

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

Cohort Study

[Introduction](#)

[Learning Objectives](#)

[Student Role](#)

[Study Design](#)

[Data Collection](#)

[Data Analysis](#)

[Discussion Question](#)

[Print Module](#)

Introduction

Methodologically, the [cohort study](#) best approximates the randomized controlled trial (RCT), it samples on exposure and follows study participants through time for the development of the outcome of interest. As such, the cohort study is considered one of the leading observational epidemiological study designs. The following exercise will show you the key features of the cohort design.

Good luck and have fun!

Faculty Highlight: Geoffrey R. Howe, PhD (1942-2006)



Dr. Geoffrey Howe was Professor of Epidemiology at the Mailman School of Public Health of Columbia University. He was Director of the Epidemiology Unit of the National Cancer Institute of Canada.

Dr. Howe's research activities have focused primarily in the areas of the relationship between radiation and cancer and other diseases, as well as nutritional epidemiology, particularly as it relates to cancer risk. He was also the Principal Investigator for an NCI-funded contract that provides scientific support for three studies of the consequences of the Chernobyl accident in 1986.

Read more about Dr. Howe's work

1. [Howe GR, Zablotska LB, Fix JJ, Egel J, Buchanan J. Analysis of the Mortality Experience amongst U.S. Nuclear Power Industry Workers after Chronic Low-Dose Exposure to Ionizing Radiation. Radiation Research 162, 517-526. 2004.](#)
2. [Silvera S, Rohan TE, Jian M, Terry PD, Howe GR, Miller AB. Glycemic index, glycemic load, and pancreatic cancer risk. Cancer Causes and Control. 16:431-436. 2005.](#)

[Next »](#)



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

Cohort Study

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

Learning Objectives

- A. Learn to apply the main features of a [cohort study](#)
 1. Formulate a research hypothesis
 2. Define the ['at risk' population](#)
 3. Define [eligibility criteria](#) for study participants
 4. Define exposure and [outcome](#)
 5. Understand cohort studies based on the timing of events
 - [Prospective](#)
 - [Retrospective](#)
- B. Conduct data analyses consistent with a cohort design
 1. Calculate and interpret the [risk ratio](#) using frequencies
 2. Calculate and interpret the [rate ratio](#) using [person-time](#) information
 3. Calculate and interpret the [standardized incidence ratio](#) (SIR)
- C. Interpret your findings
 1. Understand the differences between a risk ratio and a rate ratio
 2. Discuss how subgroup analyses influence our certainty about study results
 3. Discuss the value of age standardization

[« Previous](#) || [Next »](#)



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

Cohort Study

- Introduction
- Learning Objectives**
- Student Role**
- Study Design
- Data Collection

Student Role

Your internship at the EpiVillage Department of Health is progressing well. Your supervisor, Dr. Morissa Zapp, has been called in to investigate a sudden increase in [Susser Syndrome cases](#). Her goal is to design and conduct a cohort study to investigate the possible causes of Susser Syndrome. She calls

Data Analysis**Discussion Question****Print Module**

you into her office and asks you to do some preliminary investigative work and to report back to her with the results. You return to your office and immediately turn to the WEPI1 website to see if there are any news reports about the outbreak. Luckily you log on just in time to read through Stew O'Neil's report.

"Good afternoon, I'm Stew O'Neil with WEPI1 News... Doctors at the Epiville General Hospital report what appears to be a dramatic increase in the number of patients suffering from a cluster of neurological symptoms consistent with a diagnosis of Susser Syndrome, a rare and debilitating disease. Since March of this year, the number of diagnosed cases has been dramatically rising and health officials are concerned. We spoke with hospital officials earlier today and the cause of this increase in the number of Susser Syndrome cases is unknown.



"Doctors report that the disease pathway is poorly understood. Medical experts, however, believe that its occurrence is linked to an environmental exposure and may lead to permanent structural damage in the brain."

Channel 1, in an exclusive report, has uncovered that a significant number of diagnosed individuals are employed by Glop Industries, the manufacturing and production giant. Glop Industries is the largest employer in Epiville and the manufacturer of the antibacterial cleaning solution, SUPERCLEAN. When asked by Channel 1 reporters about the number of employees being diagnosed with Susser Syndrome, representatives from Glop had no comment.

Based on your own research and the newscast, you decide to investigate Glop Industries. The [Epiville Chamber of Commerce](#) website provides you with the following information:

- Information about Susser Syndrome from the Epiville Department of Health Webpage
- Information about Glop Industries which produces the suspected causal agent, SUPERCLEAN

Your first stop is at Glop Industries, located in the Epiville Industrial Park. Upon entering, you flash the powerful Epiville Department of Health Identification Card (carefully covering the word "intern" with your thumb), and ask to speak with the plant manager. You are immediately greeted by the plant manager, Ms. Dolores Doll, who is very responsive to your questions.

"Over the years, Glop Industries has produced more than 30 products, ranging from household cleaning supplies to pre-packaged frozen dinners to soft-drinks. Now, most of the plant space is being used to produce SUPERCLEAN."

"We produce over 1000 gallons a week. It can be used to clean dishes, floors, clothing -- I even use it in the shower, and I'll tell you, I've never felt so clean. We are still trying to fine-tune the production line to keep up with demand and so, right now, we do have a bit of spillage -- we lose about 5 percent of the product that way. At any rate, a little SUPERCLEAN in the air never hurt anyone - it keeps us feeling clean and the factory smelling great."



« Previous || Next »



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

Cohort Study

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

Study Design

From your studies, you know that the first step in any research plan is to generate a solid hypothesis to guide the investigation. So, while the WEPI1 newscast and Ms. Doll provided you with interesting information, you decide to visit a local hospital to inquire about individuals with Susser Syndrome.

Upon initial review of the cases, it does appear that a number of affected individuals did, in fact, work at the Glop Industries manufacturing plant. However, other individuals not associated with the factory have been affected as well (albeit in smaller numbers). You learn that Glop industries keeps meticulous information about its employees' work histories, which you decide to use in your study. With this exposure information from the employee records, you want to conduct a cohort study. Since both the exposure and [outcome](#) have already occurred, and since you have access to the exposure data collected prior to the disease outcome, you decide to design a [retrospective cohort study](#) (Please see Aschengrau & Seage pp. 147, and 206-208 for more information).

1. Based on the facts as presented, which do you think is the best hypothesis to investigate in this retrospective cohort study?

- a. Those who develop Susser Syndrome are more likely to have participated in the manufacturing of SUPERCLEAN than those who did not develop Susser Syndrome.
- b. Those who are exposed to chemicals involved in the production of SUPERCLEAN (via direct exposure at the factory) have a higher risk of developing Susser Syndrome than those who are not exposed.
- c. Residents of Epiville have a higher risk of developing Susser Syndrome compared with the residents of a neighboring community.

2. Based on your hypothesis, what would be the best way to define exposure?

- a. Provide all workers at the Glop Industries manufacturing plant with individual air quality instruments to take daily readings in order to compile weekly doses of exposure to SUPERCLEAN.
- b. Ask workers about their professional activities at the factory and estimate their exposure to SUPERCLEAN.
- c. Look for sources of information at the factory which record individual worker exposures throughout their employment.

Your supervisor assembles a team to begin the investigation. After a little groundwork, you find that the employee health clinic at Glop Industries keeps records of annual medical examinations for all

employees beginning with their hiring date. You also learn that the factory's human resources department has records of each worker's employment history which you can use to determine exposure to chemicals involved in the SUPERCLEAN production. Among the 40 job positions at the factory, only 5 are directly involved with the production of SUPERCLEAN.

After talking with some environmental experts and epidemiologists, you believe that an individual needs to have been exposed to SUPERCLEAN for at least 6 months before a sufficient dose of the chemical accumulates and physiological changes start taking place. Thus, exposure to SUPERCLEAN production chemicals for less than 6 months will not lead to Susser Syndrome.

You are presented with the job descriptions that are exposed to SUPERCLEAN and Glop Industries' air monitoring records.

Job Category	Maximum allowable level of exposure to SUPERCLEAN	Number of workers ^b
A	120-150 ppm ^a	800
B	150-175 ppm	200
C	175-200 ppm	500
D	200-225 ppm	150
E	≥225 ppm	250

a. ppm = parts per million

b. Exposed to SUPERCLEAN for at least 6 months, started working at Glop Industries on or before September 2002

After some deliberation, you define the exposed groups as low, medium or high exposure (depending on the maximum allowable level of exposure to SUPERCLEAN for their job category) and the unexposed group as employees either not involved with SUPERCLEAN production or those working less than 6 months in SUPERCLEAN production.

You now have the basic framework of your retrospective cohort study. You have redefined your hypothesis to incorporate your assumptions about the induction period and you have clearly defined your exposure variable. You are obviously excited to get out there and begin collecting data but you must first determine who is eligible for the study.

3. How would you define [eligibility criteria](#) for study participants? [Aschengrau & Seage pp. 203-205]

- a. Everyone working at the factory is eligible
- b. Only those who have worked at the factory as of September 1, two years ago, AND had been on the job for at least 6 months AND who were shown to be healthy at their initial or annual health check-ups as indicated by employee medical records
- c. Exclude workers who in the last three months exhibited symptoms of the disease

4. On what would you base your definition of Susser Syndrome?[Aschengrau & Seage pp. 217-219]

- a. Neurological symptoms alone
- b. Self-diagnosis of the participants
- c. Combination of neurological symptoms and laboratory tests

« Previous || Next »



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

Cohort Study

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

Data Collection

You have defined your hypotheses, exposure, participant eligibility criteria, and outcome... It's time to collect data.

5. What is the best source of information to assess Susser Syndrome (the outcome) among exposed and unexposed persons?

- a. Diagnosis of Susser Syndrome via hospital charts (based on neurological assessment and lab results)
- b. Complaints of neurological symptoms identified from employee health records
- c. Complaints of neurological symptoms based on information by the human resources department about medical leave of absence.

Your supervisor reads over your work and compliments you on a job well done. Before the study can begin and the data can be collected, however, your supervisor informs you of all the [administrative work](#) that must take place prior to the actual data collection.

[« Previous](#) || [Next »](#)



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

Cohort Study

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

Data Analysis

Now that you have collected the data, you quickly glance over the information and realize that there are a number of ways to analyze it. The most appropriate analysis of the data collected in this study employs the use of [person-time](#) as a way of taking into account the fact that subjects may have been followed for varying amounts of time (Please see Aschengrau & Seage pp. 220-221).

Learn more about [person-time calculations](#). In our retrospective cohort study, all individuals will enter the study at the same moment in time (September 1, two years ago). However, not all will exit at the same time. How can they exit the study? Any number of ways, including:

- a. The development of Susser Syndrome (once they have the disease, they are no longer at risk of developing it);
- b. Death from other competing causes;
- c. Loss to follow-up (Please see Aschengrau & Seage pg. 219-220).

[Loss to follow-up](#) presents a unique challenge in epidemiological studies. Clearly, without regular contact with study participants, it may not be possible to estimate when, and if, a person developed the disease of interest. In these situations, your calculations may be severely compromised. Epidemiologists employ two different estimates of effect to assess exposure-disease relationships in cohort studies: the [risk ratio](#) and the [rate ratio](#) (Please see Aschengrau & Seage pp. 67-69). Since this is your first real work as a budding epidemiologist, you decide to analyze the data using both measures of effect and later on compare them.

6. Calculation of the risk ratio from person-time information. [Aschengrau & Seage, Chapter 3]

The data collected by your team yield the following information:

- Number of cases among exposed - 74
 - Number of cases among unexposed - 120
 - Total number of exposed individuals - 1,900
 - Low exposure group - 1,000
 - Medium exposure group - 650
 - High exposure group - 250
 - Total number of unexposed individuals - 7,400
- a. How would you present the data in the 2x2 table?
 - b. Calculate the risk of disease among the exposed. The formula for calculating risk is:
(Number of exposed cases per 2-yr time period) / (Total number of exposed persons per 2-yr time period)
 - c. Calculate the risk of disease among unexposed
 - d. Calculate risk ratio
 - e. Interpret your findings

Intellectually curious?

In the preceding example, you estimated the magnitude of risk due to exposure to SUPERCLEAN by

comparing those with exposure to those without exposure. However, the exposure data could be characterized more accurately by dividing into three exposure categories, i.e., low, medium and high exposure. If the risk increases with the increase in exposure level, then one can conclude that there is a dose-response relationship in the data, i.e. biological dose gradient. The presence of the dose-response relationship strengthens our conviction that the relationship is causal.

Please calculate the incidence risk in the three exposure groups using the following data:

Level of Exposure	Disease +	Disease -	Total
Low	20	980	1000
Medium	30	620	650
High	24	226	250
Unexposed	120	7280	7400

Check your answers here.

7. Calculation of the rate ratio [Aschengrau & Seage, Chapter 3].

The data collected by your team yield the following information:

- Number of cases among exposed - 74
 - Number of cases among unexposed - 120
 - Number of exposed person-time of observation (PYO)- 3,675
 - Low exposure group- 2,000 PYO's
 - Medium exposure group- 1,225 PYO's
 - High exposure group- 450 PYO's
 - Number of unexposed PYO's- 14,550
- a. How would you present the data in the 2x2 format?
 - b. Calculate the incidence rate among the exposed. The formula for calculating incidence rate is: (Number of exposed cases during 2-yr time period) / (PYO's among exposed persons during 2-yr time period)
 - c. Calculate the incidence rate among the unexposed.
 - d. Calculate the rate ratio.
 - e. Interpret your findings.

8. Calculation of rate ratio in different age strata.

The data collected by your team yield the information:

Age Group	Exposed		Unexposed	
	Number of Cases	PYO	Number of Cases	PYO
< 30	43	2,188	75	9,249
≥ 30	31	1,487	45	5,301
Total	74	3,675	120	14,550

- a. Calculate the rate of disease among the exposed in each age group
- b. Calculate the rate of disease among the unexposed in each age group
- c. Calculate the rate ratio in each age group
- d. Interpret your findings
- e. Does the association between SUPERCLEAN and Susser Syndrome seem to vary by age group?

If you had chosen instead to compare the rate of Susser Syndrome in the exposed workers at Glop Industries to the rate of Susser Syndrome in the general population (e.g. the city of Epiville), the resulting rate ratio would be called the [Standardized Incidence Ratio](#) (SIR).

[Learn more on how to calculate the standardized incidence ratio \(SIR\) here.](#)

9. After putting exhaustive effort into data analysis, you present your findings to your supervisor. What should you tell her?

- a. The elevated estimates (both risk and rate ratios) do not support your hypothesis that exposure to SUPERCLEAN production is associated with Susser Syndrome.
- b. The exposure to SUPERCLEAN production is the definite cause of Susser Syndrome. Those elevated rates are very convincing.
- c. The data clearly suggest an association between exposure to SUPERCLEAN at Glop Industries and successive development of Susser Syndrome. I think we might want to explore other potential exposure sources as well and try to improve exposure measurement.

[« Previous](#) || [Next »](#)



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

Cohort Study

[Introduction](#)
[Learning Objectives](#)
[Student Role](#)
[Study Design](#)
[Data Collection](#)

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.

Data Analysis

Discussion Questions

Print Module

- What would you have done to improve the design of this retrospective cohort study? Propose a new study using a prospective design to investigate the relationship between SUPERCLEAN and Susser Syndrome.
- Why do we need to look at the differences in age distribution of risks in the cohort and how should we interpret our findings?
- Do you think the results of this cohort study are suggestive or conclusive about the effect of the SUPERCLEAN on Susser Syndrome? Is there a need to conduct further studies?

[« Previous](#) |

