown in expressions (2) and (3), we ensure the common odds ratio OR_C across L. It is then easy to show that

$$OR_{crude} = \frac{[q_{00} \times q_{A0} \times (1 - P_L) + q_{01} \times q_{A1} \times P_L] \times [P_{10} \times P_{A0} \times (1 - P_L) + P_{11} \times P_{A1} \times P_L]}{[q_{10} \times P_{A0} \times (1 - P_L) + q_{11} \times P_{A1} \times P_L] \times [P_{00} \times q_{A0} \times (1 - P_L) + P_{01} \times q_{A1} \times P_L]}$$

$$OR_C = \frac{P_{10} \times q_{00}}{P_{00} \times q_{10}} = \frac{P_{11} \times q_{01}}{P_{01} \times q_{11}}$$

$$OR_m = \frac{[q_{00} \times (1 - P_L) + q_{01} \times P_L] \times [P_{10} \times (1 - P_L) + P_{11} \times P_L]}{[q_{10} \times (1 - P_L) + q_{11} \times P_L] \times [P_{00} \times (1 - P_L) + P_{01} \times P_L]}$$

Table 2. Decomposition of the total discrepancy between the crude OR and the conditional OR.

$P(Y=1\mid A=0, L=0)$	$P(Y=1\mid A=0, L=1)$	OR_C	OR_m	OR_{crude}	$P(L=1)$	P_{A1}	P_{A0}	Noncollapsibility effect $\log(OR_m)-\log(OR_C)$	Confounding bias $\log(OR_{crude})-\log(OR_m)$	Total discrepancy $\log(OR_{crude})-\log(OR_C)$
0.2	0.6	2.667	2.253	2.253	0.45	0.500	0.500	−0.169	0	−0.169
0.2	0.6	2.667	2.253	2.667	0.45	0.556	0.455	−0.169	0.169	0
0.2	0.9	2.667	1.764	2.302	0.45	0.556	0.455	−0.413	0.266	−0.147

The decomposition of the total discrepancy for a few specific numerical examples is shown explicitly in Table 2. This table demonstrates the phenomena of "noncollapsibility without confounding" and of "confounding without noncollapsibility."

All the conditional ORs are homogeneous and equal to 2.667. In the first row of Table 2 confounding bias equals 0, and the total discrepancy equals the noncollapsibility effect ("noncollapsibility without confounding"). In the second row the noncollapsibility effect and confounding bias cancel each other out, thus leading to equality of the crude and conditional ORs in the presence of confounding ("confounding without noncollapsibility"). The last row shows a data example in which the noncollapsibility effect and the confounding bias both exist and the two components do not cancel each other out. eTables 1 to 3 in Online Appendix 1 (available at: http://www.smm.sagepub.com) provide hypothetical data for each row of Table 2. Our analytical approach can also be used to assess these conditions for the absence of a noncollapsibility effect and to examine the condition for the phenomenon of confounding without noncollapsibility. The verification of the conditions is presented in Online Appendix 2.

Other formulae to quantify the noncollapsibility effect have been presented previously in the literature.[9,22,24] In Online Appendix 3, we compare our analytical approach result with the other related formulae in the literature under various scenarios. The noncollapsibility effect calculated using our analytical approach is equivalent to the value obtained using the formula proposed by Samuels[22] and is reasonably close to the results obtained via the approximation suggested by Neuhaus.[24] All of the approaches give the same qualitative findings.

3 A graphical approach

In practice, one might be also interested in the nature of the relationships between the noncollapsibility effect and the effects of A (OR_C) and L (OR_L) on Y. We show in Online Appendix 4 that P_{01} can be expressed as a function of the baseline risk P_{00}, the effect of A and L on Y, the prevalence of L, and the exposure probability conditional on L. Therefore, with the constraint of homogenous ORs, we can express the noncollapsibility effect as a deterministic function of all the other parameters, i.e. P_{00}, OR_C, OR_L, P_{A1}, P_{A0}, and P_L. Again, $\log\left(\frac{OR_m}{OR_C}\right)$ was used to measure the magnitude of the noncollapsibility effect. Since there is no simple mathematical summary of the noncollapsibility effect as a function of these relevant parameters, a graphical approach was implemented. Flanders and Khoury illustrated the relationship between the magnitude of confounding and parameters graphically, and derived bounds if some of the relevant parameters are not specified.[25] Inspired by this graphical approach, we investigated the relationship between the noncollapsibility effects and the related parameters in different scenarios.

Noncollapsibility behavior can be explored in different scenarios by plotting the effect as a function of the baseline risk and different combinations of values of the other parameters. A range of different exposure effects and different covariate effects was specified. A strong association was represented by OR equal to 5 or 0.2, while a moderate association was represented by an OR of 2 or 0.5. OR values greater than 1 represent harmful effects, whereas values less than 1 represent protective effects.

The efigures provided in Online Appendix 5 display noncollapsibility under different parameter settings. The noncollapsibility effect is negative when the effect of A is harmful, and the noncollapsibility effect is positive when the effect of A is protective. This corresponds to the well-known fact that the marginal OR is always shifted toward the null compared to the conditional OR in the absence of confounding or when confounding is adjusted for appropriately.[22] One can make many additional observations from the figures that are relevant for anticipating the direction and magnitude of the noncollapsibility effect.

eFigure 1 (Online Appendix 5) shows the scenario in which there is no confounding, L is a baseline covariate, A is randomly assigned with probability 0.5 regardless of L (i.e. $P(A=1|L=1)=P(A=1|L=0)=0.5$), and the prevalence of L is 0.5. Results for other values of the prevalence of L are similar to eFigure 1 (Online Appendix 5), but the noncollapsibility effect tends to become more modest as the prevalence of L moves away from 0.5, which implies that one will observe a larger noncollapsibility effect as the prevalence of L approaches 0.5. Furthermore, the noncollapsibility effect is clearly symmetric in eFigure 1 (Online Appendix 5). That is, for any two settings (1) and (2) of the parameters, the magnitude of the noncollapsibility effect is equal in magnitude but with the opposite sign (on the log scale) if $OR_C(1)=\frac{1}{OR_C}(2)$, $OR_L(1)=\frac{1}{OR_L}(2)$, and $P_{00}(1)=1-P_{00}(2)$.

If we compare the noncollapsibility effect with $OR_C=u$ and $OR_L=v$ to the effect with $OR_C=v$ and $OR_L=u$ (for all $v>u>1$), it is observed that the first of these two noncollapsibility effect is always further from 0, which reveals the novel result that the effect of L plays a more important role in the noncollapsibility effect than does the effect of A.

After fixing either one of the A or L effects, the noncollapsibility effect becomes smaller as the magnitude of the other effect gets smaller. The noncollapsibility effect is quite modest when the effect of L and the effect of A are both small. It can also be observed that as the baseline risk goes to 0, the noncollapsibility effect disappears. It is well known that when the baseline risk is small the OR can be used to estimate the RR, and that collapsibility represents nonconfounding when the association measure is an approximate RR since there is no noncollapsibility effect.[21] However, it may be less well appreciated that as the baseline risk goes to 1, the noncollapsibility effect also disappears.

One can therefore observe an important noncollapsibility effect if the following conditions are jointly satisfied: (a) the association between the covariate and the outcome is strong, (b) the association between the exposure and the outcome is strong, (c) neither level of the outcome is rare, and (d) neither level of the covariate is rare. Furthermore, Online Appendix 5 provides figures that show the relationship between the noncollapsibility effect and the marginal outcome probability under different scenarios, including the presence or absence of confounding. Some interesting observations from those figures are discussed there. For example, eFigure 7 (Online Appendix 5) shows the relationship between the noncollapsibility effect and the baseline risk in the presence of confounding with different combinations of A and L effects and the effect of L on A (denoted by OR_{LA}). The noncollapsibility effect is again symmetric. For parameter settings (1) and (2), the magnitude of the effect is equal in magnitude but with opposite sign (on the log scale) for $OR_C(1)=\frac{1}{OR_C}(2)$, $OR_L(1)=\frac{1}{OR_L}(2)$, $OR_{LA}(1)=OR_{LA}(2)$, and $P_{00}(1)=1-P_{00}(2)$.

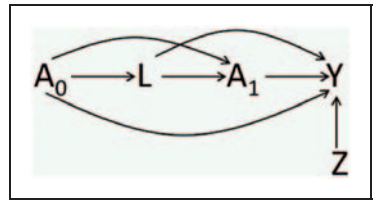

Figure 2. Time-varying confounding scenario with baseline variable Z.

4 Time-varying confounding

In observational studies, exposures may vary over time, and confounding variables may also be intermediate between exposure and outcome. Will the noncollapsibility effect be amplified in the presence of such time-varying confounding? Given the challenge in using the analytical approach in a more complex scenario, observational cohort studies were simulated to explore the noncollapsibility effect.

Figure 2 shows a simple scenario, with just two time points, no unmeasured confounding and no loss to follow-up. Let A_0 and A_1 denote the binary exposure at the first and second time points, respectively, and let Y denote the binary outcome. L is a time-varying confounding variable. If one is interested in the cumulative effect of exposure (denoted by $A = A_0 + A_1$), a MSM can be employed to assess the marginal effect of exposure in this longitudinal setting with inverse-probability-of-treatment weighted (IPTW) estimation. On the other hand, the SLRM adjusted for L always gives a biased estimate of the cumulative effect of exposure.[26] There is therefore no natural way to investigate the noncollapsibility effect with respect to the time-varying confounder L. However, we can add another baseline variable Z, which has an effect on the outcome but is independent of all other variables in the simulation study. We can thereby study the noncollapsibility effect with respect to Z and compare it to the effect from a point-exposure study with similar features.

We simulated an observational cohort study with $N = 100{,}000$ subjects who were randomly assigned to be exposed at the first time point with probability $P(A_0 = 1) = 0.5$, where $A_0 = 1$ indicates exposed and $A_0 = 0$ indicates unexposed. We let L be a factor that is affected by A_0 and also has an effect on A_1, so L was generated from a Bernoulli distribution with $P(L = 1) = \text{expit}(\alpha_0 + \alpha_1 A_0)$, where $\text{expit}(x) = \frac{\exp(x)}{1+\exp(x)}$. Consequently, we generated A_1 from a Bernoulli distribution with $P(A_1 = 1) = \text{expit}(\beta_0 + \beta_1 L + \beta_2 A_0)$, and Y from a Bernoulli distribution with $P(Y = 1) = \text{expit}(\gamma_0 + \gamma_1 L + \gamma_2 Z + \gamma_3 (A_0 + A_1))$. Z (the predictor of outcome) was generated from a Bernoulli distribution with $P(Z = 1) = 0.5$.

To estimate the cumulative effect of $A = A_0 + A_1$ on Y and measure the noncollapsibility effect of Z, we specified two models to analyze the data. After adjusting for L marginally by IPTW, Z is adjusted marginally or conditionally in model 1 and model 2, respectively. By comparing model 1 with model 2 we obtain the noncollapsibility effect of Z with respect to cumulative exposure A.

$$\text{Model 1: } \text{logit}(P(Y = 1)) = \omega_0 + \omega_1 A \qquad \textit{weights} = \frac{P(A_1|A_0)}{P(A_1|A_0, L, Z)}$$

$$\text{Model 2: } \text{logit}(P(Y = 1)) = \omega_0 + \omega_1 A + \omega_2 Z \qquad \textit{weights} = \frac{P(A_1|A_0)}{P(A_1|A_0, L)}$$

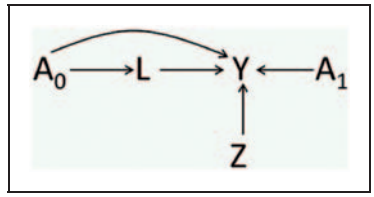

Figure 3. Point-exposure scenario with baseline variable Z.

A similar point-exposure study (Figure 3) was also simulated in order to provide a reference for the longitudinal study in Figure 2. The structure is similar without the relationship between L, A_0 and A_1. Instead, A_1 was generated from a Bernoulli distribution with $P(A_1 = 1) = 0.5$. We keep L in Figure 3 in order to have the same randomness as in the time-varying confounding case. For the point-exposure study, the noncollapsibility effect with respect to A is measured by comparing model 3 and model 4. L was left unadjusted, since it is a mediator rather than a time-varying confounding variable.

$$\text{Model 3: } \text{logit}(P(Y=1)) = \omega_0 + \omega_1 A \qquad weights = \frac{P(A_1|A_0)}{P(A_1|A_0, Z)}$$

$$\text{Model 4: } \text{logit}(P(Y=1)) = \omega_0 + \omega_1 A + \omega_2 Z$$

In the simulation, α_0, β_0, and γ_0 are parameters of no particular interest, hence we are willing to assume arbitrarily that $\alpha_0 = \beta_0 = \gamma_0 = 0$. We set α_1, β_1, β_2, and γ_1 to be log(5), log(2), log(0.5), or log(0.2). We ran the simulation with permutations of those values, so as to illustrate the scenario with strong or moderate effects and with harmful or protective effects. γ_2 and γ_3 were set to log(0.2) and log(5), respectively. For each setting, we generated 1000 independent random samples. All simulations were performed in R version 2.14.1 running on a Linux platform. The R code is provided in Online Appendix 6.

Results are presented in Table 3, showing the mean noncollapsibility effect (on the log scale). To compare the noncollapsibility effect in time-varying confounding and point-exposure studies, we present the differences between the noncollapsibility effects in the two scenarios along with 95% confidence intervals. These confidence intervals for the differences always include 0 and fall within the range ± 0.04, indicating that the noncollapsibility effect of Z with respect to the cumulative effect under the time-varying confounding scenario is comparable to that from the point-exposure study after appropriately adjusting the time-varying confounding variable L.

Informed by the observations from the point-exposure study, we anticipate that when both the effects of Z and A are in the same direction and large in magnitude, the absolute risks will approach 1 or 0, and thus the magnitude of the noncollapsibility will become trivial. Therefore, in the process of exploring the noncollapsibility effect of Z in the time-varying confounding scenario, we generally set the effect of Z on Y to OR = 0.2, and the effect of A on Y as OR = 5. Using eFigure 1 (Online Appendix 5), we can predict symmetric behavior of the noncollapsibility effects when the effects of Z and A are in the opposite direction with the same magnitude. One of the parameter settings in Table 3 again confirms this. Moreover, the main objective of the comparison was to determine whether time-varying confounding has any influence on the noncollapsibility effect. We have already described the other basic features of the noncollapsibility effect from the point-exposure study and therefore did not conduct a comprehensive simulation using all combinations of the parameters. The settings presented should be representative of the noncollapsibility effect in general time-varying confounding scenarios.

Table 3. Comparison of noncollapsibility effect between the time-varying confounding scenario and the point-exposure study.

						Noncollapsibility effect			
e^{α_1}	e^{β_1}	e^{β_2}	e^{γ_1}	e^{γ_2}	e^{γ_3}	Time-varying confounding[a]	Point-exposure study[b]	Difference	95% Confidence Interval
2	2	2	2	0.2	5	−0.167	−0.176	0.009	−0.004, 0.023
5	2	2	2	0.2	5	−0.171	−0.181	0.010	−0.004, 0.024
0.5	2	2	2	0.2	5	−0.160	−0.167	0.007	−0.005, 0.020
0.2	2	2	2	0.2	5	−0.157	−0.164	0.007	−0.006, 0.020
2	5	2	2	0.2	5	−0.164	−0.176	0.012	−0.002, 0.025
2	0.5	2	2	0.2	5	−0.177	−0.176	−0.001	−0.015, 0.013
2	0.2	2	2	0.2	5	−0.188	−0.176	−0.012	−0.026, 0.002
2	2	5	2	0.2	5	−0.163	−0.176	0.013	−0.002, 0.027
2	2	0.5	2	0.2	5	−0.175	−0.176	0.001	−0.012, 0.014
2	2	0.2	2	0.2	5	−0.181	−0.176	−0.005	−0.019, 0.008
2	2	2	5	0.2	5	−0.124	−0.131	0.007	−0.006, 0.019
2	2	2	0.5	0.2	5	−0.174	−0.178	0.004	−0.008, 0.016
2	2	2	0.2	0.2	5	−0.137	−0.135	−0.002	−0.012, 0.008
2	2	2	2	0.1	5	−0.327	−0.340	0.013	−0.005, 0.031
2	2	2	2	0.2	10	−0.219	−0.235	0.016	−0.004, 0.036
2	2	2	2	5	0.2	0.175	0.179	−0.004	−0.016, 0.007

[a]Noncollapsibility effect is measured by the difference between the estimates from Model 1 and Model 2.
[b]Noncollapsibility effect is measured by the difference between the estimates from Model 3 and Model 4.

The Z effect only enters the process and affects Y at the final time point and is independent of any other variables in the time-varying confounding scenario. Under the assumptions that there is no unmeasured confounding and no loss to follow-up, the results are expected to be comparable with the point-exposure study whenever time-varying confounding is controlled appropriately.

5 Example

To illustrate the phenomenon of collapsibility in practice, we consider an example from the Promotion of Breastfeeding Intervention Trial (PROBIT), a cluster-randomized study of a breastfeeding promotion intervention.[27] In a follow-up study, Kramer et al.[28] studied the association between breastfeeding behavior and infant weight at one year, measured by weight-for-age Z-scores. Table 4 gives a cross-tabulation of breastfeeding status during the first 6 months of life, and a binary variable indicating a positive Z-score.

While the original PROBIT design was a randomized trial, analysis of the association between breastfeeding and outcomes is potentially confounded.[29] We adjusted for several potential confounders: randomization status, geographical location (urban/rural, eastern versus western Belarus), maternal education, maternal smoking and alcohol consumption, atopic history, breastfeeding of previous children, cesarean section, the baby's gender, illness during the first 6 months of life, along with maternal age and birth weight.

We provide a crude (unadjusted) model, a marginal model (adjusted using inverse probability weighting), and a conditional model (adjusted using logistic regression) for the comparison between early weaning and exclusive breastfeeding to 6 months. Subjects who were in-between early weaning

Table 4. Cross-tabulation of breastfeeding status and weight-for-age Z-score from the PROBIT study. For illustrative purposes, subjects with missing values on any of the variables considered in the analyses are excluded.

	Weight-for-age $Z < 0$	Weight-for-age $Z \geq 0$	Total
Early weaning (<1 month)	298	980	1278
Some breastfeeding ≥1 month but not exclusive to 6 months	3226	11,172	14,398
Exclusive breastfeeding ≥6 months	75	188	263
Total	3599	12,340	15,939

Table 5. Estimated coefficients and odds ratios for the three models: The association between exclusive breastfeeding to 6 months versus early weaning and weight-for-age Z-score from the PROBIT study. For illustrative purposes, subjects with missing values on any of the variables considered in the analyses are excluded.

Model	β	Standard error	Odds ratio	95% CI
Marginal	–0.4119	0.0789	0.662	0.568, 0.773
Conditional	–0.5713	0.1901	0.565	0.389, 0.820
Crude	–0.2715	0.1518	0.762	0.566, 1.026

and exclusive breastfeeding to 6 months were excluded in this analysis. Results are given in Table 5. The total discrepancy of 0.30 between the crude and conditional effect estimates (on the log scale) can be decomposed into confounding bias of 0.14 and collapsibility effect of 0.16. When we considered adjustment for a single variable, leaving the others marginal, results were similar. This suggests that noncollapsibility can be as important as confounding under reasonable scenarios.

6 Discussion

Noncollapsibility has been used as an operational definition of confounding,[15] but it has been shown that under this definition there is a risk of detecting confounding even when the covariate is not associated with the exposure. Another definition of confounding is based on a counterfactual model:[20] "Confounding is present if our substitute population imperfectly represents what our target would have been like under the counterfactual condition." [30] It follows from this definition that confounding and noncollapsibility are different concepts,[31–33] "Collapsibility is based on the observed distribution alone, while non-confounding is a property of potential outcome distributions."[34] We emphasize that the counterfactual definition of confounding depends on the target population and focuses on causal inference, while collapsibility depends on the selected parameter and has no definitive implication for causality or confounding.[4,7]

We therefore proposed to estimate a marginal OR by using a MSM to adjust for confounding, so confounding bias could be evaluated by comparing the adjusted marginal OR and the crude OR and the magnitude of noncollapsibility effect could be measured by comparing the conditional OR and the adjusted marginal OR.

Consider the causal structure from Figure 1. "Simpson's paradox" often refers to the phenomenon that the crude OR and the conditional OR are on opposite sides of the null. Our study showed that the difference between the crude and the conditional OR can be decomposed into

a confounding bias and a noncollapsibility effect. Whenever the conditions for collapsibility of the OR are not met, the noncollapsibility effect pushes the conditional OR further from the null relative to the marginal OR, regardless of the confounding bias. Therefore, L has to be a confounding variable for Simpson's paradox to occur. The reversal of the crude and conditional OR can therefore be attributed to two distinct factors occurring jointly, confounding and a noncollapsibility effect.

As does the RD, the OR has the attractive property of symmetry under interchange of both exposure levels and outcome levels. Nonetheless, it only approximates the RR in studies where the outcome is rare in all exposure and confounder categories, or under special designs such as the case-cohort study.[35] When the outcome is common, the OR must be reported as a unique measure of association, rather than as an approximation of RR. As demonstrated in this paper, the divergence between noncollapsibility and confounding is another pronounced drawback of the OR, since noncollapsibility makes the causal interpretation of the OR difficult.[21] The OR estimate is specific to the covariates that are conditioned on in the model and subject to the outcome prevalence in the analysis dataset. It has therefore been suggested that the OR should be only reported with explicit reference to the covariates in the model.[36] That is to say, OR values cannot be compared in the same dataset across different models, nor can they be compared across different samples using the same model. Furthermore, if one considers the strata to be the characteristics of the individual, conditional ORs do not represent the averages of individual effects (i.e. the average odds that a person would get the outcome if exposed divided by the odds that the same person would get the outcome if unexposed), whenever there are other unmeasured risk factors within strata which cause an important noncollapsibility effect.[21] Therefore, it would be inappropriate to use ORs for clinical prediction of risk, and estimates for the RR or RD will be biased if they are derived from transformations of the OR.[37]

In this paper, we used analytical and graphical approaches to measure and describe the separate effects of noncollapsibility and confounding bias. We obtained the same results as Samuels, mainly because the two formulae are closely related.[22] Although Samuels developed the formula for the model in which A is independent of L, after we extend it to a more general case, it agrees with our results. Gail's results do not agree as closely with the others, presumably because it is only the first term of a Taylor expansion, and thus the approximation is less precise.[9] Although all of these approaches provide the same or similar results, all of the previous approaches concentrate on the situation where A is independent of L, whereas our approach gives a general formula to measure the noncollapsibility effect with or without confounding. In addition, the equations by Neuhaus and Gail are derived by estimation and approximation, while the formula we developed is for measuring the true value of noncollapsibility effect and provides an explicit form of the noncollapsibility effect.

Some assumptions and limitations of our study should be acknowledged. In the analytical approach, we implemented a deterministic process to measure the noncollapsibility effect. We studied the true value of the noncollapsibility effect in a source population with no sampling variability. However, random error and sampling variability are inevitable in practice. When one measures the noncollapsibility effect in finite samples, the precision of the estimate is also of interest in practice. Bootstrapping would be one method to examine the variability of the noncollapsibility effect.

Furthermore, we explored the noncollapsibility effect in very simplified scenarios, whereas reality is often much more complicated. For example, there can be more than two strata of L, and the noncollapsibility behavior could be different depending on how the strata were collapsed over. It is also possible that there is more than one baseline variable in a study, while we have only observed

the noncollapsibility effect of a single variable. When L is a vector, further study can be performed to investigate the noncollapsibility effect over some variables in L given that the others are conditionally or marginally adjusted.

Our graphical approach represented the analytical noncollapsibility effect with figures; however, the derivation of the formula expressed by the other parameters is implicit. Along with the figures, a concise and understandable formula is needed to convey more precise information from a mathematical and statistical perspective. We present the relationship between the noncollapsibility effect and the baseline risk or marginal outcome probability in two-dimensional figures while fixing the values of the other parameters. It may be of interest to extend this to express noncollapsibility simultaneously as a function of multiple parameters.

For the time-varying confounding scenario, we also focused on relatively limited settings. For example, the intercepts of the models were set arbitrarily to zero, and only a few values of the parameters were specified. Further research should aim at exploring more possible values of the parameters to make the comparison more complete.

In conclusion, the noncollapsibility of the OR should neither be ignored nor confused with confounding. The noncollapsibility effect depends on a variety of parameters, i.e. the baseline risk, the effect of the exposure, the prevalence, and the effect of the covariate. Particularly, the effect of the covariate plays a more important role than does the effect of the exposure, a result that has not been reported previously. Lastly, simulation results suggested that the noncollapsibility effect over a baseline variable in a time-varying confounding scenario is comparable in magnitude to the point-exposure study if the time-varying confounder was adjusted appropriately in the model.

Acknowledgements

The authors thank Dr Michael Kramer for access to the PROBIT data. PROBIT was supported by grants from the Thrasher Research Fund, the National Health Research and Development Program (Health Canada), the United Nations Children's Fund, and the European Regional Office of the World Health Organization. The authors thank Dr Tibor Schuster and Dr Sander Greenland for helpful comments on the manuscript.

Funding

Dr Kaufman was funded by the Canada research Chairs Program. Dr Platt was supported by a Chercheur-national (National Research Scholar) award from the Fonds de la Recherche en Sante du Quebec (FRSQ) [grant number 24919] and core funding to the McGill University Health Centre Research Institute from the FRSQ.

Conflict of Interest Statement

The Authors declare that there is no conflict of interest.

References

1. Whittemore AS. Collapsibility of multidimensional contingency tables. *J R Stat Soc Series B Stat Methodol* 1978; **40**: 328–340.
2. Ducharme GR and LePage Y. Testing collapsibility in contingency tables. *J R Stat Soc Series B Stat Methodol* 1986; **48**: 197–205.
3. Greenland S and Mickey RM. Closed-form and dually consistent methods for inference on collapsibility in 2 * 2*K and 2* J* K tables. *J R Stat Soc Series C Appl Stat* 1988; **37**: 335–343.
4. Greenland S, Robins JM and Pearl J. Confounding and collapsibility in causal inference. *Stat Sci* 1999; **14**: 29–46.

5. Newman SC. *Biostatistical methods in epidemiology*. New York: Wiley, 2001, pp.49–50.
6. Clogg CC, Petkova E and Shihadeh ES. Statistical methods for analyzing collapsibility in regression models. *J Educ Behav Stat* 1992; **17**: 51–74.
7. Greenland S and Morgenstern H. Confounding in health research. *Annu Rev Public Health* 2001; **22**: 189–212.
8. Groenwold RHH, Moons KGM, Peelen LM, et al. Reporting of treatment effects from randomized trials: A plea for multivariable risk ratios. *Contemp Clin Trials* 2011; **32**: 399–402.
9. Gail MH, Wieand S and Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* 1984; **71**: 431–444.
10. Wermuth N. Parametric collapsibility and lack of moderating effects in contingency tables with a dichotomous response variable. *J R Stat Soc Series B Stat Methodol* 1987; **49**: 353–364.
11. Shapiro SH. Collapsing contingency tables—a geometric approach. *Am Stat* 1982; **36**: 43–46.
12. Geng Z and Li G. Conditions for non-confounding and collapsibility without knowledge of completely constructed causal diagrams. *Scand J Stat* 2002; **29**: 169–181.
13. Hernán MA, Clayton D and Keiding N. The Simpson's paradox unraveled. *Int J Epidemiol* 2011; **40**: 780–785.
14. Greenland S and Pearl J. Adjustments and their consequences—collapsibility analysis using graphical models. *Int Stat Rev* 2011; **79**: 401–426.
15. Grayson DA. Confounding confounding. *Am J Epidemiol* 1987; **126**: 546–553.
16. Greenland S, Morgenstern H, Poole C, et al. RE: Confounding confounding (letter). *Am J Epidemiol* 1989; **129**: 1086–1089.
17. Becher H. The concept of residual confounding in regression models and some applications. *Stat Med* 1992; **11**: 1747–1758.
18. Yanagawa T. Case-control studies: Assessing the effect of a confounding factor. *Biometrika* 1984; **71**: 191–194.
19. Guo J and Geng Z. Collapsibility of logistic regression coefficients. *J R Stat Soc Series B Stat Methodol* 1995; **57**: 263–267.
20. Newman SC. Commonalities in the classical, collapsibility and counterfactual concepts of confounding. *J Clin Epidemiol* 2004; **57**: 325–329.
21. Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. *Am J Epidemiol* 1987; **125**: 761–768.
22. Samuels ML. Matching and design efficiency in epidemiological studies. *Biometrika* 1981; **68**: 577–588.
23. Janes H, Dominici F and Zeger S. On quantifying the magnitude of confounding. *Biostatistics* 2010; **11**: 572–582.
24. Neuhaus JM, Kalbfleisch JD and Hauck WW. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *Int Stat Rev* 1991; **59**: 25–35.
25. Flanders WD and Khoury MJ. Indirect assessment of confounding: Graphic description and limits on effect of adjusting for covariates. *Epidemiology* 1990; **1**: 239–246.
26. Hernán MA, Brumback B and Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 2000; **11**: 561–570.
27. Kramer MS, Chalmers B, Hodnett ED, et al. Promotion of Breastfeeding Intervention Trial (PROBIT): A randomized trial in the Republic of Belarus. *JAMA* 2001; **285**: 413–420.
28. Kramer MS, Guo T, Platt RW, et al. Breastfeeding and infant growth: Biology or bias? *Pediatrics* 2002; **110**: 343–347.
29. Kramer MS, Moodie EM, Dahhou M, et al. Breastfeeding and infant size: Evidence of reverse causality. *Am J Epidemiol* 2011; **173**: 978–983.
30. Maldonado G and Greenland S. Estimating causal effects. *Int J Epidemiol* 2002; **31**: 422–429.
31. Miettinen OS and Cook EF. Confounding: Essence and detection. *Am J Epidemiol* 1981; **114**: 593–603.
32. Greenland S and Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol* 1986; **15**: 413–419.
33. Pang M, Kaufman JS and Platt RW. Mixing of confounding and non-collapsibility: A notable deficiency of the odds ratio. *Am J Cardiol* 2013; **111**: 302–303.
34. Thomas R. Measure of dependence, confounding and collapsibility. SISMID 2011; Causal Model Lecture 3 (unpublished course notes obtained from the author).
35. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 1986; **73**: 1–11.
36. Norton EC. Log odds and ends. NBER working paper series. Working Paper 18252. http://www.nber.org/papers/w18252.
37. Greenland S. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. *Am J Epidemiol* 2004; **160**: 301–305.