

ERIC Notebook

September/October 2000

Issue 14



Michel Ibrahim, MD, PhD
Director of Education
Program

Lorraine Alexander, DrPH

Carl Shy, MD, DrPH

Gayle Shimokura, MSPH

Sherry Farr, Graduate
Research Assistant

www.sph.unc.edu/courses/eric



Ron Horner, PhD
ERIC Director

<http://hsrd.durham.med.va.gov/ERIC/>

The ERIC Notebook is funded by the Department of Veterans Affairs (DVA), Veterans Health Administration (VHA), Cooperative Studies Program (CSP), to promote the strategic growth of the epidemiologic capacity of the DVA.

Confounding Bias, Part I

Confounding is an important concept in epidemiology because, if present, it can cause an over- or under-estimate of the observed association between exposure and disease. The distortion introduced by a confounding factor can be large, and it can even change the apparent direction of an effect. However, unlike selection and information bias, it can be adjusted for in the analysis.

What is confounding?

Confounding is the distortion of the association between an exposure and disease outcome by an extraneous, third variable called a confounder. Since the exposure of interest is rarely the only factor that differs between exposed and non-exposed groups, and that also affects disease frequency, confounding is a common occurrence in etiologic studies.

Confounding is also a form of bias. Confounding is a bias because it can result in a distortion in the measure of association between an exposure and disease.

Confounding may be present in any study design (i.e., cohort, case-control, observational, ecological), primarily because it's not a result of the study design. Of all study designs, ecological studies are the most susceptible to confounding, because it is more difficult to control for confounders at the aggregate level of data. In all other cases, as long as there is available data on potential confounders, they can be adjusted for in the analysis stage.

Confounding should be of concern under the following conditions:

1. Evaluating an exposure-disease association.
2. Quantifying the degree of association between an exposure and disease. For example, quantifying the extent that being overweight increases the risk of CHD. When a precise estimate of effect is the goal, adjusting for confounding is imperative. In one study, a rate ratio of 4 may become 3.7 after

controlling for age, whereas in another study, a rate ratio of 4 may change to 1.2 after controlling for age.

3. Multiple causal pathways may lead to the disease. If there is only one way to contract the disease, confounding cannot occur. This criterion is almost always met as diseases can inevitably be caused by different agents, different transmission routes, or different biological or social mechanisms.

Thus, control of confounding is needed for the following research questions:

Does being overweight increase the risk of coronary heart disease (CHD) -- independently of cholesterol, hypertension, and diabetes?

Does tobacco advertising entice adolescents to experiment with tobacco independently of whether or not their parents smoke?

Assessing Confounding

The following are tools with which to assess for presence of confounding.

Is confounding present?

Each potential confounder has to meet two criteria before they can be confounders:

Criterion 1 is that the potential confounder must be a known risk factor for the disease.

Broadly speaking, a risk factor is any variable that is

1. Already known to be "causally related" to the disease (though not necessarily a direct cause)

AND

2. Antecedent to the disease on the basis of substantive knowledge or theory, and/or on previous research findings.

The confounding factor must be predictive of disease occurrence apart from its association with exposure; that is, among unexposed (reference) individuals, the potentially confounding factor should be related to disease risk.

Confounder \implies **Disease**

With an epidemiological data set, one can calculate whether or not a potential confounder is a risk factor using the following mathematical formula.

Criterion 1 for Confounding--Mathematical formula

Criterion 1 for confounding is the following: among the unexposed, there should be an association between the confounder and the disease.

To convert this to a mathematical equation, the first thing to realize is that Criterion 1 involves calculating a measure of association ("there should be an association between the confounder and the disease"). Examples of measures of association are relative risk, odds ratio, and risk difference--the type of measure depends on the type of data available, and the scale on which the measure of association is assessed (additive or multiplicative scale). We will assume we want to measure differences between groups on a multiplicative scale only. This measure of association will be calculated among the unexposed population only.

For a prospective cohort study (incidence data) where we want to measure the association on a multiplicative scale, we will calculate the following risk ratio (RR):

$$RR_{CD/notE} = \frac{\text{Rate of new cases among population A}}{\text{Rate of new cases among population B}}$$

where the rate of new cases = the number of new cases divided by the total number of susceptible individuals. Population A is comprised of all individuals who have the confounder (C+) but who are unexposed (E-), and population B is comprised of all individuals who don't have the confounder (C-) or the exposure (E-).

For a case-control study using odds ratios (OR), the formula for Criterion 1 is:

$$OR_{CD/notE} =$$

$$\frac{\text{Odds that cases have confounder among population F}}{\text{Odds that controls have confounder among population F}}$$

Where the odds that the cases have the confounder = the number of cases with the confounder (C+) divided by the number of cases without the confounder (C-) and where population F is comprised of all individuals who are not

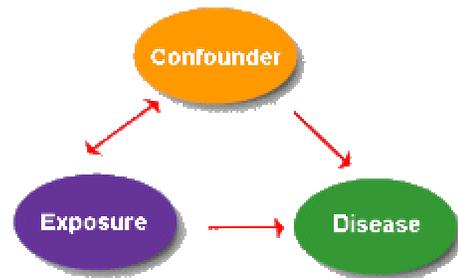
exposed (E-).

Now that the relative risk or odds ratio for the association between the confounder and disease among the unexposed has been calculated, how is it interpreted?

For the confounder to be a risk factor, the measure of association has to be greater than 1 (if a deleterious association), or less than 1 (if a protective association).

Age and smoking status, for example, are widely considered to be risk factors for lung cancer, even though the mechanisms by which both variables are determinants of this disease are not well understood. On the other hand, race is not considered to be a risk factor for lung cancer. Unnecessary adjustment of variables that are not confounders can lower precision and may even introduce bias into the estimate of effect.

Criterion 2 is that the potential confounder must be associated with the main exposure, but not as a result of the exposure. In other words, all potential confounders should be working independently and not as part of the proposed exposure-disease pathway. One can calculate whether or not a potential confounder is associated with the main exposure using a mathematical formula.



Criterion 2 for Confounding--Mathematical formula

Criterion 2 for confounding is the following: the distribution of the confounding variable differs between exposed and unexposed groups.

To convert this to a mathematical equation, the first thing to realize is that Criterion 2 involves calculating a measure of association.

For a prospective cohort study, we will calculate the following risk ratio:

$$RR_{EC} =$$

$$\frac{\% \text{ individuals with confounder (C+) among population A}}{\% \text{ individuals with confounder (C+) among population B}}$$

where Population A will be comprised of all individuals who are exposed (E+), and where Population B will be comprised of all individuals who are unexposed (E-).

For a case-control study using odds ratios (OR) the formula for Criterion 2 is:

$$OR_{crude} = \frac{\text{Odds of controls having the confounder (C+) among population A}}{\text{Odds of controls having the confounder (C+) among population B}}$$

where odds of controls having the confounder (C+) = number of controls having the confounder (C+) divided by the number of controls not having the confounder (C-). Population A is comprised of all individuals who are exposed (E+), and population B is comprised of all individuals who are unexposed (E-). Note the additional inclusion criteria for case-control studies: the individuals included in this calculation must include only those who have the potential to be cases (the control group).

Now that the relative risk or odds ratio for the association between the confounder and exposure has been calculated, how is it interpreted? For the confounder associated with the exposure, this association has to be greater than 1 (for a deleterious association) or less than 1 (for a protective association).

To decide whether a variable is working independently of the association of interest, there must be a biological or social mechanism to causally link the exposure of interest to the disease outcome. Such decisions should be made on the basis of the best available information, including nonepidemiological (i.e., clinical) data. This criterion is obviously satisfied if the confounding factor precedes the exposure and disease.

For instance, if interested in assessing the association between physical inactivity and cardiovascular disease (CVD), body weight should not be controlled for if being overweight may be an intermediary step in the causal pathway between physical inactivity and CVD.

Physical inactivity → Being overweight → CVD

In contrast, if the proposed causal pathway is independent of body weight, then body weight can be considered a potential confounder. If intervening variables are controlled for in the analysis, it may reduce or eliminate any indications in the data of a true association between disease and exposure.

The next issue of ERIC Notebook will discuss ways to control for confounders in epidemiological studies.

Self-Evaluation

Q1: In a cohort study of air pollution exposure (AP) and risk of bronchitis, you believe that smoking status may be a confounder. Use the 2x2 tables below to find out if smoking status is a confounder in the air pollution-bronchitis association.

	Bronchitis	No bronchitis	Total
Exposed to high AP	178	1129	1307
Not exposed to high AP	79	1262	1341

Among high AP:

	Bronchitis	No bronchitis	Total
Smoker	168	880	1048
Nonsmoker	10	249	259

Among low AP:

	Bronchitis	No bronchitis	Total
Smoker	34	177	211
Nonsmoker	45	1085	1130

- Is smoking an independent risk factor for bronchitis?
- Is smoking differentially distributed between high and low air pollution groups?
- Compare the crude RR with the RR's stratified by smoking status. Is the crude RR confounded by smoking status?

Answers:

1.a. To examine smoking as an independent risk factor for bronchitis, the relative risk (RR) of smoking, among the low AP group, must be calculated.

The RR of smokers contracting bronchitis is:

$$RR = (34/211)/(45/1130) = 4.0$$

An RR of 4.0 means that smoking is a strong predictor of bronchitis.

b. Among the high AP group there are 1048/1307=80% smokers. Among the low AP group there are 211/1341=16% smokers. Thus, smoking is differentially distributed among the high and low AP exposure groups.

c. The crude RR (not stratified by smoking status) = Risk of bronchitis from high AP exposure / Risk of bronchitis from low AP exposure.

$$RR_{crude} = (178/1307)/(79/1341) = 2.3$$

The RR of contracting bronchitis among smokers with high AP exposure is:

$$RR = (168/1048)/(34/211) = 1.0. \text{ It appears as though high AP exposure doesn't affect the risk of bronchitis for smokers.}$$

The RR of contracting bronchitis among nonsmokers with high AP exposure is:

$RR=(10/259)/(45/1130)=1.0$. It also appears as though nonsmokers face no increased risk of bronchitis due to high AP exposure.

The crude RR of 2.3 and the smoking-specific RR of 4.0 are both above the stratified RR's of 1.0. Thus, the crude RR is confounded by smoking status.

Glossary

Confounding bias – A systematic distortion in the measure of association between exposure and disease caused by missing the effect of the exposure of primary interest with extraneous risk factors.

Special Note:

Beginning with this issue, the ERIC notebook will be published several times a year.

Look for our next issue: Part II of Confounding!

If you would like to receive ERIC Notebook please fill out the form below:

Name: _____

Degree(s): _____

Address: _____

City, State, Zip: _____

Telephone Number: _____

Fax Number: _____

E-mail Address: _____

Please fax to: 919-416-5836 – Attn: Beth Armstrong **Or**

Mail to : Beth Armstrong, ERIC Program Manager, VA Medical Center (152), 508 Fulton Street, Durham, NC 27705

BETH ARMSTRONG, ERIC PROGRAM MANAGER
VA MEDICAL CENTER (152)
508 FULTON STREET
DURHAM, NC 27705

