

ERIC Notebook

November/December 2000

Issue 15



Michel Ibrahim, MD, PhD
Director of Education
Program

Lorraine Alexander, DrPH

Carl Shy, MD, DrPH

Gayle Shimokura, MSPH

Sherry Farr, Graduate
Research Assistant

Sandra Deming, MPH,
Graduate Research Asst.

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 II

Calculating and adjusting for confounding

The previous issue of ERIC Notebook, "Confounding Bias, Part I", discussed two criteria for identifying confounders in a study. Criterion one is that the confounder must be a known, or suspected, risk factor for the disease of interest. Criterion two requires the confounder to be associated with the main exposure of interest, but not be along the causal pathway between exposure and disease. Once actual confounders have been identified, the next step is to evaluate how much the confounders bias the study results. To do this, an analysis where confounding is ignored, the "crude" measure of association, is compared to analyses that have been corrected for distortions due to confounding, the "adjusted" measure of association. Methods to calculate adjusted measures of associations differ by the need to control each confounder individually or all confounders simultaneously.

Before calculating an adjusted measure of association using stratified analyses, one must first assess the presence of effect modification. When effect modification is present, it can be difficult to ascertain whether or not confounding is occurring.

This methodology makes two assumptions:

- First, the data are obtained by simple random sampling rather than by some more restrictive subject selection procedure, like matching.
- The second assumption is that the exposure, disease, and confounder variables are all dichotomous (i.e., having only two strata). If the variables are in a continuous format, they can either be dichotomized, or they must be adjusted for simultaneously to calculate the true measure of association.

What is effect modification?

When risk estimates of an exposure-disease relationship stratified by a confounder are sufficiently different from one another (i.e., RR level 1=4.0 and RR level 2 =0.2), they suggest that two different exposure-disease relationships may be operating, one in each level of the confounder.

Two examples of effect modification are:

- A breast cancer education program (the exposure) that is much more effective in reducing breast cancer in rural areas than urban areas. Here, the area (rural or urban) is the effect modifier.
- The finding that a reduction in regional public transportation services (the exposure) affects individuals with little or no access to a car much more than those individuals with access to a car. In this example, having access to a car is the effect modifier.

Effect modification is different from confounding, where instead of "competing" with the exposure of interest in explaining the etiology of a disease, the effect modifier identifies subpopulations that are particularly susceptible to the exposure of interest.

Is effect modification present?

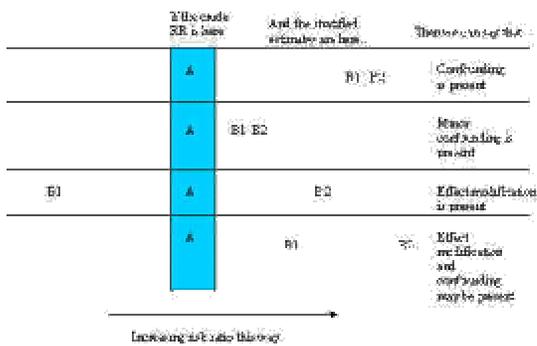
To calculate whether effect modification is playing a role in the study, first calculate three measures of association:

A = The overall, crude measure of association of the exposure-disease association.

B1 = The measure of the exposure-disease association among all study participants who have a history of the confounding variable (C+).

B2 = The measure of the exposure-disease association among all study participants who do not have a history of the confounding variable (C-).

Use the figure below as a guide on how to interpret the meaning of these three measures of associations. As a general rule, if B1 and B2 are basically equal in value, but different from A, then confounding is present and effect modification is not present. Effect modification is present when B1 and B2 are different from one another, and at least one (B1 or B2) is different from A. Both effect modification and confounding can occur simultaneously.



Note: In this figure, B1 and B2 are interchangeable.

If effect modification is present, confounding cannot be adjusted for because stratified analysis, the procedure for controlling for effect modification, is the same procedure needed to adjust for confounding. In this situation, more complex statistical techniques are needed, but will not be discussed in the scope of this issue.

Calculating adjusted summary estimates

If no effect modification is present, then stratum-specific estimated effects can be pooled to form a summary estimate of effect across strata. This summary estimate represents an adjusted risk ratio (a risk ratio adjusted for confounding).

Although there are many ways to calculate the adjusted RR, presented here is the Mantel Haenzel procedure, which is the most common pooling procedure.

For case control studies, the summary OR, using the Mantel Haenzel method, is calculated. Odds ratios from at least three strata (strata 1, strata 2, and a generic strata z) are pooled using a weighting scheme, “W”. The values for W come from the typical 2x2 table containing the four cells a,b,c, and d, and a total sum of n:

	Disease	No disease
Exposed	a	b
Unexposed	c	d

$$n=a+b+c+d$$

$$mOR = \frac{W_{strata1} \times OR_{strata1} + W_{strata2} \times OR_{strata2} + \dots + W_{strataz} \times OR_{strataz}}{W_{strata1} + W_{strata2} + \dots + W_{strataz}}$$

where

$$W_{strataz} = \frac{b_{strataz} \times c_{strataz}}{n_{strataz}}$$

This particular formula assumes that $(b_{strataz} \times c_{strataz}) / n_{strataz} \neq 0$ for all strata.

For prospective cohort studies where risk ratios are calculated, a similar formula applies:

$$mRR = \frac{W_{strata1} \times RR_{strata1} + W_{strata2} \times RR_{strata2} + \dots + W_{strataz} \times RR_{strataz}}{W_{strata1} + W_{strata2} + \dots + W_{strataz}}$$

where

$$W_{strataz} = \frac{b_{strataz} \times n_{exposed\ in\ strataz}}{n_{strataz}}$$

This particular formula assumes that $b_{strataz} \times n_{exposed\ in\ strataz} \neq 0$ for all strata.

Further details on pooling estimates across strata using the Mantel Haenzel procedure are described in Kleinbaum et al., Epidemiologic Research: Principles and Quantitative Methods, 1982, p.342-51.

Calculating adjusted measures when all confounders are assessed simultaneously

Simultaneous control of two or more variables can give different (and potentially more interesting) results from those obtained by controlling for each variable separately. Simultaneous control of confounders better emulates the natural environment where exposures, diseases, and confounders of interest are found, than does individual control of confounders.

Simultaneous control of several confounders to calculate adjusted measures is done through mathematical modeling.

To control for confounding using mathematical modeling, simply include the confounding variables as independent variables in the model. The simplicity of this method of adjustment for confounding is one of the attractive features of using mathematical models in epidemiology.

Although many types of mathematical models are available, there is generally only one type of model that is appropriate for the goals of that specific data analysis and for the type of data available. The most common mathematical model used in epidemiology is logistic regression.

This model has the general format of
 $y = a + b_1x + b_2z_2 + \dots + b_iz_i$.

Individual-level data needs to be provided on:

1. y : the disease outcome in a dichotomous format
2. x : the exposure
3. z_2 to z_i : the confounders

to the statistical software package (any one will do) and the computer will produce:

1. a : the y-axis intercept
2. b_1 : the coefficient for the exposure variable
3. b_2 to b_i : the coefficients for each confounder that is controlled in the model.

These coefficients (except for a) are very useful, as they can be transformed into odds ratios. The odds ratio obtained from b_1 of this model is an interpretable measure of association describing the relationship between the exposure (x) and disease (y) after adjustment for confounding variables (z_2 to z_i).

The biggest disadvantages to using mathematical models are the assumptions that must be met by the dataset in order to use them—often the data may not conform to all of them. It is advisable to do regression diagnostics sometime during the data analysis stage to check these assumptions.

Is adjustment for confounding necessary?

If the adjusted effects are markedly different from the crude effect (typically a 10 or 15% change from crude to adjusted), then confounding is present and should be controlled for. Report the cut-off used (5%, 10%, or 50%) in the selection of confounders for adjustment.

If the adjustment of confounding variables changes the results only slightly (less than 10%), then the tendency would be to ignore its influence, since the more variables controlled for, the less precise (the wider the confidence intervals) the study results will be. The benefits of ignoring the minor confounders would outweigh the costs.

Also consider whether it is important to control for potential confounders such as age, simply because many readers would not trust results that are not adjusted for age. This distrust stems from knowledge that age is strongly related to disease and mortality rates (similar comments would apply to sex).

Control of Confounding

A. In the analysis phase:

Once data has been collected, there are two options for control of confounding: Stratified analysis or mathematical modeling. Both methods were described above when calculating the effect of confounding on the measure of association. Briefly, stratified analysis pools the measure of association calculated in each stratum of the confounder into one summary estimate. Mathematical modeling uses a more complex approach, and makes more assumptions, than stratified analysis.

B. In the design phase:

Some confounders should be controlled for in the study design stage of a study, rather than in the analysis stage. It may be necessary to do this if the confounder is very strong and when the anticipated sample size will be large enough to deal with it in the design stage. Some study designs are more favorable for controlling for confounding than others.

Restriction, matching, and randomization are common techniques used to minimize confounding in the design phase. These techniques are not exclusive to one another. Several different control methods may be used at once.

1. Restriction

Confounding can be controlled for by restricting the study population to those who are unexposed to one or more confounding variables. An example of restriction is to restrict a study population to nonsmokers when studying the association of environmental radon with lung cancer. Restriction is ideal when the exposure-disease relationship has strong confounders, because it can be an efficient, convenient, inexpensive, and straight-forward method of controlling for confounding. However, the restricted variable, for instance smoking in the given example, cannot be assessed for confounding. Restriction may not always be logistically feasible because the sample size of available study participants is decreased, sometimes to the point that a study cannot be done.

2. Matching

Confounding can also be controlled through matching on the confounder variable(s). Matching involves constraining the control group (for case-control studies) or the unexposed group (for cohort studies) such that the distribution of the confounding variable(s) within these groups are similar (or identical) to the corresponding distribution within the index group (the case group for case-control studies or the exposed group for cohort studies). Matching can be viewed as imposing a "partial restriction" on the values of the confounding variables, since only the control or unexposed group is restricted.

- For instance, a group of 30 HIV-positive military recruits from various parts of the country (20% East Coast, 30% West Coast, 50% Central US) has been identified with which to study behavioral risk factors for HIV infection, independent of location. Therefore, the control group should be 30 HIV-negative military recruits with the same location distribution as the group of HIV-positive military recruits (20% East Coast, 30% West Coast, 50% Central US). This would be accomplished through matching the controls to the cases by location, by selecting only HIV-negative military recruits who contribute to the pre-determined location distribution.

Analysis of matched data requires special consideration, because the control, or unexposed, group is not a random sample of study participants; they should be considered to be a biased sample. Techniques for analyzing matched data include conducting the data analysis separately for each level of the confounder (stratified analysis) and using conditional logistic regression.

When considering matching, consider four factors:

1. Precision (generally increased with matching)
2. Cost (generally lowered with matching, because a smaller sample size is needed)
3. Feasibility (can be increased with matching)
4. Flexibility in deciding whether to match

Also keep in mind that variables matched cannot be assessed for confounding.

3. Randomization

Randomization is an ideal method for controlling for confounding because this method can control both known and unknown confounders. However, because randomization requires that the exposure status of individuals be assigned to study participants, observational study designs such as cross-sectional, cohort, case-control and ecological studies cannot use randomization to control for confounding. For controlled

clinical trials however, randomization is a common method to control for confounding.

Use of randomization for control for confounding presumes that random classification of individuals into x number of groups will produce an x number of groups that have an equal (or similar) distribution of confounders. For example, the theory of randomization says that given two randomly selected groups of students, each group will have an equal percentage of females, an equal percentage of individuals with white-colored shirts, an equal percentage of brown-eyed individuals, and so forth. Thus, randomization, if done correctly, will produce homogeneous groups of individuals. When an exposure is applied to one of these homogeneous groups, but not to the other, the only difference between the two groups is their exposure status. In this situation, confounding, the unequal distribution of a risk factor between exposed and non-exposed groups, cannot occur.

The key to proper control of confounding through randomization is having a sufficiently large sample size in each randomized group. Rothman and Greenland (1998) state that having at least 50 subjects, preferably 100 or more, will assure that potential confounders are equally distributed among each study group.

References

Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic Research: Principles and Quantitative Methods. Belmont, CA: Lifetime Learning Publications, 1982.

Miettinen OS, Cook EF. Confounding: essence and detection. American Journal of Epidemiology 114(4):593-603, 1981 Oct. Rothman KJ and Greenland S. Modern Epidemiology, 2nd edition. Philadelphia, PA: Lippincott-Raven, 1998.

Self-evaluation

Q1: A new drug to lower blood pressure is being tested. The results of the clinical trial are displayed below:

	Lowered BP	No change in BP	Total
Drug	70	30	100
Placebo	40	60	100

There is speculation, however, that the new drug may not be as effective in overweight individuals as in normal weight individuals. The results stratified by weight are presented below.

Among Overweight (BMI \geq 25)

	Lowered BP	No change in BP	Total
Drug	30	20	50
Placebo	20	30	50

Normal weight (BMI<25)

	Lowered BP	No change in BP	Total
Drug	40	10	50
Placebo	20	30	50

- a) Calculate the crude measure of association between the drug and blood pressure.
- b) Calculate the stratum-specific associations between the drug and blood pressure.
- c) Based on your results, assess whether effect modification, confounding, or both are present in this study.

Answers

1.a. $RR_{crude} = (70 \times 60) / (40 \times 30) = 3.5$

b. $RR_{overweight} = (30 \times 30) / (20 \times 20) = 2.3$

$RR_{normal} = (40 \times 30) / (20 \times 10) = 6.0$

c. Because the RR for overweight subjects is less than the crude RR and the RR for normal weight subjects is greater than the crude RR, weight is an effect modifier and not a confounder. Therefore, the study results must be presented stratified by weight, and cannot be pooled or adjusted for weight.

Glossary

Effect modification – a variation in the magnitude of a measure of exposure effect across levels of another variable

Randomization – random assignment of subjects to exposure categories

Matching – the selection of controls, or unexposed subjects, that are identical, or nearly so, to the cases, or exposed subjects, with respect to the distribution of one or more potentially confounding factors

From: Modern Epidemiology, Rothman KJ and Greenland S, 1998

THIRD ANNUAL EPIDEMIOLOGY SUMMER SESSION

The three Epidemiologic Research and Information Centers (ERIC) are pleased to announce June 4 through June 8, 2001 as the dates for the third annual Summer Session in Epidemiology. The 2001 Session is being hosted by the ERIC at Durham, NC and will be held at the University of North Carolina in Chapel Hill, NC. The 2001 Summer Session is open to administrators, clinicians, and researchers employed by the VA.

Further details regarding course topics and applications will be distributed this winter. For more information, call Beth Armstrong at 919-286-6936 or email betharmstrong@mindspring.com.

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

Upcoming Topics

THIRD ANNUAL EPIDEMIOLOGY SUMMER SESSION

The three Epidemiologic Research and Information Centers (ERIC) are pleased to announce June 4 through June 8, 2001 as the dates for the third annual Summer Session in Epidemiology. The 2001 Session is being hosted by the ERIC at Durham, NC and will be held at the University of North Carolina in Chapel Hill, NC. The 2001 Summer Session is open to administrators, clinicians, and researchers employed by the VA.

Further details regarding course topics and applications will be distributed this winter. For more information, call Beth Armstrong at 919-286-6936 or email betharmstrong@mindspring.com.



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