



Cohort Study

Step 1: Learning Objectives

- A. Apply the principles of cohort studies:
 1. Define population at risk of disease
 2. Define eligibility criteria for study participants
 3. Define exposure - Timing and measurement
 4. Define outcome - Ascertainment of outcome
- B. Employ steps in data analysis of cohort studies to analyze the data:
 1. Administrative procedures before start of the study
 2. Calculate relative risk based on simple counts - Interpret your estimate
 3. Calculate relative rate from person-year information - Interpret your estimate
 4. Calculate relative risk in exposure subgroups - Interpret your estimate
 5. Calculate standardized incidence ratio - Interpret your estimate
- C. Explain your findings and discuss problems in data analysis:
 1. Reconcile the differences between relative risk and relative rate estimates
 2. Give your suggestions for carrying out this retrospective cohort study
 3. Analyze results of exposure subgroup analysis and suggest how they influence our certainty about the results of the study
 4. Discuss the value of age standardization
 5. Compare the values of crude and age-adjusted relative rates with standardized incidence ratio and explain the difference
 6. Design a prospective cohort study which would investigate the relationship between exposure and outcome
 7. Discuss the value of statistical data analysis for conclusions about causality of the exposure-outcome relationship

Step 2: Introduction to the Study

Susser Syndrome, a rare and debilitating neurological disease, is striking the People of Epiville!

You have just begun your internship at the Epiville Department of Health and your supervisor has been called in to investigate the possible causes of the sudden increase in Susser Syndrome cases. After careful thought, she wants you to lead the investigation and report to her continuously. Armed with your trusty Gordis textbook and your love of epidemiology, you decide upon a plan of action.

Step 3: Student Role - Your Plan of Action

You need to gather the following background information found on the various web pages.

- [the Epiville Chamber of Commerce web site](#)
- Information about [Glop industries](#)
- Information about [Susser Syndrome at the Epiville Department of Health Web Page](#)

Listen to the WEPI1 which provides background to your investigation (text of the newscast is also available). Based on your own research and the newscast, you decide to investigate Glop Industries.

Interview Transcript

Reporter

"Welcome to WEPI Channel 1 news. I am Lenny Regression. Now to Health & Medicine. Doctors at Epiville General Hospital report what appears to be a dramatic increase in the number of patients suffering from a cluster of neurological symptoms including dizziness, double-vision, fainting spells, and difficulty in concentration. Medical experts indicate that these symptoms are consistent with a diagnosis of Susser Syndrome, a rare and debilitating disease found only sporadically in Epiville in years past. However, 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 characterized; however, medical experts believe that its occurrence is linked to an environmental exposure, which 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 such products as the energy bar Endurobrick, the sports drink Quench-it, and the antibacterial cleaning solution, Superclean, among other products. When asked by Channel 1 reporters about the number of employees being diagnosed with Susser Syndrome, representatives from Glop had no comment.

The Epiville Department of Health is apparently aware of the Susser Syndrome cases and is investigating the matter further to determine if there is indeed an increase as suspected and to determine the cause of this increase. Medical experts and public health officials are expected to be working around the clock.

In the meantime, hospital officials warn that those suffering from dizziness, double vision, and fainting spells should immediately seek medical attention in the hospital emergency room. Stay tuned as Channel 1 News continues to investigate this Susser Syndrome outbreak. If you would like to obtain more information on the situation, go to the Epiville town home page."

You enter the Glop Industries, flash the powerful Epiville Department of Health Identification Card (being sure to cover the word "intern" with your thumb), and ask to speak with the plant manager. You are immediately greeted by Ms. Dolores Doll who is very responsive to your questions.

Interview Transcript

Plant Manager

"Over the years, Glop Industries has produced more than 30 products, ranging from household cleaning supplies to pre-packaged frozen dinners to soft-drinks. Currently, we are producing three items. Quench-it and Endurobrick are being manufactured under a limited production run. Most of the plant space is being used to produce Superclean. We got the Superclean account in early 2000 and have been pumping it out ever since - over a thousand gallons a week. Its a great product...real versatile. I know because we use it at home in all sorts of ways - cleaning the dishes, cleaning the floor, cleaning the clothes - I even use it in the shower and I'll tell you, I've never felt so clean. Anyway, that fresh lemony scent you smell, that's Superclean. 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 which is actually a pretty good number but we are hoping to bring it down a bit. At any rate, a little Superclean in the air never hurt anyone - it keeps us feeling clean and the factory smelling great. Incidentally, I really have nothing to say about employees taking sick leave - you'll have to talk to the president about that and I believe he's vacationing in Monte Carlo right now."

Questions in Steps 3 and 4 require you to demonstrate your knowledge of epidemiological terminology and concepts, including study design and data collection. In **Step 5 Data Analysis** you will perform calculations of the measures of effect and explain your findings. Select what you think is the best answer for each question. At the same time, read carefully through the explanations of both correct and incorrect answers. Finally, answer the discussion questions in Step 6 found at the end of the exercise. Bring your answers to your seminar section and be prepared to discuss them in class. Please proceed to Step 3. Good luck and have fun!

Step 4: Study Design

As a first step, you want to generate a solid hypothesis to guide the investigation. In order to do that, you need some background information first. Initially, you decide to contact people in the hospital to inquire about the individuals with this illness. On 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, are affected as well, albeit in smaller numbers. You decide to design a retrospective cohort study.

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

a. Those who are exposed to chemicals involved in the production of Quench-it (via direct exposure at the factory) have a higher risk of developing Susser Syndrome than those who are not exposed.

Incorrect

The recent appearance of Susser Syndrome does not necessarily coincide with the longer production history of Quench-it at Glop Industries. Another exposure variable is more promising.

b. Those who develop Susser Syndrome are at a higher risk of having been involved in the production of SUPERCLEAN than those who did not develop Susser Syndrome.

Incorrect

When conducting a cohort study, we are interested in comparing outcomes between exposed and non-exposed groups and thus interested in estimating the rate of disease development. This proposed hypothesis implies the comparison of exposure status between the diseased and non-diseased groups and would be more appropriate of a case-control study design.

c. Those who are exposed to chemicals involved in the production of SUPERCLEAN (via direct exposure at the factory) have a higher rate of developing Susser Syndrome than those who are not exposed.

Correct

When generating a hypothesis, we need to be specific without being so stringent as to limit study participation. Here, we specify the exposure of interest and indicate we wish to estimate the rate of disease development in the exposed and unexposed.

d. Residents of Epiville are at higher rate of developing Susser Syndrome than the residents of the neighboring community.

Incorrect

This hypothesis is too general and broad in scope. Findings of a study based on this hypothesis would do little to elucidate the cause of Susser Syndrome.

Your supervisor assembles a team to begin the investigation. You must determine the available sources of information. After a little groundwork, you find that the employee health clinic at Glop Industries keeps records of comprehensive annual medical examinations of all

employees beginning with their hiring date. You also learn that the factory's human resources department has records of each worker's employment position, which you can use as an indicator of the level of exposure (duration and intensity of exposure) to chemicals involved in SUPERCLEAN production. Given the availability of these data, your team decides to compare individual workers involved in the production of the SUPERCLEAN ("exposed") with workers involved with the production of other products at the factory ("unexposed"). Exposure is considered to be cumulative and will be calculated for every worker from September 1st, 2000 (time when SUPERCLEAN production started) to the present (September 1st, 2002). To simplify the study design (after all, you have just begun studies!), you decide to enroll a cohort in which all study participants enter into the study at the same time (September 1st, 2000). No individuals will be allowed to enter after the start of the study. Obviously, study participants may not be followed for the entire length of the study (2 years).

2. Based on the above data, how should you compare the individual workers (i.e. how would you define the exposed and unexposed groups)? [See Gordis, Ch. 8, pg. 133-135]

- a. Keep the exposure variable as dichotomous; thus, the employees involved in the production of SUPERCLEAN are exposed and those not involved in the production are unexposed.

Incorrect

While this can certainly be done, we have the advantage of knowing both the duration and intensity of exposure. We should use this information to better characterize and strengthen the evidence of any association between exposure and outcome.

- b. Categorize the exposed group based on the level and duration of exposure.

Correct

When possible, categorization can improve our ability to characterize the exposure- outcome relationship as well strengthen the evidence of association.

Due to the complexities of quantifying chemical exposure, you decide it will be classified into 4 categories. Among the 40 job positions at the factory, only 5 positions work directly with the production of SUPERCLEAN and can be considered potentially "exposed." Additionally, after talking with some environmental experts and epidemiologists, you believe that an individual needs to have been exposed for a minimum of 6 months before a sufficient dose of the harmful chemical accumulates and any physiological changes can take place leading to the possible development of Susser Syndrome. This, you recall, is termed the induction time. This is the time between exposure and disease development during which a number of causes have to occur to result in the disease. For example, we may believe that Susser Syndrome is the result of the combination of a) exposure to a certain threshold of manufacturing chemicals; b) genetic mutation; c) a decrease in the production of a certain neurotransmitter. Thus, Susser Syndrome cannot occur until all three causes occur which may take months or even years. Any occurrence of disease within this induction period cannot be attributed to that particular exposure. Often, the induction period in an individual is unknown and we are forced to make educated guesses. For this study, you are assuming a

minimal induction period of 6 months. Exposure to SUPERCLEAN production chemicals of less than 6 months will not lead to Susser Syndrome.

Therefore, you decide to define the exposed and unexposed groups as follows:

Unexposed

- not working in one of the 5 positions directly involved in SUPERCLEAN production.
- having worked less than 6 months in one of these 5 positions during the two-year follow-up period.

Exposed

- low exposure (working 6 to 12 months in one of the 5 exposed job categories).
- medium exposure (working 12 to 18 months in one of the 5 exposed job categories).
- high exposure (18 to 24 months in one of the 5 exposed job categories).

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 and the underlying population to be sampled. You are obviously excited to get out there and begin collecting data but you have not yet finished designing the study. You must first determine who is eligible for the study.

3. How would you define eligibility criteria for study participants? [See Gordis, Ch.8, pg. 133-135]

a. everyone from the selected sample is eligible

Incorrect

These eligibility criteria will include workers who have been on the job for less than two years; remember that we decided to enroll only those who already worked at the factory as of September 1st, 2000.

b. only those who have worked at the factory for at least two years AND who were shown to be healthy at their initial or annual health check-ups (mandatory for all employees of the factory) as indicated by employee medical records

Correct

We only want to capture incident cases - those who develop the disease for the first time after the exposure starts and the study begins. It is important to include only new cases of disease and not pre-existing, prevalent cases, as it may lead to spurious findings. We also want to enroll only those who already worked at the factory as of September 1st, 2000) [Refer to Gordis, Ch.3, pg. 32-36]

- c. exclude workers who in the last three months exhibited symptoms of the disease

Incorrect

We need to exclude people who were sick at the start of follow-up, not those who became sick during the follow-up period.

Not only must the exposure variable be defined, but so too the outcome variable. Susser Syndrome is a complicated disorder with many symptoms.

4. On what would you base your definition of Susser Syndrome? [See Gordis, Ch.3, pg. 48-51]

- a. the neurological symptoms alone

Incorrect

These symptoms are not very specific and have a broad differential diagnosis. Consequently, individuals who do not actually suffer from the disorder may be incorrectly classified as having the disorder. These cases would be false positives [Gordis, pg. 65] Alternatively, not all Susser Syndrome patients will present with the full panoply of neurological symptoms. Thus, a number of true cases may be missed. These cases would be false negatives. [Gordis, pg. 65]

- b. the self-diagnosis of the participants

Incorrect

Not only can Susser Syndrome not be self-diagnosed, but the accuracy of self-diagnosis in general is always suspect and can lead to either an over- or underestimate.

- c. results of a CT scan

Incorrect

The CT scan is an expensive and time-consuming procedure and not necessary to diagnose all cases. Furthermore, it is unclear if a CT scan will pick up cases at the early stage of their development. Thus, the exclusive reliance on a CT scan may miss many early cases and only include the most severe ones. In other words, the CT scan would result in a number of false negatives and thus have low sensitivity. [Gordis, pg. 65]

- d. combination of neurological symptoms and blood test

Correct

This is the most accurate and economical diagnosis. Thus, only those symptomatic individuals with confirmed blood test diagnosis should be considered as suffering from Susser Syndrome.

Step 5: Data Collection

Now that you have defined the appropriate variables and determined the information you want to collect, you are ready to determine exactly how you are going to collect your data. You need to collect information on the exposure variable and information on the outcome variable.

5. What is the best source for gathering information on the exposure variable? [See Gordis, Ch.3, pg. 37; Ch.18, pg. 283-285]

- a. Collect information from the human resources department on all employees and their job categories as of September 1st, 2002
- b. Assess baseline exposure at the beginning of the study period (September 1st, 2000)
- c. Collect monthly updated files in the human resources department, beginning with the start of the study through the end of the study

Incorrect

This does not provide us with the adequate exposure information because it gives us no data on the duration of exposure.

Incorrect

Again, this tells us nothing about duration of exposure as individuals would continue to be exposed after the baseline assessment.

Correct

This will tell us both the job category of the individual and the length of exposure. Thus, we will be able to estimate the duration of exposure.

6. What is the best source for gathering information on the outcome variable? [See Gordis, Ch.3, pg. 35-36]

- a. diagnosis of Susser Syndrome identified from the hospital charts of the local hospital and based on the neurology consultation and supported by lab results

Correct

We will not miss any cases as all subjects with disorder will end up at the local academic hospital and all cases will be valid.

b. complaints of neurological symptoms identified from the records of the employee health clinic

Incorrect

We might miss persons with the syndrome who quit work at the factory and might include persons whose symptoms are not actually full-blown Susser Syndrome.

c. complaints of neurological symptoms based on information provided by the human resources department about medical leave of absence

Incorrect

We might miss many legitimate cases and obtain many invalid cases.

7. What is the best source for gathering information on the outcome variable? [See Gordis, Ch.3, pg. 35-36]

a. look through the records of the local hospital dated from September, 2000 to September, 2002 to identify those with disorder and see if there is information in the chart about their employment

Incorrect

This is not the best choice because employment information might be missing from the charts; in addition, subjects who had already quit their job at the factory as a result of the illness would be missed.

b. link a computer database containing cohort members from Glop Industries with the discharge database from the local hospital

Correct

Modern linkage techniques allow us to identify approximately 95% of people based on their name and birth date. Because the local hospital treats all potential cases, the only cases you are likely to miss are those who moved out of the area before treatment or those patients discharged after the study end date.

Your supervisor reads over your recommendations and decides to send out two teams of investigators - one to Glop Industries and the other team to the hospital. However, before the teams can leave and the data can be collected, your supervisor instructs you of all the administrative work that must be in order. You need to:

1. Get approval of your study from your Institutional Review Board (IRB) - this will ensure that the study adheres to the ethical principles of conducting public health research and that the rights of study participants are protected.
2. Prepare a budget and get funding.

3. Develop an operations manual to be used by all study personnel which will describe the standardized procedures for collecting data, managing data, etc. This is used to ensure quality control.
4. Design a consent form for study participants - it should clearly indicate your goals; it should also explain risks, benefits and expenses that participants might incur if they decide to participate in your study; it should also state how participants would benefit from their participation and how you plan to make use of the data once it is collected.
5. Design a questionnaire that will be used to collect data from the study participants.
6. Hire and train data extractors.
7. Design a data management plan (how and where paper forms will be stored, when and how they will be entered into computer database, when periodical checks of the data will be performed to spot possible problems in the study).
8. Design a data analysis plan and think about publicizing your findings.

Having received the necessary IRB approval, designed the consent form, and planned out all of the logistics, the teams are deployed. The data begin to file back to the Department of Health and must now be collated and carefully entered into the computer database. Despite your leadership role, you are still the intern and thus have the humble yet crucial responsibility of data entry. Once all of the data are entered, you can proceed to the analysis stage where the association stated in your hypothesis is characterized and tested.

Step 6: Data Analysis

Before crunching the numbers, you quickly glance over the data and realize that an appropriate analysis of the data collected in this study employs the use of person-years as a way of taking into account the fact that subjects may be followed for varying amounts of time (see Gordis, Ch.5, pg.83-85). This allows the researcher to account for those who dropped out of the study and no longer contribute to person-years at risk due to a variety of reasons (moved away, refused to participate, died from unrelated causes, etc.). At the end of follow-up period, all person-years are summed up to represent the cumulative time at risk for disease. The time at risk for each person will be calculated from the time the individual entered the study until the time he/she exits the study. As previously stated, all individuals will enter the study at the same moment in time. However, no 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 endpoint, they are no longer at risk of developing it);
- b. death; or
- c. loss-to-follow-up, meaning they choose to no longer participate in the study.

Loss-to-follow-up results in data not being collected for the epidemiological study. We may not know when the study participants dropped out and thus we may not know whether they developed the disease. It becomes impossible to directly calculate person-years. In these situations, epidemiologists may use simple counts of subjects to calculate measures of effect. This is obviously not the best choice, but it provides an estimate of the true measure of effect. This is your first real work as a budding Epidemiologist and you decide to analyze the data using both simple counts and person-years. It is time to get to work!

8. Calculation of the rate ratio based on simple counts. [See Gordis, Ch.3, pg. 32-33, Ch.10, pg. 159-162, 171]

The data collected by your team yield the following counts:

- Total number of exposed individuals - 1900
 - low exposure group - 1000
 - medium exposure group - 650
 - high exposure group - 250
- Total number of unexposed individuals - 7400
- Number of exposed diseased (all people who develop Susser Syndrome among the exposed) - 74
- Number of unexposed diseased - 120

a. The first step is to tabulate the data in the classic 2x2 table. How would you do this?

Disease +	Disease -	Total	
Exposed	74	1826	1900
Unexposed	120	7280	7400

b. Calculate cumulative incidence among all exposed

Cumulative incidence in exposed = $74 / 1900 = 0.039$ (or 39 cases per 1000 exposed per 2 years or 20 cases per 1000 exposed per year)

c. Calculate cumulative incidence among unexposed

cumulative incidence in unexposed = $120 / 7400 = 0.016$ (or 8 cases per 1000 unexposed per year.)

d. Calculate rate ratio

cumulative incidence in exposed / cumulative incidence in unexposed = $0.039 / 0.016 = 2.44$

e. Interpret your finding

Those who are exposed to chemicals involved in SUPERCLEAN production for at least 6 months have a 2.4 times higher rate of developing Susser Syndrome than those who are not exposed to SUPERCLEAN production.

9. Calculation of the rate ratio from person-year information. [See Gordis, Ch.10, pg. 159- 162, 171]

- Number of exposed person-years of observation (PYO) - 3700, i.e.,
 - low exposure group - 2000 PYO
 - medium exposure group - 1250 PYO
 - high exposure group - 450 PYO
- Number of unexposed person-years - 14500 PYO
- Number of exposed cases - 74
- Number of unexposed cases - 120

a. Again, how would you present the data in the 2x2 format?

	PYO	Number of Cases
Exposed	74	3700
Unexposed	120	14,500

b. Calculate cumulative incidence among all exposed

Inc rate in exposed = $74 / 3700 = 0.02$ (or 20 cases per 1000 PYO)

c. Calculate cumulative incidence among unexposed

Inc rate in unexposed = $120 / 14500 = 0.0083$ (or 8 cases per 1000 PYO)

d. Calculate rate ratio

Inc rate in exposed / Inc rate in unexposed = $0.02 / 0.008 = 2.4$

e. Interpret your finding

Those who are exposed to chemicals involved in SUPERCLEAN production for at least 6 months have a 2.4 times higher rate of developing Susser Syndrome than those who are not exposed to SUPERCLEAN production.

The above analyses are called "crude analyses." They suggest that there is an association between involvement with SUPERCLEAN production and the development of Susser Syndrome. You decide to better characterize this association using the information you have collected detailing the exposure sub-groups.

10. Calculation of rate ratio in exposure sub-groups. [See Gordis, Ch.10, pg. 159-162, 171]

- Number of exposed person-years of observation (PYO) - 3700, i.e.,
 - low exposure group - 2000 PYO
 - medium exposure group - 1250 PYO
 - high exposure group - 450 PYO
- Number of unexposed person-years - 14500 PYO
- Number of exposed cases - 74
 - low exposure group - 32
 - medium exposure group - 30
 - high exposure group - 12

a. There is too much information here to present in the simple 2x2 format. How would you present the data in the table according to different exposure sub-groups?

*See table below

b. Calculate incidence rate among exposed by level of exposure

Inc rate in exposed (Low exposure) = $32 / 2000 = 0.016$ (or 16 cases per 1000 PYO)

Inc rate in exposed (Medium exposure) = $30 / 1250 = 0.024$ (or 24 cases per 1000 PYO)

Inc rate in exposed (High exposure) = $12 / 450 = 0.027$ (or 27 cases per 1000 PYO)

c. Calculate incidence rate among unexposed

Inc rate in unexposed = $120 / 14500 = 0.0083$ (or 8 cases per 1000 PYO)

d. Calculate rate ratio at each level of exposure

Inc rate in exposed / Inc rate in unexposed (Low exposure) = $0.016 / 0.0083 = 1.93$

Inc rate in exposed / Inc rate in unexposed (Medium exposure) = $0.024 / 0.0083 = 2.89$

Inc rate in exposed / Inc rate in unexposed (High exposure) = $0.027 / 0.0083 = 3.25$

e. Interpret your finding

Those who are exposed to low levels of exposure have a 1.9 times higher rate of developing a neurological disorder than

those who are not exposed.

Those who are exposed to medium levels of exposure have a 2.9 times higher rate of developing a neurological disorder than those who are not exposed.

Those who are exposed to high levels of exposure have a 3.3 times higher rate of developing a neurological disorder than those who are not exposed.

- f. What is this pattern of increase in the rate ratio consistent with?

Taken together, these findings are consistent with a dose-response relationship. As the dose of exposure increases, the rate of disease development also increases.

*** Answer to question 10a**

Exposure Group	Low Exposure		Medium Exposure		High Exposure	
	Exposed	Unexposed	Exposed	Unexposed	Exposed	Unexposed
Number of Cases	32	120	30	120	12	120
PYO	2000	14,500	1250	14,500	450	14,500

11. Calculation of rate ratio in different age strata. [See Gordis, Ch.10, pg. 159-162, 171]

The crack team of field agents has presented you with the data on the age distribution of all subjects in the cohort, detailed as follows:

Age Group	Number of Cases in Exposed	PYO in Exposed	Number of Cases in Unexposed	PYO in Unexposed
Younger than 30	17	1000	30	4257
30 - 39	26	1200	45	5037
40 - 49	21	1000	40	4606
50 and older	10	500	5	600
Total	74	3700	120	14,500

a. Calculate incidence rate among exposed in each age group

Inc rate in exposed = $17 / 1000 = 0.017$ (or 17 cases per 1000 PYO) in the age group younger than 30

Inc rate in exposed = $26 / 1200 = 0.0216$ (or 20 cases per 1000 PYO) in the age group 30-39

Inc rate in exposed = $21 / 1000 = 0.021$ (or 20 cases per 1000 PYO) in the age group 40-49

Inc rate in exposed = $10 / 500 = 0.02$ (or 20 cases per 1000 PYO) in the age group 50+

b. Calculate incidence rate among unexposed in each age group

Inc rate in unexposed = $30 / 4257 = 0.0070$ (or 7 cases per 1000 PYO) in the age group younger than 30

Inc rate in unexposed = $45 / 5037 = 0.0089$ (or 9 cases per 1000 PYO) in the age group 30-39

Inc rate in unexposed = $40 / 4606 = 0.0087$ (or 9 cases per 1000 PYO) in the age group 40-49

Inc rate in unexposed = $5 / 600 = 0.0083$ (or 8 cases per 1000 PYO) in the age group 50+

c. Calculate rate ratio in each

Inc rate in exposed / Inc rate in unexposed = $0.017 /$

age group	<p>$0.0070 = 2.4$ in the age group younger than 30</p> <p>Inc rate in exposed / Inc rate in unexposed = $0.0216 / 0.0089 = 2.4$ in the age group 30-39</p> <p>Inc rate in exposed / Inc rate in unexposed = $0.021 / 0.0087 = 2.4$ in the age group 40-49</p> <p>Inc rate in exposed / Inc rate in unexposed = $0.02 / 0.0083 = 2.4$ in the age group 50+</p>
d. Interpret your findings	<p>Those who are exposed and are younger than 30 have a 2.4 times higher rate of developing a neurological disorder compared to those who are not exposed and are younger than 30.</p> <p>Those who are exposed and are 30-39 years old have a 2.4 times higher rate of developing a neurological disorder compared to those who are not exposed and are 30-39 years old.</p> <p>Those who are exposed and are 40-49 years old than 30 have a 2.4 times higher rate of developing a neurological disorder compared to those who are not exposed and are 40-49 years old.</p> <p>Those who are exposed and are 50 or older than 30 have a 2.4 times higher rate of developing a neurological disorder compared to those who are not exposed and are 50 or older.</p>
e. Does the association between exposure and outcome seem to vary by age group?	No.

12. Calculation of standardized incidence ratio (extra credit). [See Gordis, Ch.3, pg. 54-

56]

You have data available from the local department of health on the annual incidence rate of the neurological disorder in Epiville. These data would allow you to calculate the standardized incidence ratio (indirect method) to determine if the incidence among SUPERCLEAN employees is higher than the incidence in the general population. Because the age distribution of the general population is quite different from the age distribution of the working population you have to take into account the age structure of the respective groups.

Age Group	Incidence Rate
Younger than 30	0.0039
30 - 39	0.0052
40 - 49	0.0047
50+	0.0062

- a. Calculate the number of observed cases (total of cases among exposed and unexposed) and PYO in each age strata

Age Group	Observed Cases	Observed PYO
Younger than 30	47	5257
30 – 39	71	6237
40 – 49	61	5606
50+	15	1100
Total	15	1100

- b. Calculate the number of expected cases in each strata

*See table below

c. Calculate standardized incidence ratio (SIR)

Age Group	Observed Cases	Expected Cases
Younger than 30	47	21
30 – 39	71	32
40 – 49	61	26
50+	15	7
Total	194	86

d. How do you interpret your findings?

Factory employees exposed to SUPERCLEAN have a 2.3 times higher incidence rate of neurological disorder than the general population of Epiville.

*** Answer to question 12b**

Age Group	Incidence Rate in the General Population	Observed PYO	Expected Cases
Younger than 30	0.0039	5257	$0.0039 \times 5257 = 21$
30 – 39	0.0052	6237	$0.0052 \times 6237 = 32$
40 – 49	0.0047	5606	$0.0047 \times 5606 = 26$
50+	0.0062	1100	$0.0062 \times 1100 = 7$
Total	194	86	86

13. After putting in an exhaustive effort of data analysis, you present your findings to your supervisor. What should you tell her?

a. It looks like we were chasing a red herring. In my opinion, there does not appear to be any relationship between working with SUPERCLEAN and the development of Susser Syndrome. Let's look at another source of exposure.

Incorrect

The consistently elevated risks do support our hypothesis that SUPERCLEAN production may be associated with Susser Syndrome.

b. The data regarding the possible association between SUPERCLEAN and the development of Susser Syndrome is totally inconclusive. I think we should repeat the study with more participants.

Incorrect

The elevated rates are not totally inconclusive and suggest an association between the exposure and the outcome.

c. The exposure to SUPERCLEAN production is the definite cause of Susser Syndrome. Those elevated rates are very convincing.

Incorrect

SUPERCLEAN appears to be associated with the development of Susser Syndrome. However, as detailed in Gordis [CH.13, pg. 193-195], to move from association to causation requires a substantial amount of epidemiological evidence as well as biological plausibility. At this stage in the investigation, we are far from having enough.

d. The data clearly suggest an association between exposure to SUPERCLEAN production and later development of Susser Syndrome. I think we might want to explore other potential exposure sources to be sure as well as further characterizing this association.

Correct

The data do suggest an association; however, we need to check the statistical significance of these findings as they may be due to chance. Furthermore, it is important to rule out other potential exposures as they may confound the findings.

Step 7: Seminar Discussion Questions

Carefully consider the following questions related to your work above. Write down your answers and be prepared to discuss them in seminar.

1. How and why the results of Q8 are different from the results of Q9? Give examples of the events that could have influenced the results.
2. What would you have done to improve the design of this retrospective cohort study?
3. What does the crude rate ratio tells us?
4. What does the monotonic increase in rate ratio among various exposure groups in Q10 tells us? How does it influence our certainty about the results of this study?
5. Why do we need to look at the age distribution of risk in the cohort (see Q11) and how should we interpret our findings?
6. Design a prospective cohort study which would investigate the relationship between exposure to SUPERCLEAN and Susser Syndrome. Use information from the Study Design section to guide yourself through the necessary steps.
7. Do you think the results of this cohort study are suggestive or conclusive about the effect of the SUPERCLEAN on neurologic morbidity?
8. Is there a need to conduct further studies?

Questions for the Intellectually Curious

1. What does SIR in Q12 tells us?
2. Why do you think SIR is lower than the crude rate ratio?
3. Let us assume that Susser Syndrome was a reportable disease. In other words, physicians and hospitals would be required to report all patients diagnosed with Susser Syndrome to the Epiville Department of Health (this is a form of passive surveillance). How would your data collection change if this were the case? What are the advantages and disadvantages of such change, vis-a-vis hospital discharge data?