

[Welcome](#) [Module Index](#) [Glossary](#) [Epiville Chamber of Commerce](#) [About this site](#) [Requirements](#)

Bias

[Introduction](#)

[Learning Objectives](#)

[Student Role](#)

[Study Design](#)

[Data Analysis Questions](#)

[Discussion Questions](#)

[Print Module](#)

Introduction

Bias is defined as a systematic error which results in an incorrect or invalid estimate of the measure of association, and addressing bias in [observational studies](#) is a central goal of the field of epidemiology. The following exercise will introduce several different types of bias, discuss various ways of minimizing or eliminating it, as well as possible impact on study results.

Faculty highlight: Steven D. Stellman, PhD



Dr. Steven Stellman's research has included studies of tobacco-related cancers, dietary risk factors for cancer, and environmental factors in breast cancer. More recently, he has applied geographic information systems (GIS) in the studies of physical and mental health of Vietnam veterans in relation to exposure to Agent Orange and military combat, and served as the Research Director for The World Trade Center Health Registry through the NYC Department of Health.

For this module we will focus on a recent paper published by a team of investigators involved with the Long Island Breast Cancer Study Project, examining the association between home pesticide use and breast cancer. In particular, we will focus on the potential association between lawn/garden pesticide use and risk of breast cancer in an article by the study team led by Dr. Stellman and Dr. Susan Teitelbaum, a former student in our Department.

Read more about Dr. Stellman's work in the following articles

- [Muscat JE, Takezaki T, Tajima K, Stellman SD. Charcoal cigarette filters and lung cancer risk in Aichi Prefecture, Japan . Cancer Sci. 2005 May;96\(5\):283-7.](#)
- [Muscat JE, Britton JA, Djordjevic MV, Citron ML, Kemeny M, Busch-Devereaux E, Pittman B, Stellman SD. Adipose concentrations of organochlorine compounds and breast cancer recurrence in Long Island, New York. Cancer Epidemiology Biomarkers Prev. 2003 Dec;12\(12\):1474-8.](#)
- [Stellman JM, Stellman SD, Christian R, Weber T, Tomasallo C. The extent and patterns of usage of Agent Orange and other herbicides in Vietnam. Nature. 2003 Apr 17;422\(6933\):681-7.](#)

Good luck and have fun!

[Next »](#)



[Welcome](#) [Module Index](#) [Glossary](#) [EpiVille Chamber of Commerce](#) [About this site](#) [Requirements](#)

Bias

[Introduction](#)
[Learning Objectives](#)
[Student Role](#)
[Study Design](#)
[Data Analysis Questions](#)
[Discussion Questions](#)
[Print Module](#)

Learning Objectives

A. Describe key features of the two types of bias and give examples of each:

1. Selection Bias
 - Define [selection bias](#)
 - Describe types of selection bias
 1. Recall bias
 2. Interviewer bias
 3. Loss to follow-up
 4. Differential and non-differential misclassification
1. Observation/Information Bias:
 - Define [observation bias](#)
 - Describe types of observation bias
 1. Control-selection bias
 2. Self-selection bias
 3. Differential surveillance, diagnosis or referral bias

B. Identify types of biases specific for different study designs:

1. [Randomized clinical trials \(RCT\)](#)
2. [Case-control studies](#)
3. [Cohort studies](#)

C. Understand how the magnitude and direction of bias can affect study results

D. Describe ways to minimize bias in the design phase of a study:

1. Develop an explicit case definition
2. Enroll all cases in a defined time and region from a defined source population
3. Strive for high participation rate (minimize loss to follow-up and drop-out rate)
4. Take precautions to ensure your sample is a good representation of the underlying population
5. Handling of misclassification (observer bias, measurement bias)
6. Other methods

[« Previous](#) || [Next »](#)



[Welcome](#) [Module Index](#) [Glossary](#) [Epiville Chamber of Commerce](#) [About this site](#) [Requirements](#)

Bias

[Introduction](#)
[Learning Objectives](#)
[Student Role](#)
[Study Design](#)
[Data Analysis Questions](#)
[Discussion Questions](#)
[Print Module](#)

Student Role

Reading and understanding scientific articles takes a critical mind. You must carefully evaluate the methodologies employed and conclusions that are formed. At the end of the semester you will be asked to critique a published article. To start you on this path, below we illustrate an example of some of the core elements of a critique of a recently published paper on the association of self-reported residential pesticide use and breast cancer by Teitelbaum et al. (2007). Pay particular attention to the areas that are highlighted. If you have any questions, do not hesitate to discuss them with your seminar leader.

Synopsis

[Susan L. Teitelbaum, Marilie D. Gammon, Julie A. Britton, Alfred I. Neugut, Bruce Levin, and Steven D. Stellman \(2007\). Reported Residential Pesticide Use and Breast Cancer Risk on Long Island. New York, American Journal of Epidemiology, 165:643-651.](#)

Study Aim

To "investigate whether self-reported lifetime residential organochloride pesticide use increases breast cancer risk among women living on Long Island, NY between August 1996 to July 1997."

Null Hypothesis

Organochloride pesticides are not associated with breast cancer risk, i.e. an odds ratio of 1.0 comparing cases to controls on the probability of organochloride pesticides use.

Alternative Hypothesis

Organochloride pesticides are associated with increased breast cancer risk, i.e. an odds ratio greater than 1.0 comparing cases to controls on the probability of self-reported organochloride pesticides use.

Study Design

Type of study

The Long Island Breast Cancer Study Project is a population-based case-control study.

Study population

The study population consisted of all adult female residents of either Nassau or Suffolk Counties, residing in these areas between August 1, 1996 and July 31, 1997.

Method of selection/sampling of subjects:

Sources of data

Cases were identified through a 'super-rapid' identification network. This network "was established to ascertain potentially eligible case women with newly diagnosed breast cancer" from all hospitals in Long Island and three large tertiary care hospitals in New York City. Cases were included if their primary residence at the time of diagnosis (between August 1, 1996 and July 31, 1997) was in Nassau or Suffolk Counties of New York State. Controls were ascertained in two ways: among those <65 years of age, women were recruited through random digit dialing. Among those 65 years of age and older, women were recruited through Medicare and Medicaid rosters. Women were eligible to be controls if they did not have a lifetime history of breast cancer and lived in the two catchment counties between August 1, 1996 and July 31, 1997. Controls were frequency matched to cases at 5-year age intervals. Of the eligible cases, 74.3% completed the interview. Of the eligible controls, 62.8% completed the interview.

Measurements of exposures

Information included in the present study was collected through in-person structured questionnaires. Respondents self-reported lifetime exposure history to a vast array of potential chemical hazards that may increase risk for breast cancer. Questions elicited detailed information about the types of pesticides used, the duration and amount of use, the proximity of the respondent to the chemical, personal application of the pesticide (versus family member, etc.), and location of the pesticide use (home, lawn, garden, etc.). Questionnaires were administered by trained interviewers. Overall pesticide use (the sum of lifetime applications of all 15 categories), the two combined groups (lawn and garden and nuisance pest), and each of the 15 individual categories were considered in the analyses. Lifetime applications were categorized based on the control distribution.

Measurement of outcomes

Cases were women newly diagnosed with a first primary in situ or invasive breast cancer confirmed by the physician and the medical record. Controls were confirmed as not having had a history of physician-diagnosed breast cancer.

Measurement of potential confounders

To eliminate potential confounding effects of age, controls were frequency-matched to cases in the design stage of the study. Geographic location was controlled by restriction; only women residing in two specific counties were included for participation. Other potential confounders were measured via in-person structured questionnaires. These potential confounders included: "race, marital status, religion, household income, age at menarche, parity, age at first birth, lactation, menopausal status, oral contraceptive use, hormone replacement therapy use, first-degree family history of breast cancer, history of benign breast disease, body mass index (weight in kilograms divided by height in meters squared) at the reference age and at age 20 years, alcohol use, smoking status, and physical activity." Finally, highest level of educational attainment was included to eliminate potential confounding by socioeconomic status.

Findings

For the present module we will focus on the association between lawn/garden pesticide use and breast cancer. The numbers below were extracted from Table 1 of Teitelbaum et al. (2007). We use these numbers to calculate a crude (i.e., unadjusted) odds ratio.

	Breast cancer +	Breast cancer -	Total
Ever used lawn/garden pesticides	1,254	1,231	2,485
Never used lawn/garden pesticides	240	305	545
Total	1,494	1,536	3,030

$$OR = (1254/240)/(1231/305) = 1.29$$

The odds of breast cancer among those who have ever used lawn/garden pesticides is 1.29 times that of those who have never used lawn/garden pesticides. Alternatively, the odds of ever using lawn/garden pesticides among cases is 1.29 times that among controls.

When controlled for age and educational status, the association between lawn/garden pesticide use and breast cancer was 1.34 (Table 1 in the text). This indicates that the odds of breast cancer among those who have ever used lawn/garden pesticides is 1.34 times that of those who have never used lawn/garden pesticides. The 95% confidence interval for this odds ratio ranged from 1.11 to 1.63, a fairly narrow interval indicating that the estimate has good precision. Furthermore, because the null hypothesis value of 1.0 is not included in the confidence interval, we can conclude that the evidence presented here does not refute the hypothesis that lawn/garden pesticide use is associated with breast cancer.

[« Previous](#) || [Next »](#)



[Welcome](#) [Module Index](#) [Glossary](#) [Epiville Chamber of Commerce](#) [About this site](#) [Requirements](#)

Bias

[Introduction](#)
[Learning Objectives](#)
[Student Role](#)
[Study Design](#)
[Data Analysis Questions](#)
[Discussion Question](#)
[Print Module](#)

Study Design

1. Investigators recruited both cases and controls from a defined catchment area in the general population. This is often difficult to do in the absence of a comprehensive case registry. Suppose that investigators had recruited all cases of breast cancer from a comprehensive national registry, between August 1, 1996 and July 31, 1997. What type of bias might be introduced if investigators kept the same set of controls as were reported in the Teitelbaum et al. (2007) study?

- a. Selection bias
- b. Recall bias
- c. Self-selection bias

Interactive exercise: Hypothetical selection of cases and controls for the hospital-based case-control study investigating the potential association between lawn/garden pesticide use and breast cancer to show selection procedures and their effects.

2. Suppose that you are designing a case-control study on the association between lawn/garden pesticide use and breast cancer using subject selection methods described in

the interactive module. Which method of accruing cases and controls do you think is the most practical?

- a. Situation 1: Cases were recruited at the Epiville General Hospital, and since it is a tertiary-care facility, they came from different areas, including outside of Epiville. Controls are selected from those attending the Fancypants™ weight loss center at the Star Hospital.
- b. Situation 2: Cases were recruited at the Epiville General Hospital and came from different areas, including outside Epiville. Controls are selected from the admissions office of the Epiville General Hospital with diagnoses of any cancer other than breast cancer that are not related to lawn/garden pesticide use or with diagnoses of non-cancer diseases.
- c. Situation 3: Cases were recruited at the Epiville General Hospital and came from different areas, including outside Epiville. Controls are selected from those attending the free clinic at the Epiville General Hospital

3. What potential bias could have been introduced if you found out that those who interviewed cases took 30 minutes longer on average than those who interviewed controls?

- a. Selection bias
- b. Information bias
- c. Volunteer bias
- d. Loss to follow-up

4. Teitelbaum et al. (2007) reported a significant association between lifetime use of lawn/garden pesticides and breast cancer compared to no lifetime use of lawn/garden pesticides. The duration of exposure varied among both cases and controls. For example, within the cases, some individuals may have reported being exposed more than 5 or 10 years prior to diagnosis, whereas others may have reported current exposure to these pesticides. What potential problems could arise when trying to measure exposures that happened over different time periods?

- a. Misclassification bias
- b. Volunteer bias
- c. Surveillance bias

5. What effect (if any) would you expect if the interviewers were aware of the disease status of the study subjects?

- a. It would benefit the validity of the results since the interviewers would understand more precisely how the exposure is related to the disease and collect better data for the cases.
- b. The results would likely not change.
- c. It could damage the validity of the results by introducing interviewer bias.

[« Previous](#) || [Next »](#)



[Welcome](#) [Module Index](#) [Glossary](#) [Epiville Chamber of Commerce](#) [About this site](#) [Requirements](#)

Bias

[Introduction](#)
[Learning Objectives](#)
[Student Role](#)
[Study Design](#)
[Data Analysis Questions](#)
[Discussion Questions](#)
[Print Module](#)

Data Analysis Questions

In the examples you have just completed, you were faced with the major types of bias. There are many different ways of categorizing biases. We choose to classify biases based on two general schemes: whether the bias is caused by selection of individuals into the study, or by the information obtained from the participants during the study (selection vs. information bias), and whether the resulting bias is likely to be non-differential or differential.

Selection Bias

An error due to selection of cases and controls based on differing criteria that are related to exposure status, or selection (or follow-up) of exposed and unexposed individuals in a way that is related to the development of the outcome. Self-selection and volunteer bias, among others, fall under the category of selection bias.

Information or Observation Bias

Bias arising from errors in exposure or disease classification (also known as measurement error). Recall bias is the most common form of information bias.

Information bias can manifest in two ways:

Non-differential:

- If misclassification of exposure is UNRELATED to disease
- If misclassification of disease is UNRELATED to exposure
- Effect = Bias towards the null (OR and RR closer to 1.0)

Differential:

- If misclassification of exposure is RELATED to disease
- If misclassification of disease is RELATED to exposure
- Effect = Bias can go in either direction from the null; it can inflate or attenuate your effect estimates (OR and RR)

Understanding the effects of differential and non-differential misclassification are among the most subtle and challenging aspects of studying bias. For some examples of these types of misclassification, work through the following scenarios.

Examples of Misclassification

Please Note: Everything that follows will be based on the 2x2 table of the association between lawn/garden pesticide use and breast cancer reported in Teitelbaum et al. (2007). For the purpose of this example, we are going to assume that the effect measure reported in the study is the "truth". That is, there was no misclassification in the Teitelbaum et al. (2007) study. Obviously, we never know the "truth" in epidemiology, and certainly at least some misclassification is likely in the measures reported in Teitelbaum et al. (2007). But for the purpose of the exercise, let's assume that exposure and disease were measured perfectly. Disease status (breast cancer) will be labeled as D+ (breast cancer) and D- (no breast cancer). Exposure to lawn/garden pesticides will be labeled as E+ and E-.

The "Truth"			
	Breast cancer +	Breast cancer -	Total
Ever used lawn/garden pesticides	1,254	1,231	2,485
Never used lawn/garden pesticides	240	305	545
Total	1,494	1,536	3,030

Let us assume that the "true" estimate of the association between lawn/garden pesticide use and breast cancer is $OR = (1254/240)/(1231/305) = 1.29$. Let's see how different types of misclassification can bias the estimate.

Scenario #1:

Some of the cases were confused about how you defined lawn/garden pesticide use. Because they were extraordinarily concerned with finding the cause of their breast cancer, 10 of the cases reported that they had been exposed when in fact they had not been exposed. Therefore, among those with breast cancer, you have misclassified 10 people who are in "truth" unexposed as exposed.

		Misclassification	
		D+	D-
E+			
E-	10		

6. What is the effect of this misclassification on the odds ratio?

Check your answers here.

Scenario #2:

Now suppose that 10 of the cases forgot about their lawn/garden pesticide exposure, while all of the controls still remembered perfectly. Therefore, among those with disease, you have misclassified 10 people who in "truth" were exposed as unexposed.

Misclassification

	D+	D-
E+	10	
E-		

7. What is the effect of this misclassification on the odds ratio?

Check your answers here.

Scenario #3:

When compared with the "truth", you have misclassified 10% of your D+E+ ('a' cell: people exposed to lawn/garden pesticides **with** a diagnosis of breast cancer) as D+E- ('c' cell: people not exposed to lawn/garden pesticide with a diagnosis of breast cancer), and misclassified 40% of your D-E+ ('b' cell: people exposed to lawn/garden pesticides **without** a diagnosis of breast cancer) as D-E- ('d' cell: people unexposed to lawn/garden pesticides **without** a diagnosis of breast cancer).

This may happen in a case control study when both cases and controls under-recall their exposure status, but controls under-recall at a greater proportion than cases. This is common in case control studies; cases have likely been giving more thought to the potential exposures that may have caused their diagnosis.

Misclassification

	D+	D-
E+	10%	40%
E-		

8. What is the effect of this misclassification on the odds ratio?

Check your answers here.

Scenario #4:

When compared with the "truth" you have misclassified 15% of ALL persons exposed to lawn/garden pesticide as non-exposed, and you have misclassified 10% of ALL non-exposed as exposed.

This could have happened in the Teitelbaum et al. (2007) study. For instance, one common lawn/garden pesticide was left off of the questionnaire, the interviewers did not ask about it and all those exposed to this particular pesticide were classified as unexposed. Thus, the exposure would not be based on diseased status - all individuals were not asked the question.

Misclassification

	D+	D-
E+	15%	15%
E-	10%	10%

9. What is the effect of this misclassification on the odds ratio?

Check your answers here.

Scenario #5:

When compared with the "truth" you have misclassified 50% of ALL individuals truly diagnosed with breast cancer as non-diseased, and you have misclassified 30% of ALL truly non-diseased as having been diagnosed with breast cancer. For misclassification this serious to occur, you would have used a very bad measure of breast cancer in your study!

Misclassification

	D+	D-
E+	50%	30%
E-	50%	30%

10. What is the effect of this misclassification on the odds ratio?

Check your answers here.

[« Previous](#) || [Next »](#)



Bias

- Introduction
- Learning Objectives
- Student Role
- Study Design
- Data Analysis Questions
- Discussion Questions**
- Print Module

Discussion Questions

Carefully consider the following questions. Write down your answers (1 - 2 paragraphs) for question # 1 within a word document and submit your answers to your seminar leader. Be prepared to discuss all questions during the seminar section.

- What are some strategies in the study design and data collection phase to minimize misclassification?
- The study that we read this week used population-based controls. More often in case-control studies, cases are recruited from a specific hospital and controls are selected from individuals without the outcome of interest admitted to the same hospital. Discuss some of the strengths and limitations of using hospital controls in case-control studies. Would using friend controls (i.e., asking the cases to nominate a friend to serve as a control) be a more appropriate choice than hospital controls?
- In the study of pesticide use and breast cancer, researchers evaluated pesticide use during the entire life of the respondents. How would you expect the results to have been affected if the researchers collected exposure history spanning the previous 20 years only?

Questions for the Intellectually Curious:

Think about what you know about p-values and 95% confidence intervals. A traditional interpretation of a 95% confidence interval is that "we are 95% confident that the true effect estimates lies within these bounds." Do 95% confidence intervals incorporate any information about bias and misclassification?

[« Previous](#) |

