ELSEVIER

A Positive or a Negative Confounding Variable? A Simple Teaching Aid for Clinicians and Students

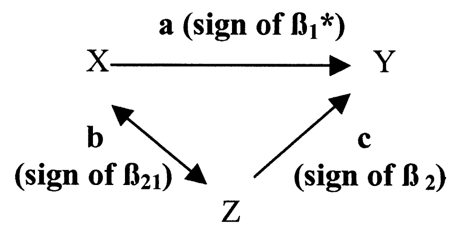
ABLA MEHIO-SIBAI, PhD, MANNING FEINLEIB, MD, DrPH, TAREK A. SIBAI, Bsc, AND HAROUTUNE K. ARMENIAN, MD, DrPH

Anticipating the direction of a confounding variable can be problematic especially to introductory students. Using elementary rules of mathematics, we describe below a simple instructional tool for deriving the direction of confounding bias. The tool is illustrated with examples and a heuristic mathematical justification is also described.
Ann Epidemiol ■.

KEY WORDS: Positive Confounding, Negative Confounding, Bias, Teaching Aid.

INTRODUCTION

Confounding is a central concept in observational epidemiology, and anticipating the role of confounding variables, as positive or negative, on effect measures is important in interpreting results. In particular, we need to judge the direction of the effect of a confounder since failure to adjust for it can lead either to an over- or under-estimate of the primary association of interest. Earlier, Vander Stoep and colleagues (1) used a didactic visual device to help introductory students understand how a third variable will affect an association between an exposure and a binary outcome. In their recent book, Szklo and Javier-Nieto (2) summarize in a table (Table 5–8) the expectations of changes brought about by adjustment for a confounder based on the direction of association between the confounder and both exposure and outcome. However, both presentations are limited to a certain scenario where the relation between exposure and outcome is positive; and, hence, information provided is not applicable to the situation where exposure decreases the likelihood of the outcome. Using elementary rules of mathematics we describe below a more comprehensive and simpler instructional aid to be used by students and researchers, including non-epidemiologists, to ascertain the direction of the confounder. A heuristic mathematical justification of this analogy is also described.



The ßs refer to the magnitude of associations between variables

OVERVIEW: MATHEMATICAL ANALOGY

Consider, in a relation between an exposure (X) and an outcome (Y), a covariate (Z). The three-way relation between these variables is symbolized by the following figure, with the confounder (Z) being a variable associated with the exposure and itself being an independent risk factor for the outcome.

Each of the variables (X, Y, Z) can either be positively (+ve, i.e., increases the likelihood of the other variable) or negatively (−ve, i.e., decreases the likelihood) associated with the other. We apply knowledge of the direction of the relation (a) between X and Y and that between Z and both X (b) and Y (c), to determine the effect of ignoring the confounder on the magnitude of the crude relation between exposure and outcome. The direction of the net effect corresponds to the sign resulting from multiplication of the three respective relations (sign of triple product, Table 1), such that a positive resultant indicates positive confounding, and a negative resultant indicates negative confounding (Table 1).

Demonstrations

As an example, consider a study to determine the effect of education (Z) on the association between smoking (X) and cardiovascular disease (CVD, Y). The effect of confounding bias depends on the direction of the association between smoking and CVD (a is positive) as well as on the direction

Department of Epidemiology and Population Health, Faculty of Health Sciences, American University of Beirut, Beirut, Lebanon (A.M-S.); Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD (M.F., H.K.A.); and Faculty of Medicine, American University of Beirut, Beirut, Lebanon (T.A.S.).

Address correspondence to: Dr. Abla Mehio-Sibai, Department of Epidemiology and Population Health, Faculty of Health Sciences, American University of Beirut, Beirut, Lebanon. Tel.: +961-1-744-470; Fax: +961-3-646-688. E-mail: ansibai@aub.edu.lb

Received February 19, 2004; accepted October 5, 2004.

360 Park Avenue South, New York, NY 10010

1047-2797/05/$–see front matter
doi:10.1016/j.annepidem.2004.10.004

ARTICLE IN PRESS

TABLE 1. Anticipating the direction of the confounder based on the direction of the associations between exposure, outcome, and covariate

Case	Direction of association* a	b	c	Sign of confounding bias (b*c)	Sign of triple product (a*b*c)	Direction of confounder†
1.	+	+	+	+	+	Positive
2.	+	+	−	−	−	Negative
3.	+	−	+	−	−	Negative
4.	+	−	−	+	+	Positive
5.	−	+	+	+	−	Negative
6.	−	+	−	−	+	Positive
7.	−	−	+	−	+	Positive
8.	−	−	−	+	−	Negative

A positive sign (+) indicates a positive association or a risk factor.
A negative sign (−) indicates a negative association or a protective factor.
*Each of the relations a, b, and c specifies direction of the following relations: primary relation of interest, association between confounder and exposure *or* association between confounder and outcome.
†A positive confounder: the unadjusted estimate of the primary relation between exposure and outcome will be pulled further away from the null hypothesis than the adjusted measure. A negative confounder: the unadjusted estimate will be pushed closer to the null hypothesis.

of the association between education and both smoking (b is negative) and CVD (c is negative). The resultant sign of the three relations is, in this example, positive. Consequently, we anticipate a spurious strengthening of the exposure–outcome relation between smoking and CVD if education were not taken into account (case 4 in the table).

Alternatively, consider the role of smoking (Z) as a confounder in the relationship between alcohol (X) and CVD (Y). With the signs of the interrelationships being negative, positive, and positive for a, b, and c, respectively, one can directly deduce a negative confounding effect with a diminution of the magnitude of the alcohol–CVD relation if smoking were not taken into account (case 5 in the table).

A Heuristic Mathematical Derivation

The strength of the exposure–disease relation may be defined by the crude effect (β_1*) of X on Y with

$$Y = \beta_0^* + \beta_1^* X + e^*$$

The relation can also be defined by the adjusted effect (β_1) of X on Y, taking into account covariate, Z,

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + e$$

And the association of Z with X may be represented by

$$Z = \beta_0^{**} + \beta_{21} X + e^{**}$$

where β_{21} represents the regression of Z on X.

The total effect of X on Y is a function of the direct effect and the regression of Z on X, i.e.,

$$\begin{aligned} Y &= \beta_0 + \beta_1 X + \beta_2(\beta_0^{**} + \beta_{21} X + e^{**}) + e \\ &= (\beta_0 + \beta_2\beta_0^{**}) + (\beta_1 + \beta_2 \beta_{21})X + (\beta_2 e^{**} + e) \end{aligned}$$

Consequently,

$$\beta_1^* = \beta_1 + \beta_2 \beta_{21}$$

The bias due to confounding is the product $\beta_2 \beta_{21}$, and the confounded estimate, β_1*, will be different from β_1, the adjusted one, whenever β_2 and β_{21} are both non-zero, i.e., when Z, the confounding variable, is associated with both X and Y. The resultant effect of the confounder on the exposure–outcome relation depends on the sign of each of the estimates β_1 and $\beta_2 \beta_{21}$.

In the case when the resultant sign of $\beta_2 \beta_{21}$ (sign of confounding bias b*c, Table 1) is similar to that of β_1, the net effect is to inflate the crude estimate and β_1* would be pulled further away from the null hypothesis. In such a case, the confounder is said to have a positive effect. In contrast, in the case when the resultant sign of $\beta_2 \beta_{21}$ is different from β_1, the net effect is to deflate the crude estimate and β_1* is pushed closer to the null hypothesis, and the confounder is said to have a negative effect.

In extreme cases, a negative confounder may result not only in a change in the strength of the association, but also, in a divergence of its direction as well. This occurs in the case of a strong confounder where the magnitude of the product $\beta_2 \beta_{21}$ is larger than that of β_1. Hence, β_1* will be different from β_1 both in magnitude and sign.

When *either* β_2 or β_{21} is equal to zero, β_1* becomes equal to β_1 and confounding is not present. For example, randomization in clinical trials usually makes the distribution of potential confounders similar among the categories of main exposure of interest, thus rendering $\beta_{21} = 0$.

CONCLUSION

The table presented should provide teachers, students, and researchers a brief and straightforward derivation for predicting the direction of confounding bias (positive or negative), more comprehensive and simpler than those earlier presented in the literature. An understanding of how an uncontrolled potential confounder is likely to affect the primary association of interest is very crucial in cases where information on the confounding variable *was not or could not be obtained* (3). To this end, the results observed and published for an association in which the missing confounder is judged as having a negative weak effect are more likely to represent a conservative estimate and a stronger argument for a true association than in the case of a positive confounding. In contrast, failure to account for a negative but strong confounder can lead to results that may be misleading.

REFERENCES

1. Vander Stoep A, Beresford SAA, Weiss N. A didactic device for teaching epidemiology students how to anticipate the effect of a third factor on an exposure–outcome relation. Am J Epidemiol. 1999;150:221.
2. Szklo M, Javier-Nieto F. Identifying non causal associations: Confounding. In: Epidemiology: Beyond the Basics. Gaithersburg, MD: Aspen Publishers, Inc.; 2000:203.
3. Hennekens CH, Buring JE. Analysis of epidemiologic studies: Evaluating the role of confounding. In: Epidemiology in Medicine. Boston/Toronto: Little Brown and Co.; 1987:287–323.