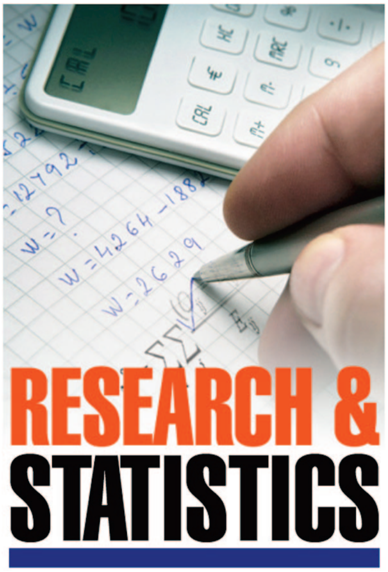
PediatricsinReview®

Understanding Confounding in Research

Kantahyanee W. Murray and Anne Duggan

Pediatrics in Review 2010;31;124
DOI: 10.1542/pir.31-3-124



Understanding Confounding in Research

Kantahyanee W. Murray, PhD,* Anne Duggan, ScD†

Author Disclosure
Drs Murray and Duggan have disclosed no financial relationships relevant to this article. This commentary does not contain a discussion of an unapproved/investigative use of a commercial product/device.

Case Study

The association between indoor smoking and asthma exacerbation is well established. You learn that one of your asthmatic patients is exposed regularly to indoor smoking. Your patient's mother attributes her daughter's asthma exacerbations to the high level of air pollution in the community in which they live. You think that indoor smoking is a major contributor to her asthma exacerbations but are unable to convince the mother. You decide to search the literature for studies that have examined whether air pollution confounds the relationship between indoor smoking and asthma exacerbations.

Introduction

When establishing relationships of cause and effect, internal validity represents the truthfulness of conclusions about causal relationships. Internal validity means that a true cause-and-effect relationship exists between an exposure (the cause) and outcome (the effect) variable. Confounding is one of several threats to the internal validity of a research study. (1) Confounding is defined as a possible source of bias in studies in which an unmeasured third variable (the confounder) is related to the exposure of interest (although not causally) and causally related to the outcome of interest. (2)

Understanding confounding is critical in determining what inferences can be drawn from study findings. Is indoor smoking a cause of asthma exacerbation or is a third factor (eg, air pollution) the true cause? Failure to detect confounding can lead to wrong conclusions. Several issues must be considered when evaluating the approaches used by researchers to deal with confounding. This article describes confounding and strategies to address confounding during both study design and data analysis. The article also describes the strengths and weaknesses of each strategy.

Criteria for Confounding

For air pollution to confound the relationship between indoor smoking (exposure) and asthma exacerbation (outcome) in children, the following criteria must be satisfied:

1. Air pollution is a risk factor for asthma exacerbation.
2. Air pollution is associated with indoor smoking.
3. Air pollution is not an intermediate step in the causal path between indoor smoking and asthma exacerbation.

If these three criteria are met, a causal relationship between indoor smoking and asthma exacerbation does not exist (Fig. 1). Indoor smoking and asthma exacerbation are associated.

In certain situations, the first, but not the second, criterion for confounding is met and the third factor (eg, air pollution) is associated only with the outcome (eg, asthma exacerbation). When added to a statistical model, the third factor is an additional variable that may help to ex-

*Ruth H. Young Center for Children and Families, University of Maryland School of Social Work, Baltimore, Md.
†Department of General Pediatrics & Adolescent Medicine, The Johns Hopkins University School of Medicine, Baltimore, Md.

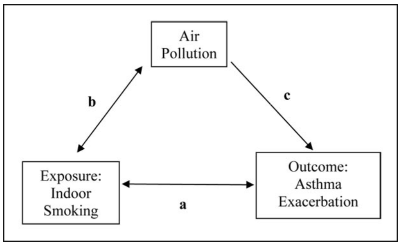


Figure 1. Hypothetical graphic representation of how air pollution confounds the association between indoor smoking and asthma exacerbation. Paths a and b (double-headed arrows) denote relationships of association. Path c (single-headed arrow) denotes a causal relationship.

plain the outcome. In this scenario, the third factor does not confound the relationship between the exposure and the outcome. This type of relationship among the variables is distinct from confounding.

Confounding is not Mediation

Confounding often is confused with mediation. Both confounders and mediators are related causally to the outcome. Although both confounders and mediators are related to the exposure, the relationship is causal for mediators, but not for confounders. As shown in Figure 2, an arrow leads from the exposure to the mediator, indicating that the mediator is an effect of the exposure variable.

Dealing With Confounding in the Study Design

In testing for true causal effects, it is essential to prevent, minimize, or

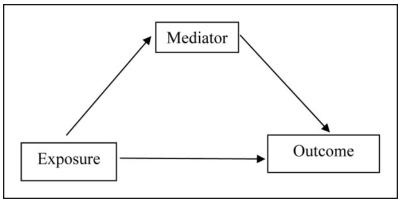


Figure 2. Graphic representation of a mediation model.

measure the effects of confounding. One approach is through study design. Specification and matching are strategies to handle confounding in the study design. Specification (also called restriction) involves restricting study enrollment to participants who have a specific value or level of the confounding variable. (3)(4) For example, if air pollution is believed to be a confounding factor, a study can enroll only individuals living in low-pollution areas. If an association between smoking and asthma exacerbation is observed, air pollution cannot be the cause because it has been held constant. One disadvantage of specification is that it decreases sample size and sample heterogeneity, thus reducing statistical power and external validity (generalizability).

Matching is another study design strategy to prevent confounding. One approach is individual-level matching, using a case-control study design in which cases and controls are matched in regard to the confounding factor. (4) In the case example, cases and controls could be matched by pollution exposure status. One advantage of this approach is that it can be extended to allow matching on several different potential confounders. One key disadvantage of matching is that it precludes study of the effect of the confounder(s) on the outcome.

Dealing With Confounding After Study Completion

Stratification and multivariate techniques are two analytic strategies to handle confounding. Stratification involves subdividing subjects by levels of a potential confounding variable, then testing for the association of the exposure with the outcome within each stratum. (3) Is indoor smoking associated with asthma exacerbation regardless of the level of air pollution? To answer this question, consider the following hypothetical study of the associations among air pollution, indoor smoking, and children's asthma exacerbation. The variable air pollution is categorized into two strata (low-level and high-level air pollution), and the indoor smoking variable is a dichotomous measure (exposed to indoor smoke and not exposed to indoor smoke). As shown in the Table, the hypothetical study reveals that the prevalence of asthma exacerbation is higher for children exposed to indoor smoking in the high-level air pollution strata compared to children exposed to indoor smoking in the low-level air pollution strata. In this example, the odds of asthma exacerbation are higher for children who are exposed to indoor smoking compared with those who are not exposed to indoor smoking, regardless of level of air pollution.

In many instances, a researcher uses stratification to deal with known confounders. Cases are stratified by the known confounding variable(s), and the researcher tests the association between the exposure and the outcome within each stratum. One major weakness of the stratification strategy is that it may not be feasible to handle multiple confounders. As the number of strata increase, sample size within each stratum decreases, reducing statistical power. Another weakness is that stratification might not adequately control for confounding. (3) For example, dichotomizing a continuous variable to create two strata (eg, low pollution versus high pollution) might result in too crude a measure to control adequately for pollution as a confounder.

Researchers commonly use multivariable statistical techniques to adjust for confounders when examining relationships between cause and effect variables. Techniques such as

Table. Hypothetical Example of Stratification by Air Pollution Level

Level of Air Pollution Exposure	Exposed to Indoor Smoke	Asthma Exacerbation Yes	Asthma Exacerbation No	Odds Ratio
High Pollution	Yes	50	20	$\frac{50 \times 30}{50 \times 20} = \frac{1500}{1000} = 1.5$
	No	50	30	
	Total	100	50	
Low Pollution	Yes	10	16	$\frac{10 \times 216}{90 \times 16} = \frac{2160}{1440} = 1.5$
	No	90	216	
	Total	100	232	

multivariable logistic regression or structural equation modeling permit an understanding of how much variability in an outcome is accounted for by a confounder. (4) Multivariable statistical techniques also permit researchers to control for more factors than stratification. One potential drawback is that multivariable techniques require readers to understand how to interpret the meaning of adjusted odds ratios and regression coefficients as well as how statistical significance was determined.

Summary

- When confounding is present, a real association does exist between the exposure and the outcome. However, the hypothesized exposure is not the cause of the outcome. A third variable, the confounder, is the direct cause of the outcome. The confounder also is related to the exposure, although not as a *cause* of the exposure. (4)
- Researchers use several different strategies to address confounding.
- Dealing with confounding is necessary, but not sufficient, for demonstrating internal validity. Other sources of systematic or random error should be addressed as well.

References

1. Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet*. 2002;359:248–252

2. DeAngelis C. *An Introduction to Clinical Research*. New York, NY: Oxford University Press; 1990

3. Gordis L. *Epidemiology*. 4th ed. Philadelphia, Pa: Saunders Elsevier; 2009

4. Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman RB, eds. *Designing Clinical Research: An Epidemiologic Approach*. 3rd ed. Baltimore, Md: Lippincott Williams & Wilkins; 2007

Clarification

Two problems have been pointed out with the quiz questions for the article on respiratory failure in the December issue (*Pediatr Rev*. 2009;30:470–478).

The correct answer to Question 5 is B, "cyanide toxicity," as listed in the answer key. The question states, "You suspect impaired oxygen use by the tissues," also known as cytochemical hypoxia, and the text lists cyanide poisoning as a cause of exogenous cytochemical hypoxia. What is not mentioned in the text is that high-dose sodium nitroprusside therapy can liberate cyanide into the system, which is why other drugs have largely supplanted nitroprusside in treating hypertension. *Pediatrics in Review* tries to avoid questions that require specific information not included in the article, but this question did so, which is a bit unfair to the reader. We regret the ambiguity.

The correct answer to Question 8 is C, "acute on chronic respiratory acidosis." The answer given in the answer key is wrong due to a typographical error. We regret the error.

For the quiz that accompanies the article on tuberculosis in the January issue (*Pediatr Rev*. 2010;31:13–26), the correct answer to Question 6 is A, "lymph nodes." The answer given in the answer key is wrong due to a typographical error. We regret the error.