Ecological Study

Step 1: Learning Objectives

A. Apply the principles of ecologic studies:
    1. Define study hypothesis
    2. Define the difference between individual and group characteristics
    3. Define the source population for the study
    4. Describe sources of information for ecological studies
B. Employ steps in data analysis of ecologic studies to analyze the data:
    1. Administrative procedures before start of the study
    2. Calculate incidence rates based on simple counts
        A. Interpret your estimates
    3. Plot incidence rate against exposure
        A. Interpret your plots
C. Explain your findings and discuss problems in data analysis:
    1. Explain your conclusions
    2. Give your suggestions for carrying out ecologic study
    3. Define and explain ecologic fallacy
    4. Describe ecological studies in terms of validity, cost effectiveness, time requirement and ability to estimate measure of effect
    5. Describe ecological studies in comparison with cohort and case-control studies
    6. 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 Porks-A-Lot Pig Farm
- Information about Epiville reservoirs
- Information about Susser Sydrome 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 Porks-A-Lot Pig Farm.

Interview Transcript

Reporter

Bill: Welcome to WEPI News. I'm Bill Williams. We're going to a breaking report from the field where we have Lynn Regression standing by at the Epiville Hospital. Lynn?

Lynn: Welcome to the Health and Medicine desk of WEPI Channel 1 news. I am Lynn Regression. The rare and debilitating disease doctors call Susser Syndrome appears to be striking Epiville residents. Doctors from the Epiville General Hospital report a dramatic increase in new cases of the disease since March of this year. Susser Syndrome is characterized by a cluster of neurological symptoms including dizziness, double vision, fainting spells, and difficulty concentrating. Historically, it has been found sporadically throughout the United States and can be fatal if left untreated.

We spoke with hospital officials earlier today and they report that the number of Susser Syndrome patients has been slowly rising over the past 2 years. However, since March, doctors have reported an unprecedented increase. The cause of this increase is unknown, although medical experts believe it to be linked to an environmental exposure.

The Epiville Department of Health is investigating the matter fully and is expected to contact various individuals and industries throughout the city. A team of epidemiologists is trying to determine what may be causing this apparent increase in Susser Syndrome cases.

Hospital officials warn that those suffering from dizziness, double-vision, and fainting spells should immediately seek medical attention in the hospital emergency room. Stay tune 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 home page."

Bill: Thanks Lynn. And in environmental news, does your tap water taste funny? Channel 1 has learned that residents in the Epiville areas serviced by the Rothman Reservoir have reported a soapy flavor to their water as well as a filmy appearance. We spoke with an Epiville Water Works Department representative who was unfamiliar with the complaint but will investigate the matter. Those who use Greenland Reservoir water have made no complaints, although officials are concerned about what appears to be increased algae growth.

You enter the pig farm, 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 facilities leader. You are immediately greeted by D. Herman Murtz who is very responsive to your questions.

Interview Transcript

Farm Facilities Leader

"You never do get used to that smell, do ya? We got over 10,000 pigs in this here farm. Pretty amazing when you think this whole place was founded just 50 years ago with just 1 hog, 2 sows, and a dream. That picture on the wall, the one next to my daddy, that's King Humbert, 300 pounds of hog muscle - I guess you could say he really started it all. Anyway, we keep this place pretty state-of-art. All of the breeding is scientifically monitored and done in our custom designed fertility barn. We got 200 individual feeding pens and 75 recreation pens to keep the pigs moving. Waste, as you might imagine, is always a problem. Back in the old days we would just hose everything out and let it run off into the Sludge River. Then, as times changed, we added a filtration system such that no solids could pass through. Just a few months ago we revamped the system again, Now, we first give all the pens a hosing, run it through a filter, and collect the initial runoff in evaporatin' pools. After the first hosing, we spray every inch of the pens with Superclean to sterilize and disinfect everything - Superclean doesn't evaporate so well so we drain it out into the Sludge River where its gets diluted. To the best of my knowledge, this two-step process is really pushin' the envelope of pig farm sterilization techniques. So far, all of our pigs are clean and healthy and all of our customers are takin' home our bacon, if you catch my drift."

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

(To answer the following multiple choice and data analysis questions, please refer to Gordis, pgs. 185-187)

The first step of your investigation is to generate a solid hypothesis. Once again, you look over the information that you have gathered regarding the Susser Syndrome cases and the reservoirs in Epiville. The data you have is at the population level. Consequently, you decide to conduct an ecological study.

You and your supervisor agree on the hypothesis. Now, in order to be thorough, you must

1. Based on the facts presented, which of the following would be the most appropriate hypothesis to investigate in this ecological study?

☐	a.	Residents of Epiville are at a higher risk of developing Susser Syndrome than residents of the neighboring towns.	**Incorrect** We do not have the data for the neighboring towns to compare with that of Epiville. This hypothesis is too broad. Furthermore, ecological studies compare rates and not risks.
☐	b.	Employees of the Porks-A-Lot pig farm have a greater risk of developing Susser Syndrome than the general population of Epiville.	**Incorrect** This hypothesis is more characteristic of a cohort study. It is too specific and we do not have the necessary data to test it.
☑	c.	The population served by the Rothman Reservoir has a higher incidence rate of Susser Syndrome than does the population served by the Greenland Reservoir.	**Correct** The ecological study relates group (or aggregate) characteristics to rates in each group. In our study, we have chosen to state that the Rothman Reservoir population will have a higher rate than the Greenland Reservoir population - the alternative would also be an appropriate hypothesis to test.
☐	d.	Those diagnosed with Susser Syndrome will have greater odds of being a Porks-A-Lot employee than those without Susser Syndrome.	**Incorrect** This hypothesis is more appropriate for a case- control study.

clearly define who is to be eligible for the ecological study.

2. Which population will be the source population for the study, given that this is an ecological study?

☐	a.	Individuals should be diagnosed with Susser Syndrome and be residents of Epiville.	**Incorrect** Recall that we are undertaking an ecological study and are thus concerned with populations and not simply individuals.
☐	b.	The study population should be that of Epiville residents.	**Incorrect** This is not specific enough and only defines one study population. We are looking for two populations which we

			would like to compare.
☐	c.	One study population should be comprised of Susser Syndrome patients residing in Epiville; a second population should be comprised of healthy Epiville residents.	**Incorrect** These definitions are too specific and more appropriate for a case-control design.
☑	d.	The study population should be comprised of Epiville residents who are serviced by either the Rothman Reservoir or by the Greenland Reservoir. Furthermore, Susser Syndrome patients should have come from this underlying population.	**Correct** This provided us with the two study groups coming from the same source population. This would allow us to compare sources of reservoir water and incidence of Susser Syndrome in the two groups.

With your hypothesis and source population defined, you must now determine from exactly where the data will come from and how it should be collected.

Step 5: Data Collection

Ecological studies involve comparison and analysis of variables at the population level. They may involve direct observations of individuals which are then aggregated or summarized (to give means or proportions) or they may rely on global population measures, such as population density. The data to be used in an ecological study have typically been collected previously for other purposes and often come from such sources as population censuses and disease registries.

Your study is no exception. Rather than collecting data via fieldwork, you will journey through the often treacherous terrain of paper-based registries and computer databases. Before you begin, you must understand exactly the kind of information you are looking for so that you can narrow down the number and type databases/registries to examine.

3. On a conceptual level, define the minimal information you need to collect in order to test your ecological study hypothesis.

☐	a.	I will need to know the size of the population of Epiville and the	**Incorrect** While this information is important, more

		number of new cases of Susser Syndrome at the present time.	is needed to test your hypothesis.
☐	b.	I will need to know the number of new cases of Susser Syndrome and the size of the populations serviced by the Rothman and Greenland reservoirs.	**Incorrect** This information is certainly necessary but is not quite enough to test your hypothesis.
☑	c.	I will need to know the number of new cases of Susser Syndrome as well as their addresses. Additionally, I will need to know the geographic areas serviced by the Rothman and Greenland Reservoirs, respectively. Finally, I will need to know the population size of these areas and the amount of reservoir water used during a given time period.	**Correct** This information will allow you to cross-reference the Susser Syndrome cases with reservoir use and thus test your study hypothesis.

In order to get the information on the reservoirs, you contact the Epiville Water Works Department. They report that all of Epiville is serviced by either the Rothman or Greenland Reservoir. The geographic area served by each reservoir is subdivided into 5 sectors. The Water Works Department has a database recording the average daily water use for each sector (but not for each individual living in the sector) for each of the previous 5 years. You decide that the information regarding the most recent year is adequate.

Additionally, you have access to the most recent Epiville population census.

Finally, the Epiville Department of Health keeps a database of all reportable diseases. Susser Syndrome is reportable and as such all physicians and hospitals must report each diagnosed case along with the requisite demographic information to the health department.

4. How can all of this information be used to test your study hypothesis?

☑	a.	I can cross-reference the three databases in such a way that I will know from what reservoir sector the Susser Syndrome cases came, the population size of that sector, and the average daily water use of that population.	**Correct** By cross-referencing the databases along address/geographic location information, we can compare differences in Susser Syndrome incidence by reservoir use.

☐	b.	I can use the Water Works Department information to extrapolate the amount of water each individual in Epiville consumed and then cross reference these individuals with the Susser Syndrome registry.	**Incorrect** In an ecological study we are interested in population-based statistics and not individual based statistics. Furthermore, extrapolating the water consumption from the population level to the individual level is incorrect. This type of extrapolation is called "ecological fallacy.

5. Which of the following is a chief shortcoming of using a reportable disease registry to collect Susser Syndrome cases in our study?

☐	a.	We may not capture all of the Susser Syndrome cases occurring in Epiville.	**Incorrect** As we have defined it, all reportable diseases, including Susser Syndrome, must be reported to the Department of Health. In a perfect world, no cases would go unreported; for rare diseases and syndromes, this is usually the case.
☐	b.	We would only be capturing prevalence data and not incidence data.	**Incorrect** The registry will have a date of initial diagnosis attached to it and, if designed correctly, repeat occurrences of the disease within the same patient will also be noted. As long as we have a clearly defined time component to our study, we will be able to capture incident as well as prevalent cases.
☑	c.	We are relying on the diagnostic capabilities of the doctors and hospitals and are assuming that a case reported and recorded as Susser Syndrome is in fact such.	**Correct** We do not have the time or resources to independently verify that all reported cases have been diagnosed correctly. Rather, we educate the medical professionals as to our exact diagnostic criteria and then often act under good faith that such criteria are being used. As such, it is possible that cases included in the registry do not in fact belong.

The next step is to cross-reference your databases and extract the required information. Being no fool, you call in the database experts to help you. After a few days of computer wrangling, the databases appear to be linked and you can run your search queries.

Step 6: Data Analysis

You plan to use the Average Daily water use calculated for the previous year, the total population census calculated from the previous year, and the number of new Susser Syndrome cases reported over the previous year as contained in the reportable disease registry.

After running your database queries, you get the following printout detailing the reservoirs:

Rothman Reservoir					
	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5
Ave Daily Use (L)	165,000	189,500	142,500	88,000	130,000
Total Pop	10,300	11,300	9,500	8,000	8,900

Greenland Reservoir					
	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5
Ave Daily Use (L)	124,500	151,500	102,000	180,000	222,000
Total Pop	8,300	10,100	6,800	12,000	14,800

You also get the following printout detailing the Susser Syndrome cases over the 1-year period:

Susser Syndrome Registry		Reported over 1 year Period	
Reservoir	Sector #	# of Cases	Reported By
Rothman	1	25	Epiville General
Greenland	5	16	Epiville General
Greenland	2	10	Epiville General
Rothman	2	30	Epiville General
Rothman	5	17	Epiville General
Rothman	4	11	Epiville General
Greenland	1	8	Epiville General
Greenland	3	7	Epiville General
Rothman	3	21	Epiville General
Greenland	4	12	Epiville General

6.	**Combine the above data into a table summarizing the pertinent information. Calculate the incidence rates per 1,000 population (per 1 year) for the sectors associated with each reservoir.**	*See answer below
7.	**Using the tabulated information from question 6, create a graph showing the correlation between incidence rates of Susser Syndrome and Rothman Reservoir water intake. (Hint: model the graph after that found in Gordis pg. 186)**	**See answer below
8.	**Using the tabulated information from question 6, create a second graph showing the correlation between incidence**	***See answer below

rates of Susser Syndrome and the Greenland Reservoir water intake. (Hint: model the graph after that found in Gordis pg. 186)

***Answer to question 6**

*Incidence per 1,000 can be calculated in the following manner: (# Cases * 1,000) / (Population)*

ROTHMAN RESERVOIR

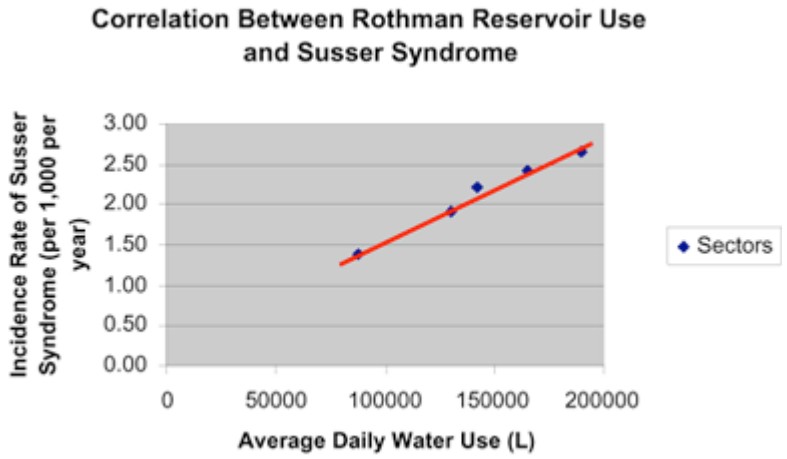
Sector	Average Daily Use (L)	Population	# Cases (over a 1-year period)	Incidence (per 1,000 per year)
1	165,000	10,300	25	2.43
2	189,500	11,300	30	2.65
3	142,500	9500	21	2.21
4	88,000	8000	11	1.38
5	130,000	8900	17	1.91
TOTAL	**715,000**	**48,000**	**104**	**2.17**

GREENLAND RESERVOIR

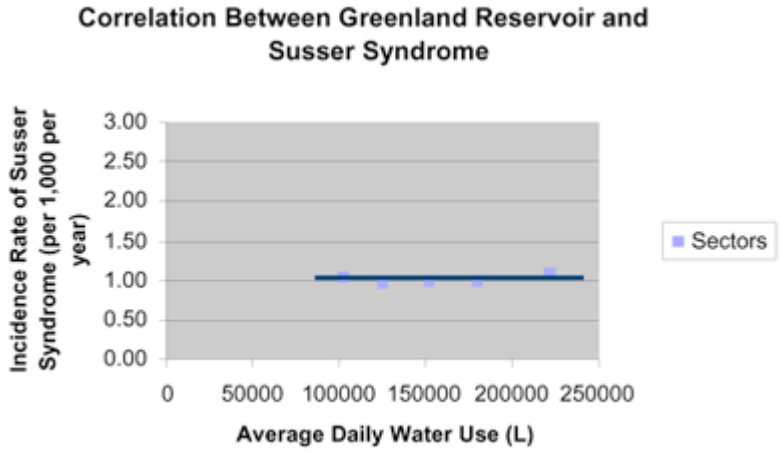
Sector	Average Daily Use (L)	Population	# Cases (over a 1-year period)	Incidence (per 1,000 per year)
1	124,500	8300	8	0.96
2	151,500	10,100	10	0.99

3	102,000	6800	7	1.03
4	180,000	12,000	12	1.00
5	222,000	14,800	16	1.08
TOTAL	**780,000**	**52,000**	**53**	**1.02**

***Answer to question 7**

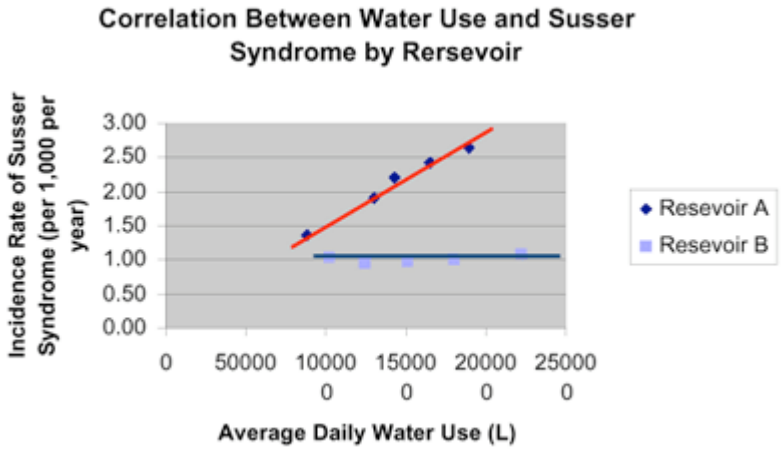


***Answer to question 8**



9. In order to simplify the visual comparison of Susser Syndrome incidence in different sectors, you decide to combine the two graphs.

Answer to question 9
You obtain the following:



10. What conclusions can you draw from the graph above?

☐	a.	The incidence rate of the population served by the Rothman Reservoir appears to be identical to the incidence rate of the population served by the Greenland Reservoir.	**Incorrect** There is clearly a difference in incidence rates between the two populations. Furthermore, the data allow us to make a more refined conclusion.)
☑	b.	Increased water use appears to be correlated with an increase in incidence rate of Susser Syndrome among populations served by the Rothman Reservoir whereas no correlation appears to exist among populations served by the Greenland Reservoir.	**Correct** The positive, increasing slope of the line fitting the Rothman Reservoir sectors suggests a dose-response relationship and displays a correlation between average daily water use and incidence rate of Susser Syndrome. The horizontal line fitting the Greenland Reservoir sectors suggests no such correlation.
☐	c.	Individuals drinking Rothman Reservoir water have approximately twice the risk of developing Susser Syndrome than individuals drinking Greenland Reservoir water.	**Incorrect** This is an example of the 'ecological fallacy' where we are applying group-level characteristics to the individuals within that group. Using the ecological study design, we can only draw

	conclusions concerning the groups or populations under analysis and we cannot draw conclusions about the individual members of the population because we do not have exposure and outcome data for each member.)

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. What are the advantages and disadvantages of ecological studies?
2. What is meant by ecological fallacy? Give some examples.
3. What is the difference between individual- and population-based studies? How is this reflected in the ecological study hypotheses?
4. What are the limitations of using secondary data sources (i.e. disease registries, census data)?

Questions for the Intellectually Curious

1. What does it mean to refer to "correlation" between variables and what kind of relationship can be determined using this?