Introduction to confounding (5.1)

From an epidemiologic standpoint, it is useful to conceptualize confounding as distinct from bias in that a confounded association, although not causal, is real⚓

If, on the other hand, the goal of the researcher is to carry out primary prevention, it becomes crucial to distinguish a causal from a noncausal association, the latter resulting from either bias or confounding. ⚓

A way to explain confounding, which is found with increasing frequency in the epidemiology literature, is based on the counterfactual model⚓

This explanation relies on the notion that the attribution of causation to a particular exposure is based on the concept that the risk of the outcome observed among individuals exposed would have been different in these same individuals had exposure been absent.⚓

Because the latter group is, by definition, unobservable ("counterfactual"), epidemiologists are forced instead to use a different group (the unexposed) as comparison; thus, the essence of epidemiologic analysis rests on the assumption that a separate unexposed group is comparable to exposed individuals if they had not been exposed. ⚓

Using an alternative terminology, confounding is described as present when the exposed–unexposed groups are nonexchangeable.⚓

Nature of association between confounder, exposure and outcome (5.2)

The common theme with regard to confounding is that the association between an exposure and a given outcome is induced, strengthened, weakened, or eliminated by a third variable or group of variables (confounders⚓

> **The confounding variable is causally associated with the outcome**
>
> **and**
>
> **noncausally or causally associated with the exposure**
>
> **but**
>
> **is not an intermediate variable in the causal pathway between exposure and outcome**

⚓

The general rule defining a confounding variable excludes the situation in which the exposure determines the presence or level of the presumed confounder. ⚓

Exceptions to the General Rule for the Presence of Confounding⚓

Although the general rule is that the confounding variable must be causally associated with the outcome (Figure 5-1), sometimes a random (statistical) association results in confounding.⚓

. Thus, for example, in a case-control study, sampling (random) variability may create an imbalance between cases and controls with regard to a given variable related to the exposure even though there is no such imbalance in the total reference population⚓

It is important to understand that this phenomenon is not exclusive to observational studies.
The notion that randomized clinical trials are free of confounding is an average concept⚓

Variables treated as confounders are occasionally surrogates of the true confounding variable(s). ⚓

As discussed previously, the potential confounder should not be an intermediate variable in the causal pathway between the suspected risk factor and the outcome. It follows that it is inappropriate to adjust for such a variable. Although this rule is generally sound, exceptions to it occur when the investigator deliberately explores alternative mechanisms that could explain the association between the exposure and the outcome of interest.⚓

After the principal link in the causality chain (hypertension) is established, however, a different question may be asked: "Does obesity cause death through mechanism(s) other than hypertension (e.g., through diabetes)?" (FIGURE 5-5). In this situation, to treat hypertension as a confounder (at least in the sense of controlling for it) is appropriate.⚓
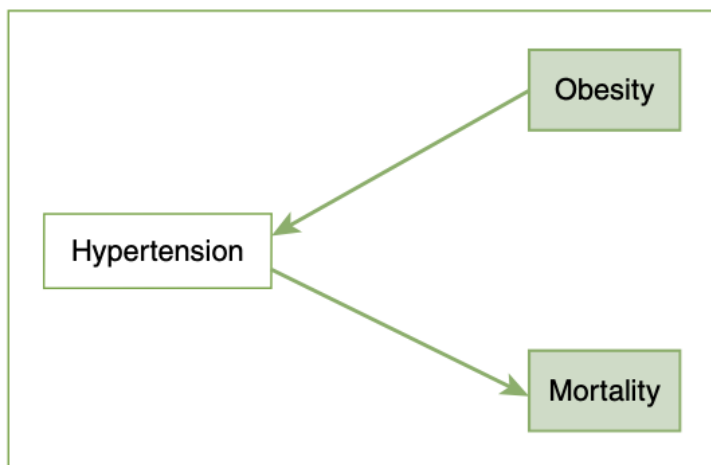
FIGURE 5-4 Schematic representation of hypertension as the only mechanism for the relationship between obesity and death.
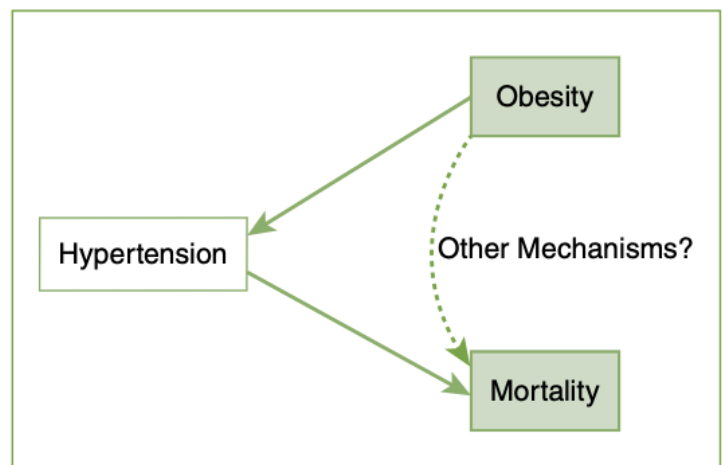

FIGURE 5-5 Schematic representation of mechanism(s) other than hypertension (e.g., diabetes) to explain the relationship between obesity and death.

$$\% \text{ Excess Risk Explained} = \frac{RR_U - RR_A}{RR_U - 1.0} \times 100 \qquad \text{(Eq. 5.1)}$$

It is important to keep in mind, however, that if a residual association persists after a potentially intermediate variable is controlled for, this does not necessarily mean that there must be other causal pathways or mechanisms; the residual association may be due to residual confounding

, the concept of proportion of effect explained by a putative mechanism (Equation 5.1) relies on the assumption that the relationship between the suspected intermediary variable and the outcome is free of confounding

Theoretical and graphical aids to frame confounding (5.3)

The importance of selecting the proper statistical model when analyzing epidemiologic data is discussed in detail in Chapter 7, Section 7.8.

For this purpose, some level of understanding (even if hypothetical) is necessary of the underlying pathophysiologic (or psychosocial) pathways representing the relationships between the suspected risk factor(s), confounder(s), and outcome(s). The more explicit this mechanistic model, the more straightforward the analytical approach and data interpretation will be.

A potentially useful analytical aid for this purpose is the directed acyclic graph (DAG)

. Originally developed in the field of artificial intelligence, this technique has been more recently used in the social science fields, including epidemiology

The DAG (also known as a "causal diagram") is a formal and more elaborate extension of traditional graphs to represent confounding, such as shown in Figure 5-2; in these graphs,

the direction of the association between the variables of interests and other unknown confounders is explicitly displayed to facilitate and guide the casual inference process.

It is important to point out, however, that although helpful, the use of DAGs is limited by their inability to handle effect modification and a large number of variables, among other limitations.[20]

Assessing the presence of confounding (5.4)

. The identification of potential confounders is usually based on a priori knowledge of the dual association of the possible confounder with the exposure and the outcome, the two poles of the study hypothesis

There are several approaches for assessing the presence of confounding, which are related to the following questions:

1. Is the confounding variable related to both the exposure and the outcome in the study?
2. Does the exposure–outcome association seen in the crude analysis have the same direction and similar magnitude as the associations seen within strata of the confounding variable?
3. Does the exposure–outcome association seen in the crude analysis have the same direction and similar magnitude as that seen after controlling (adjusting) for the confounding variable?

EXHIBIT 5-3 Association of the putative confounder (mostly outdoor occupation) with the exposure of interest (male gender) and the outcome (malaria) in the hypothetical study in Exhibit 5-2 and Figure 5-8.

Confounder versus exposure

	Mostly outdoor	Mostly indoor	Total	
Males	68	88	156	**Odds ratio = 7.8**
Females	13	131	144	
			300	

Confounder versus outcome

	Cases	Controls		
Mostly outdoor	63	18		**Odds ratio = 5.3**
Mostly indoor	87	132		
Total	150	150	300	

Perhaps the most persuasive approach to determine whether there is a confounding effect is the comparison between adjusted and crude associations.

As described in that section, the Mantel-Haenszel adjusted odds ratio in this example is 1.01. (This odds ratio is merely a weighted average of the stratum-specific odds ratios shown in Exhibit 5-4.)

*Equation 5.1 was presented as a tool to evaluate the proportion of the association explained by a potentially mediating variable. As this example illustrates, however, it may also be used to assess the proportion of the association explained by confounding; as with the previous application, the importance of a proper theoretical basis and required assumptions needs to be carefully considered (see Sections 5.3 and 5.4).

Additional Issues Related to Confounding (5.5)

Although, as discussed previously in this chapter, the most persuasive evidence supporting presence of confounding is the demonstration that the crude and adjusted estimates differ, it is useful to consider the other strategies discussed to evaluate confounding

Stratification is also a useful step when analyzing epidemiologic data, as it also allows the formulation of an a priori expectation of the effects of confounding on the association and, thus, of the effects of adjustment on the association (notwithstanding the noncollapsibility caveat, previously referred to, that even when confounding is absent the pooled measure of association may be different from the stratum-specific measures of association).

Another reason stratification is a useful analytical strategy is that, in addition to assessing confounding, it allows the examination of the presence of interaction, as discussed in more detail in Chapters 6 and 7.

5.5.2 Confounding Is Not an "All-or-None" Phenomenon

In other instances, however, the confounding effect is only partial.

a relatively strong and statistically significant relative risk was still observed—which, in addition to resulting from residual confounding, could also mean that these markers confound part, but not all, of the entirety of the observed associations

Although, by definition, a confounding variable is correlated with the exposure of interest (Figure 5-1), on occasion, the correlation is so strong that adjustment becomes difficult if not impossible 〰

This is a problem analogous to the situation known in biostatistics as collinearity 〰

. The ideal situation for effective control of confounding is that in which there is a clear-cut correlation between exposure and confounder but the variability is sufficient to allow adequate representation of all cross-tabulated cells shown in FIGURE 5-10. 〰
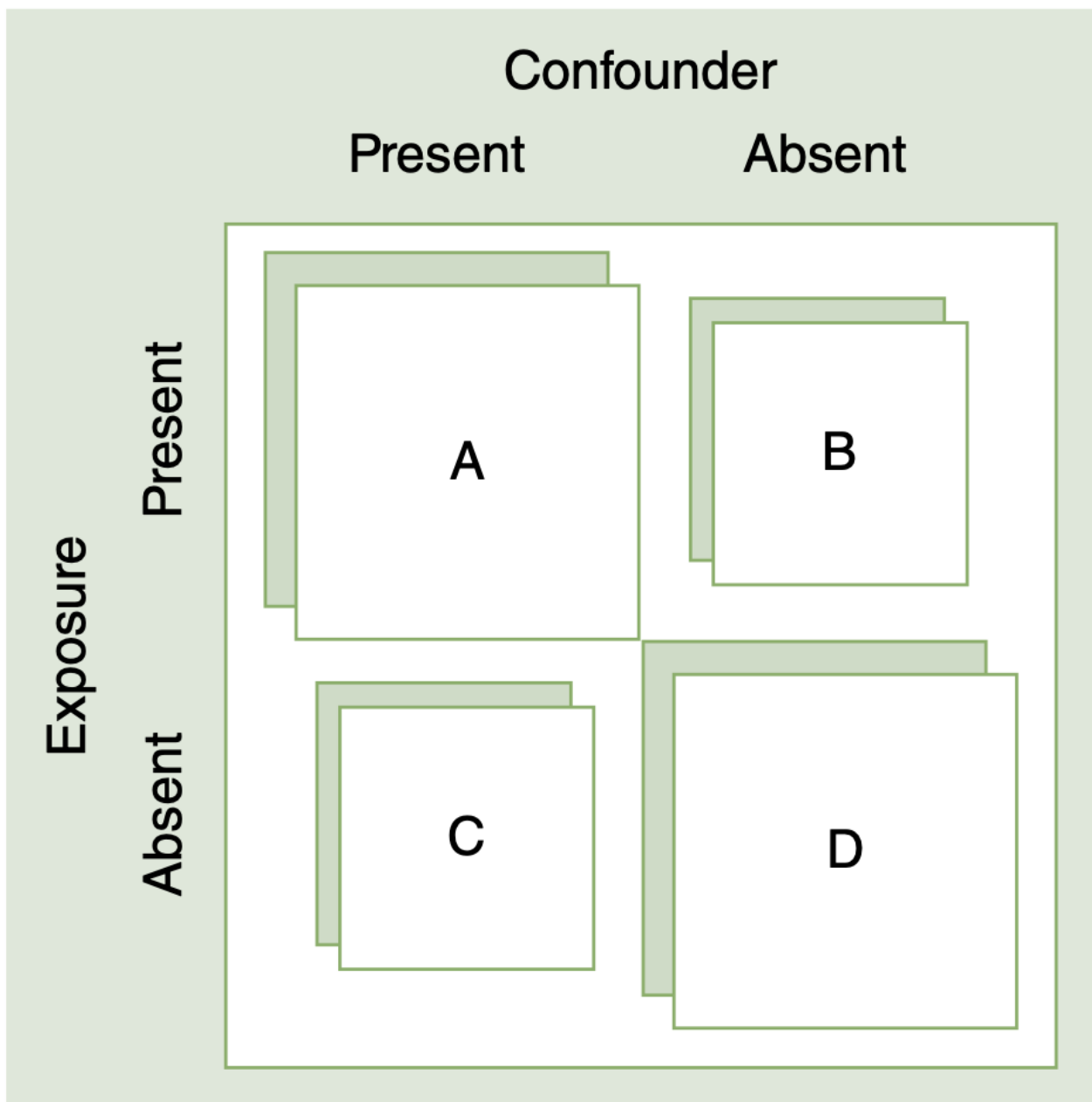
FIGURE 5-10 Correlation between an exposure of interest and a confounding factor: All four cells for a cross-tabulation of dichotomous categories are represented. In this schematic representation, the larger sizes of cells A and D denote the magnitude of the positive correlation between the exposure and the confounder.

As a corollary of the preceding discussion, it is important to underscore that it is possible to adjust for a confounder while examining the relationship between exposure and outcome only when levels of the confounder and those of the exposure overlap, a situation not always encountered in observational epidemiology studies.

In general, it must be kept in mind that when adjustment is carried out for a given confounding variable, it is also carried out for all variables related to it.

5.5.6 Statistical Significance in Assessing Confounding

Thus, even if there were no association whatsoever between occurrence of menopause and disease, if for each year of age the odds of menopause hypothetically increased from 1:1 to 1.5:1 (e.g., for an increase in menopause prevalence from 50% to 60%), a case-control age difference as small as 1 year (which might not be statistically significant if the study sample were not large) would result in an age-unadjusted menopause relative odds of 1.5.

For those who insist on using the p value as a criterion to discern whether confounding is present, it may be wiser to use more "lenient" (conservative) p values.

This strategy, however, should not replace the investigator's consideration of the strength of the associations of the suspected confounder(s) with the exposure and outcome as a means to identify confounding.

A presumed confounding variable may be confounded by other variables.