ERIC Notebook

December 1999 Issue 8

UNC SCHOOL OF PUBLIC HEALTH
DEPARTMENT OF EPIDEMIOLOGY

Michel Ibrahim, MD, PhD
Director of Education Program

Lorraine Alexander, DrPH

Carl Shy, MD, DrPH

Sherry Farr, Graduate Research Assistant

ERIC
EPIDEMIOLOGIC RESEARCH & INFORMATION CENTER
Durham, North Carolina

Ron Horner, PhD
ERIC Director

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.

Selection Bias

Selection bias is a distortion in a measure of association (rate ratio, risk ratio, or odds ratio). Selection bias usually occurs as a result of either using improper procedures for obtaining persons from the target population to become members of the study population or as a result of factors that influence continued participation of subjects in a study.

Selection bias occurs when the association between exposure and disease is different for those who complete a study compared with those who are in the target population, the overall population for which the measure of effect is being calculated and from which study members are selected.

- In a case-control study of smoking and chronic lung disease, the association of exposure with disease will tend to be weaker if controls are selected from a hospital population (because smoking causes many diseases resulting in hospitalization) than if controls are selected from the community.

In this example, hospital controls do not represent the prevalence of exposure (smoking) in the community from which cases of chronic lung disease arise. The exposure-disease association has been distorted by selection of hospital controls.

Sources of selection bias

Selective survival and losses to follow-up – After enrollment of subjects and collection of baseline data there is usually some loss to follow-up with some individuals dropping out of the study. This biases the study when the association between risk factor and disease differs in dropouts compared with study participants.

Volunteer and non-response bias – Individuals who volunteer for a study may possess different characteristics than the average individual in the target population. Individuals who do not respond to requests to be studied generally have different baseline characteristics than responders. Bias will be introduced if the association between exposure and disease differs between study volunteers and non-responders.

Hospital patient bias (Berkson's Bias) – Berkson's bias may occur when hospital controls are used in a case-control study. If the controls are hospitalized due to an exposure that is also related to the disease under study then the measure of effect may be weakened, i.e. biased towards the null hypothesis of no association.

Healthy worker effect – Generally, working individuals are healthier than individuals who are not working. Therefore, in occupational exposure studies, where cases (or exposed subjects) are workers, controls (or non-exposed subjects) should also be workers, otherwise the association between exposure and disease will tend to be biased towards the null.

Selection probabilities

Selection bias will occur as a result of the procedure used to select study participants when the selection probabilities of exposed and unexposed cases and controls from the target population are differential and not proportional. This can occur when exposure status influences selection.

Suppose in a study of asbestos exposure and lung cancer the exposure is distributed among the cases and controls in the target population as follows:

	Diseased	Nondiseased
Exposed	100	200
Unexposed	100	400

The true OR in the target population is (100 x 400) / (100 x 200) = 2.0.

If the selection probabilities for all the cells in the table are equal at 90%, the 2x2 table of selection probabilities would look like the following.

	Diseased	Nondiseased
Exposed	"=90%	(=90%
Unexposed	$=90%	)=90%

And the 2x2 table of individuals in the case-control study will look like the following.

	Diseased	Nondiseased
Exposed	100 x .90=90	200 x .90=180
Unexposed	100 x .90=90	400 x .90=360

OR = (90 x 360) / (90 x 180) = 2.0

Here we have an example with each cell in our study population containing the same proportion of subjects as the corresponding cell in the target population. 90% of each cell is sampled. In this case, selection bias does not exist. Alpha / beta = gamma / delta where each greek letter represents the selection probability of the corresponding cell in the 2x2 table.

If the selection probabilities are unequal, but still proportional (that is, alpha / beta = gamma / delta), we still do not observe any selection bias in our study. If the selection probability is 90% among the diseased individuals and the selection probability is 70% among the nondiseased individuals the resulting 2x2 table would look like the following.

	Diseased	Nondiseased
Exposed	100 x .90=90	200 x .70=140
Unexposed	100 x .90=90	400 x .70=280

OR = (90 x 280) / (90 x 140) = 2.0

If however the selection probabilities are unequal, and also nonproportional, then selection bias will occur. The following table shows how selection bias occurs when the selection probability for the unexposed controls is different than that of the other three groups of study members.

	Diseased	Nondiseased
Exposed	100 x .90=90	200 x .90=180
Unexposed	100 x .90=90	400 x .70=280

OR = (90 x 280) / (90 x 180) = 1.6

To assess the amount of confounding in this estimate one can divide the confounded OR by the true OR, (1.6 / 2.0 = 0.80). The estimated effect of exposure on disease status is biased downward toward the null value of one.

To accurately represent the target population, we need the selection odds for exposure among the diseased (alpha/beta) to be equal to the selection odds for exposure among the nondiseased (gamma/delta). If the selection odds are different, then selection bias will distort our study measure of effect from the "truth", which in this study is 2.0. Likewise, no selection bias will occur if the selection odds for disease among the exposed is equal to the selection odds for disease among the nonexposed, i.e. if alpha/gamma = beta/delta. The two previous statements can be combined into one general principle—No selection bias will occur if the cross product of the four selection probabilities is equal to one, i.e. (alpha x delta) / (beta x gamma) = 1.

Selection bias will occur in cohort studies if the rates of participation or the rates of loss to follow-up differ by both exposure and disease status. Although we seldom can know the exposure and disease status of nonrespondents or persons lost to follow-up, it is sometimes possible to obtain these data from an external source.

Self-evaluation

Researchers conducted a cohort study to examine the relationship between baseline obesity and the incidence of non-insulin dependent diabetes. Suppose that 40% of the 1000 study subjects who were obese at baseline at the start of the study were lost to follow-up and 20% of the 1000 subjects who were non-obese at baseline were similarly lost during follow up. After 10 years, the cumulative incidence of diabetes was 120 cases in the remaining baseline obese group, and 88 in the remaining baseline nonobese group (call these two groups the "initial respondents").

Q1: Create a 2x2 table for exposed and unexposed cases and controls of the initial respondents and compute the cumulative incidence (CI) of diabetes in each exposure group and the relative risk of diabetes due to obesity.

Q2: A random sample of the subjects lost to follow up was obtained. The sample yielded an estimate of a 10% cumulative incidence of diabetes in the obese nonrespondents, and 5% in the non-obese nonrespondents. Assume that the latter results are a valid estimate of the diabetes risk in all nonrespondents. Note that this cumulative incidence is lower than that in the initial respondents. Create a 2x2 table with the new data on non-respondents combined with the initial respondents' data. Is there selection bias in this study?

Answers

1. Initial respondents

	Diabetes	No diabetes	Total
Obese	120	480	600
Non-obese	88	712	800
Total	208	1192	1400

Obese CI = 120 / 600 = 0.20
Non-obese CI = 88 / 800 = 0.11
RR = 0.20 / 0.11 = 1.82

2.
Obese nonrespondents: $400 \times 0.10 = 40$ cases
400 - 40 = 360 noncases
Non-obese nonrespondents: $200 \times 0.05 = 10$ cases
200 - 10 = 190 noncases

	Obese	Non-obese	Total
Diabetes	120+40= 160	88+10=98	258
No diabetes	480+360=840	712+190=902	1742
Total	1000	1000	2000

Obese CI: 160 / 1000 = 1.6
Non-obese CI: 98 / 1000 = 0.98
RR = 1.6 / 0.98 = 1.63

This study apparently suffers from selection bias because 1) the loss to follow-up was greater in the exposed (obese) than nonexposed cohort and 2)the proportion of persons becoming diseased differed between initial respondents and nonrespondents in both the exposed and nonexposed cohorts. The relative risk of 1.8 computed from the initial respondents is not a valid estimate of the association between obesity and risk of diabetes in the original starting population of 2000 persons.

Glossary

Bias: a systematic error in a study that leads to a distortion of the results. (Medical Epidemiology, Greenberg RS, 1993).

Target population: the overall population for which the measure of effect is being calculated, and from which study members are selected.

ERIC NOTEBOOK IS PRODUCED BY THE EDUCATIONAL ARM (MICHEL A. IBRAHIM, MD, PHD, DIRECTOR) OF THE EPIDEMIOLOGIC RESEARCH AND INFORMATION CENTER AT DURHAM, NC (RON HORNER, PHD, DIRECTOR)

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: 919-966-2089 – Attn: Sherry Farr

Mail to : UNC-CH , School of Public Health , Attn: Angie Tuck
Department of Epidemiology , Campus Box 7400, Chapel Hill, NC

Upcoming Topics

Cumulative Incidence
Randomization

Please let Sherry Farr know which topics are of special interest to you so that we can include them in a future issue.

ERIC Notebook

UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
SCHOOL OF PUBLIC HEALTH
DEPARTMENT OF EPIDEMIOLOGY
CAMPUS BOX 7400
CHAPEL HILL, NC 27599