

## Selection Bias

### Second Edition Authors:

Lorraine K. Alexander, DrPH

Brettania Lopes, MPH

Kristen Ricchetti-Masterson, MSPH

Karin B. Yeatts, PhD, MS

Selection bias is a distortion in a measure of association (such as a risk ratio) due to a sample selection that does not accurately reflect the target population. Selection bias can occur when investigators use improper procedures for selecting a sample population, but it can also occur as a result of factors that influence continued participation of subjects in a study. In either case, the final study population is not representative of the target population – the overall population for which the measure of effect is being calculated and from which study members are selected.

Selection bias occurs when the association between exposure and health outcome is different for those who complete a study compared with those who are in the target population.

### Example

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, i.e. when individuals leave the study before the end of follow-up. This biases the study when the association between a risk factor and a health outcome 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 a health

outcome 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 health outcome 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 unexposed subjects) should also be workers, otherwise the association between exposure and the health outcome will tend to be biased towards the null.

#### **Selection Bias**

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.

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 health outcome status. Although we seldom can know the exposure and health outcome status of non-respondents or persons lost to follow-up, it is sometimes possible to obtain these data from an external source.

#### **Terminology**

*Bias*: a systematic error in a study that leads to a distortion of the results. ( *Target population*: the overall population for which the measure of effect is being calculated, and from which study members are selected.

*Loss to follow-up*: when individuals leave the study before the end of follow-up.

Medical Epidemiology, Greenberg RS, 1993).

#### **Practice Questions**

*Answers are at the end of this notebook*

- 1) Researchers are planning to conduct a case-control study of the association between an occupational exposure and a health outcome. The researchers plan to study exposed workers from one factory and compare them with unexposed retirees who have never worked in a factory. A reviewer of the research proposal is worried about selection bias and in particular about the possibility of the healthy worker effect. Which of the following best represents the reviewer's concern?
  - a) Retirees should not be compared to factory workers because factory workers are under more stress than retirees
  - b) Retirees should not be compared to factory workers because factory workers' incomes differ from those of retirees
  - c) Retirees should not be compared to factory workers because factory workers are likely to need to maintain a certain level of health in order to work in a factory while retirees would not necessarily be as healthy
  - d) Retirees should not be compared to factory workers because factory workers likely live in a different city than the retirees
- 2) Researchers conducted a prospective cohort study of the association between air pollution exposure and asthma. Some study participants were lost to follow-up (dropped out of the study) over time. The researchers were able to obtain data on the exposure and the health outcome for participants who remained in the study as well as for participants who dropped out of the study. The researchers discovered that the rate of loss to follow-up did not differ when comparing exposed and unexposed groups. The researchers also found that the rate of loss to follow-up did not differ when comparing people who developed asthma and people who did not develop

asthma. Based on this information, which one of the following statements is most likely to be true?

- a) Selection bias likely occurred in this study because both exposure groups experienced loss to follow-up
- b) Selection bias likely did not occur in this study because exposure status and health outcome status did not influence whether or not people dropped out of the study
- c) Selection bias likely occurred in this study because both of the outcome groups (people with asthma and people without asthma) experienced loss to follow-up
- d) Selection bias likely did not occur in this study because people cannot choose if they are exposed to air pollution or not exposed to air pollution

### References

Dr. Carl M. Shy, Epidemiology 160/600 Introduction to Epidemiology for Public Health course lectures, 1994-2001, The University of North Carolina at Chapel Hill, Department of Epidemiology

Rothman KJ, Greenland S. Modern Epidemiology. Second Edition. Philadelphia: Lippincott Williams and Wilkins, 1998.

The University of North Carolina at Chapel Hill, Department of Epidemiology Courses: Epidemiology 710, Fundamentals of Epidemiology course lectures, 2009-2013, and Epidemiology 718, Epidemiologic Analysis of Binary Data course lectures, 2009-2013.

### Acknowledgement

The authors of the Second Edition of the ERIC Notebook would like to acknowledge the authors of the ERIC Notebook, First Edition: Michel Ibrahim, MD, PhD, Lorraine Alexander, DrPH, Carl Shy, MD, DrPH, Gayle Shimokura, MSPH and Sherry Farr, GRA, Department of Epidemiology at the University of North Carolina at Chapel Hill. The First Edition of the ERIC Notebook was produced by the Educational Arm of the Epidemiologic Research and Information Center at Durham, NC. The funding for the ERIC Notebook First Edition was provided 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.

### Answers to Practice Questions

1. Answer choice c is correct. The healthy worker effect is a type of selection bias that may occur in occupational exposure studies when the exposed cases are workers but the non-exposed study participants (controls) are not workers. In general, working individuals are healthier than non-working individuals. Health problems may actually be a reason for not working. In addition, retirees are typically older than the working population and may have more age-related health problems.
2. Answer choice b is correct. Selection bias likely did not occur in this study because exposure status did not influence whether or not people dropped out of the study. Furthermore, the health outcome status did not influence whether or not people dropped out of the study. Remember that selection bias may occur in a cohort study if the rate of participation or the rate of loss to follow-up differ by both exposure and health outcome status. Selection bias is not affected by if the exposure is an avoidable exposure or a non-avoidable exposure.