Suspected Legionnaires' Disease in Bogalusa
A *Disease Detectives* Exercise from the
Centers for Disease Control and Prevention
Teacher's Guide and Answer Key

Teacher's Guide and Answer Key

Instructions

- Prepare student by teaching the fundamental principles of epidemiology and outbreak investigation. Materials are available in "Background and Teaching Aids" on the EXCITE web site. At a minimum, students should be familiar with the basic steps of an outbreak investigation.
- One excellent format for this exercise is to divide the class into small work groups of five to ten students and have each group assign a facilitator, a recorder, and a reporter. Ask individual students from the class at large to read the narrative and questions out loud. Then have students work in their small groups to answer the questions. Finally, have the groups report their responses to the class.
- Students can use a simple statistics program, such as Excel, to calculated attack rates and relative risk.
- The exercise is in twelve parts. Each part should be distributed independently and only after students have completed the preceding part. The total exercise requires approximately 2½ hours. An estimated time to completion is included for each part.

Learning Objectives

To teach students

- the basic principles and methods of epidemiology as they relate to scientific inquiry;
- the basic concept of risk factors for health problems and the specific the risk factors for Legionnaires disease;
- how to apply the scientific method of investigating a disease outbreak; and
- the epidemiology and clinical features of Legionnaires disease.

*This exercise is adapted for high school use from an investigation conducted in Bogalusa, Louisiana, in 1989. The original case study is used each year in CDC's Epidemic Intelligence Service (EIS) Summer Course, which trains incoming EIS Officers.

Suspected Legionnaires' Disease in Bogalusa
A *Disease Detectives* Exercise from the
Centers for Disease Control and Prevention
Teacher's Guide and Answer Key

1. What additional information do you need to decide whether or not this is a real public health problem? Base your answer on the scientific method used by disease detectives when they investigate an outbreak.

Answer:
Verify the Diagnosis

- How certain is the diagnosis? Could the number of cases be the result of having a new physician in town who is over-diagnosing?
- Are any laboratory results available? If laboratory results are positive, could they reflect laboratory error (e.g., false-positives)?

Confirm an Outbreak (use descriptive epidemiology)

- What is the denominator (i.e., source population at risk) for the observed cases? What are the referral patterns? Have they changed?
- In Bogalusa, what is the background incidence of pneumonia? of legionellosis? Has a similar cluster been noted before?
- Is any additional time/place/person information available?
- Case-finding issues: Might this be the tip of the iceberg? Are cases occurring in other hospitals or areas? How active has case-finding been?

2: What else, other than a true outbreak, could account for a sudden increase in the number of cases of a disease being reported to a health department?

Answer:

- **Increased awareness of the diagnosis**, which may make the public more likely to seek medical care or local doctors more likely to diagnose the disease.
- **Availability of a new, more sensitive laboratory test.**
- **Increased testing**, resulting perhaps from a new policy in a clinic or HMO requiring that specimens be tested from more acutely ill patients.
- **Increased reporting**, brought on because there is a new physician or clinic in town, or a change in local patient referral patterns.
- **Increased denominator**, such as would occur when students return to a college town, vacationers come to a resort area, or migrant farmers arrive in a rural area.
- **Change in the disease reporting (surveillance) system**, such as a decision by the health department to contact physicians and laboratories to identify cases, rather than waiting for their reports.
- **Laboratory error**, as in the case of false positives.

3: Before leaving your office to begin the investigation in Bogalusa, what preparation do you need to make?

Answer:
Gain Scientific Knowledge

- Discuss the situation with your supervisor or someone else who is knowledgeable about Legionnaires' disease and other pneumonias and about field investigations.
- Review applicable literature.
- Assemble useful references and sample questionnaires.

.

Gather Appropriate Supplies and Equipment

- Determine available local resources (e.g., laboratory and epidemiologic expertise).
- Consult with lab staff regarding additional materials that may be needed and proper collection, storage, and transportation techniques.
- Find portable computer, dictaphone, camera, and other supplies as needed.

Clarify Expected Role in the Field (the issue of "Who's in charge?" of this group effort)

- Agree on the investigator's role. In other words, is the investigator expected to lead the investigation, provide consultation to local staff who will conduct the investigation, or simply lend a hand to local staff?)
- Who are the local contacts? Arrange to meet with local officials and contacts upon arrival.

Teacher's Note:
Additional considerations that the students may not suggest but that you might mention relate to issues that are always important when government is involved in an investigation:

- Has the local health department been notified and are they involved? The U.S. public health care system places primary jurisdiction over a problem at the local level, with the state health department being the next level of authority. CDC serves local and state health departments in investigations by providing whatever assistance is requested following an official invitation to participate.
- Who has already been involved in the investigation? Do the public and the media know? Who else should know (e.g., neighboring counties or states)?

4a: What are possible interpretations for the data in Table 1? If you wanted to intensify the investigation, what steps would you take next?

Answer:

- First, note the basic epidemiologic pattern of seasonal variation in the numbers of hospitalized cases of pneumonia from all causes: there are more cases during the winter months than other times of the year. What factors might account for this pattern? Many viral and bacterial infections of the lungs occur most commonly during the winter months. An example is infection with the well-known influenza viruses, which can cause especially severe infections of the respiratory tract in the elderly.
- What might account for the large number of pneumonia cases in October 1989? A real increase, or a spurious (apparent but not real) increase as the result of changes in surveillance, reporting requirements, new laboratory diagnostic procedures, or other factors.

4b: If you wanted to intensify the investigation, what steps would you take next?

Answer:
On the basis of the large, and likely true, increase in cases, proceed with an investigation by developing a working case definition and beginning the search for additional cases.

5: Would you want a relatively sensitive or a relatively specific case definition in this setting? With your decision in mind, develop a case definition for this outbreak.

Teacher's Note:
A "sensitive" case definition is broad enough to identify nearly all true cases ("true positives") of the disease being investigated. However, because it is so broad, a sensitive definition may also draw in similar illnesses with different causes (known as "false positives"). In contrast, a "specific" case definition is narrow enough to exclude false positives, but may exclude some true positives that have slightly unusual symptoms. So the question to the students is whether they would want a case definition that identifies the maximum number of possible cases (high sensitivity) or one that only identifies true cases (high specificity)?

If students have difficulty understanding the distinction between sensitive and specific case definitions, you might compare deciding on a sensitive case definition with a tuna fisherman's choice to cast a wide net and catch all fish, including dolphins. Deciding on a specific case definition can then be compared with using a dolphin-free net to catch only real tuna. With the second choice, you may miss some tuna, but you will be more confident about having caught only real tuna. Appendix A provides more detail regarding these concepts, including definitions of sensitivity, specificity, predictive value positive (PVP) and predictive value negative (PVN), with a table illustrating the relationships between these measurements.

In preparing students to answer this question, you may want to review the four basic components of a case definition: 1) clinical information about the disease; 2) person information; 3) place information; and 4) time information. Remind students that epidemiologists sometimes create a hierarchy of case definitions based on the certainty of the diagnosis (i.e., a definition for confirmed cases, one for probable cases, and one for possible cases).

Additional teaching point: Although the description of "person" is often broad in an outbreak investigation, emphasis on the elements of person, place, and time is always important in epidemiologic case definitions. These criteria become even more critical when the investigation leads to a case-control study because a key to control selection is seeking people from the same general population as the cases. In other words, the "control definition" should have the same criteria for person/place/time as the case definition.

Answer:
In this investigation, the primary objective is to rapidly identify the source of the outbreak, rather than to characterize its extent. A specific case definition would, therefore, be preferred.

Following is one reasonable case definition for this outbreak investigation:

Clinical: Confirmed case: laboratory confirmation, as described in the reference Control of Communicable Diseases in Man.
Possible case: hospitalized with a physician's diagnosis of suspected Legionnaires' disease.

Person: Any male or female of any ethnicity, with no age restriction. (This is a common approach when the disease under investigation is not known to be restricted to a particular gender, ethnic group, or age group. However, as noted in Part 5 of the student exercise, the case definition in this investigation actually specified people aged $20 years.)

Place: Resident or visitor of Washington Parish or adjacent parishes.

Time: Onset of illness after September 1, 1989 (or October 1).

6a. How would you go about case-finding?

Answer:
Possibilities include

- checking with the other area hospitals, laboratories, and physicians (especially infectious disease specialists and pulmonologists (lung specialists),
- talking to case-patients, and
- requesting reports through media publicity.

Note, that the latter is a "double-edged sword" with the potential both of stimulating case-finding and of causing panic or over-reporting.

6b. Do you need to find every case?

Answer:
Finding every case would be helpful for characterizing the full extent of the outbreak, but is not necessary in investigations such as this one, where your aim is to establish the source and mechanism of the outbreak in order to stop it as quickly as possible. Time and resource limitations must also be taken into account, when considering whether or not to try to find every case.

7a. How would you generate plausible hypotheses to test in this type of investigation?

Answer:
Possibilities include

- asking local public health authorities and clinical health providers what they think,
- asking case-patients and their families what they think, and
- taking stock of what you know about the disease and its causes, reservoirs, and modes of transmission—a step epidemiologists often call "rounding up the usual suspects."

7b. What, if any, are your ideas at this point?

Answer:
Students should find clues by examining the overall patterns regarding person, place, and time observed during descriptive epidemiology and the outliers on the epidemic curve. For example:

- *Person*: About two-thirds of case-patients were female. What could this mean?
- *Place*: A high proportion of case-patients resided on the east side of town. Why might this be?
- *Time*: Does the epidemic curve suggest a need to more carefully examine events from the earlier part of October (i.e., during the period of likely exposure and incubation)?
- *Outlier* cases (usually the first or very last cases in an outbreak): Does the single case that occurred on October 6 have special significance?

8. Would you use a case-control study, a cohort study, or some other method to test the hypotheses in this outbreak? Why?

Answer:
For at least two reasons, a case-control study is the preferred and most efficient method for examining the hypotheses in this outbreak:

- Data are available for only a portion of the total number of cases (referred to as a "case series"), and you don't know what specific exposure may be causing the disease. Consequently, the study must start with disease status. A case-control study allows you to use a comparison group of individuals without disease to evaluate the relationship between the disease and multiple possible exposures.
- The objective of this investigation is to rapidly determine the source of the outbreak in order to institute control measures, and the case-control study can be conducted quickly.

9. What case definition would you use for your case-control study?

Answer:
Because a substantial number of "possible" cases are eventually determined not to be true cases of Legionnaires' disease, we may prefer to use only "confirmed" cases. By including "possible" cases in the calculations, we would produce an inaccurate odds ratio for any exposure factor being analyzed (i.e., the calculated odds ratio would underestimate the true risk of disease).

10. What are some possible sources of controls?

Teacher's Note:
Before asking students to consider sources of controls, review the concept that controls should be drawn from the same population and be as similar as possible to cases, except for the presence of the disease. In other words, controls should be people who would be counted as cases in your study if they became ill.

Answer:
Possible sources of controls:

- Medical facilities: physicians' offices, hospital, etc.
- Acquaintances: family members, neighbors, friends, coworkers
- Community: population-based (e.g., by telephone random-digit dialing or population-based survey)

11a. Calculate the odds ratios for illness with Legionnaires' disease among people who visited Hospital B and those who visited the Post Office.

Answer:
Exposed to Hospital B?

	Case	**Control**	**Total**
Yes	a = 3	b = 7	10
No	c = 25	d = 49	74
Total	28	56	84

OR = (odds of disease among exposed) / (odds of disease among unexposed)
Shortcut formula for OR = AD/BC = (3 x 49) / (7 x 25) = 0.8

Exposed to Post Office?

	Case	**Control**	**Total**
Yes	7	12	19
No	20	38	58
Unknown	1	6	7
Total	28	56	84

OR = (7 x 38) / (20 x 12) = 1.1

11b. Why might the numbers of cases and controls in these two odds ratio calculations differ? How would you interpret the results?

Answer:
The numbers in the tables may differ because of data incompleteness ("Unsure" responses resulting from non-response or problems with memory recall). In calculating the odds ratio for assessing exposure to the Post Office, we have deliberately excluded the "unsure" responses; this decreases the sample size in this calculation by 7 and may decrease the power (ability to detect a statistically significant relationship), especially considering the relatively small numbers in this investigation.

Interpretation of odds ratio for Hospital B: Having visited Hospital B is associated with a reduced odds of illness (0.8). In other words, visiting hospital B was not associated with risk of being ill, but appeared to be protective against Legionnaires' disease in this population.

Interpretation of odds ratio for Post Office: Having visited the Post Office is associated with a slight increase in the odds of becoming ill (1.1).

Teacher's Note:
Emphasize that "Unknowns" need to be excluded from calculations to avoid biasing the results. The power may be decreased, thus decreasing the likelihood of identifying a truly significant association, but arbitrarily assigning unknowns to one category or another opens the door for bias in the analysis. For more information regarding sample size and power, please refer to Appendices B and C.

12. How would you interpret these data? In other words, which exposures suggest an association with illness, which one accounts for the greatest number of cases, and what are the implications?

Teacher's Note:
You may need to prompt students by emphasizing that this investigation is looking for risk factors associated with developing illness. This is easily done by scanning the table for odds ratios >1.0 and p-values <0.05. After identifying these relationships, looking for the statistically significant associations that can account for the greatest number of cases increases the investigator's suspicion that the association is important for further consideration.

Answer:
Among people with outdoor exposure to stores near paper mill cooling towers, the odds ratio was substantially and statistically increased only for those who also reported visiting Butcher Store A (OR = 3.5, p = 0.03). However, this exposure could account for only 12 (44%) of 27 cases. In contrast, exposure to Grocery Store A yielded an odds ratio of 11.6 (p<0.01) and could account for almost all of the cases (25 [93%] of 27). Although, odds ratios were also elevated for exposures to Grocery Store B and Retail Store C, the strength of association for each of these exposures was considerably less than that for exposure to Grocery Store A (i.e., smaller odds ratios and less statistical significance).

Teacher's Note:
Although the exact reason for Grocery Store A's association with illness was not immediately apparent from existing knowledge of Legionella transmission, the implication of Grocery Store A provided an interesting link to the unusual proportion of women noted in the descriptive data. In the late 1980s, women in communities such as Bogalusa usually did most of the grocery shopping, so they would have had a greater risk of exposure if the reservoir for Legionella pneumophila truly was in Grocery Store A.

13a. At this point, do you have enough information to make recommendations—in other words, have the basic criteria of causation been satisfied?

Teacher's Note:
The point of this question is to consider whether the findings of the investigation meet the criteria for causality and justify public health action**. First have the students generate the list of criteria for causality; then discuss whether each criterion is met.**

Answer:

Strength of association: Yes. The odds ratio of 11.6 is both large and statistically significant and can account for most cases.

Biologic plausibility: Maybe. Mist machines had never been implicated in a Legionnaires' outbreak before, but isolation of the organism and the machine's aerosol action make it plausible.

Temporality: Probably. Case-patients and controls were asked about exposures before disease onset; however, we cannot be certain that the mist machine was contaminated at the times of reported exposure.

Dose-response: Yes.

Consistency: No. This is a new finding. We are not aware of similar outbreaks associated with mist machines, although mist machines are widely distributed. (Since Legionnaires' disease was known to be associated with aerosolized water sources, so some students may consider this consistent.)

13b. How would you proceed with this investigation?

Answer:
On balance, the findings are consistent with the hypotheses of risk of illness being related to exposure to Grocery Store A and, within Grocery Store A, to exposure to a contaminated misting device. However, additional studies and steps can be taken to confirm these hypotheses.

14. Who needs to know about these findings? How would you go about reporting the findings?

Answer:
The primary objective of the investigation was to identify the source and mode of transmission in order to develop appropriate measures to control the outbreak and prevent further cases. Thus, we would want to eliminate the risk associated with the mist machines by cleaning or removing them. Since these machines are nationally distributed, cleaning instructions or a product recall will require wide distribution.

We need to inform

- other public health officials in Louisiana and other states,
- the Food and Drug Administration, because of its role in regulating these devices,
- mist machine manufacturers, who produce and distribute these devices, and
- grocery store operators, who use and maintain these devices.

Other groups to inform include

- community residents (through press releases, a town meeting, or a press conference);
- the physicians who initially notified the health department of the problem, and the hospital(s) that treated patients; and
- the medical/public health scientific community (through publications, bulletins, and meetings).

Teacher's Note:
Feedback of information to individuals and groups that have been involved in reporting a potential public health problem is of paramount importance. Not only do they deserve to be kept informed, but appropriate feedback encourages the sense of a "collaborative" relationship and perpetuation of timely reporting in the future.

References

CDC. Legionnaires' disease outbreak associated with a grocery store mist machine - Louisiana, 1989. MMWR 1990;39:108-110.

Mahoney FJ, Hoge C, Farley TF, et al. Legionnaires' disease associated with a grocery store mist machine. J Infect Dis 1992;165:736-9.

Benenson AS, editor. Control of communicable diseases in man.16th edition. American Public Health Association, pp. 256-8.

Appendix A: Predictive Value Theory
(as applied to case definitions)

Actual Status

Study Classification

	True Case	**Non-Case**	
Case	A = True Positive	B = False positive	All People Included as Cases Using Case Definition (A+B)
Non-Case	C = False Negative	D = True Negative	All People Excluded as Cases Using Case Definition (C+D)
	All True Cases(A+C)	All True Non-Cases(B+D)	Total (A+B+C+D)

Sensitivity - the probability that the case definition will identify a true case.

Algebraically, sensitivity = A / (A+C).

Specificity - the probability that the case definition will identify a true non-case.

Algebraically, specificity = D / (B+D).

Predictive-value positive (PVP) - the probability that a person counted as a case is a true case.

Algebraically, PVP = A / (A+B).

Predictive-value negative (PVN) - the probability that a person counted as a non-case is a true non-case.

Algebraically, PVN = D / (C+D).

Teacher's Note:
Although the above definitions have been adapted to the question of "sensitivity" versus "specificity" in case definitions for application to this case study, the concepts can be applied to a variety of disciplines. This "Predictive Value Theory" is often used to evaluate laboratory tests in health disciplines, but it could easily apply to other tests completely unrelated to medicine and public health.

Examples:
a. Laboratory Tests (within the health field): Screening for drug use among workers or for HIV exposure among blood donors will always result in some false positives and some false negatives.
b. SAT Tests (outside the health field): As a means of determining actual ability to do college work, SAT scores will not always be perfect predictors – dividing students according to test scores will result in some false positives and some false negatives when it comes to identifying students with an actual ability to do the work.

Suspected Legionnaires' Disease in Bogalusa
A *Disease Detectives* Exercise from the
Centers for Disease Control and Prevention
Teacher's Guide and Answer Key

Appendix B: Sample Size and Power
"Optional" or "Extra Credit" Exercise

You may recall that the investigators in the Bogalusa case-control study chose to use 2 controls per case; the use of more than one control per case is a means of increasing the sample size in a study to improve the "power" of the study to detect a statistically significant association. Before conducting a study of a small number of cases, it is often useful to calculate the power or ability of a study to detect, at a statistically significant level, a particular odds ratio or difference between cases and controls.

The statistical power of a case-control study is influenced by 5 factors:

1. **n**, the number of cases;

2. **c**, the number of controls per case;

3. **OR**, the odds ratio in the source population worth detecting;

4. **Po**, the proportion of exposed non-cases in the source population;

5. "**α** ("**alpha**"), the desired level of significance. The corresponding 2-tailed **Zα** " from the normal distribution is used in the formulas, e.g., for " **α** = 0.05, **Zα** = 1.96.

The calculation of a study's power involves two steps. First, we calculate **Zβ** ("Z-beta"). Second, we determine the POWER, which is equal to 1-β, by looking up in a table of standard normal cumulative probabilities the cumulative probability associated with that **Zβ**.

A formula for calculating **Zβ**, with n cases and c controls per case, is given by:

$Z_\beta = [n(p_1 - p_0)2 / pq(1 + 1/c)]½ - Z_\alpha$

where $p_1 = p_0 OR / [1 + p_0(OR - 1)]$ = proportion of cases exposed

$p = (p_1 + cp_0) / (1 + c)$ = proportion of all subjects exposed

and $q = 1 - p$

EXAMPLE
Suppose you were designing the case-control study to test the association between exposure to a particular water tower and Legionnaires' disease. You figure that you could

enroll about 50 of the cases, and that about 14% of the town's population is exposed to the water tower in question. You might be able to afford (in terms of time and resources) to enroll 3 controls per case, and you were indoctrinated that " is always 0.05. Calculate the study's power to detect a true odds ratio of 2.0.

Given: $n = 50$, $c = 3$, $p0 = 0.14$, and $OR = 2.0$

$p1 = (0.14)(2.0) / [1 + 0.14(2.0 - 1)] = 0.246$
$p = [0.246 + (3)(0.14)] / (1 + 3) = 0.167$
$q = 1 - 0.167 = 0.834$
$Z\beta = [50(0.246 - 0.14)2 / (0.167)(0.834)(1+ 1/3)]^{½} - 1.96 = -0.221$

POWER $(1-\beta)$ = cumulative probability of -0.221 = 0.413

In other words, a study of 50 cases and 150 controls would be expected a priori (that is, before the study begins and based on the estimated exposure to the water tower of 14%) to have an approximately 41% chance of detecting a statistically significant association in the study, if the underlying association between water tower exposure and Legionnaires' disease in the population were 2.0.

QUESTION B1: Using the formulas above, calculate the power of the study to detect an odds ratio of 2, 3, or 4 at an alpha of 0.05 using 1, 2, 3, 4, or 10 controls per case, as indicated in the table below.

Table. Statistical Power of a Case-Control Study with n=50, p0=0.14, and "=0.05, for different control-to-case ratios and underlying associations

	Control-to-Case Ratio				
	1	2	3	4	10
OR = 2(p1 = 0.246)			0.413		
OR = 3(p1 = 0.328)					
OR = 4(p1 = 0.394)					

QUESTION B2: Discuss the pattern illustrated by the power estimates in the table.

Standard Normal Cumulative Probabilities, Page 1 of 2

* Use this table to find the power which corresponds to Zβ For a given value of Zβ (say, -0.221), find that value to 1 decimal place in the left-most column (-0.2). The power will be in the -0.2 row. Now find the

second decimal of your Zβ across the top row (0.02). The power is in that column. The power is at the intersection of the row and column you've identified (for -0.02 and 0.02, power = 0.41, or 41%).

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129*	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Standard Normal Cumulative Probabilities, Page 2 of 2

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Appendix C: Sample Size and Power Teacher's Guide
Questions and Answers Only for Optional/Extra Credit Exercise

QUESTION B1: Using the formulas above, calculate the power of the study to detect an odds ratio of 2, 3, or 4 at an alpha of 0.05 using 1, 2, 3, 4, or 10 controls per case, as indicated in the table below.

ANSWER B1

Table 2
Statistical Power of a Case-Control Study
with n=50, p0=0.14, and **α**=0.05,
for different control-to-case ratios and underlying associations

	Control-to-Case Ratio				
	1	**2**	**3**	**4**	**10**
OR = 2 (p1 = 0.246)	0.25	0.36	0.41(example)	0.45	0.51
OR = 3 (p1 = 0.328)	0.59	0.76	0.82	0.84	0.88
OR = 4 (p1 = 0.394)	0.84	0.94	0.96	0.97	0.98

QUESTION B2: Discuss the pattern illustrated by the power estimates in the table.

ANSWER B2

a. Given a fixed number of cases, the power of a study is a function of the number of controls and the association one is trying to detect. All else being equal, a study always has more power to detect a stronger association than a weaker one.
b. Given 50 cases, the study has poor power to detect an odds ratio of 2, even with 10 controls per case. However, the study has very good power to detect an odds ratio of 4, even with only one control per case.
c. The table illustrates the general rule that very little power is gained by increasing the control-to-case ratio beyond three or four.

Teacher's Note:
Limited financial, time and personnel resources are frequent problems in choosing the number of controls in a study. However, the two most important factors determining the number of people selected as controls in a case-control study are: 1) statistical "power," and 2) public health urgency:

- First, as illustrated in the above exercise, statistical power in epidemiology is the ability of an analytic study to identify a true association between a risk factor for disease and the disease itself. In general, statistical power is directly related to sample size; the larger the sample, the more likely the study results will be able to demonstrate a statistically significant relationship between the risk factor and the disease outcome. However, there are practical and mathematical limits on the number of controls you select for each case; there is a mathematical "point of diminishing returns" that helps to guide you to the optimal number of controls. As a practical guide, a ratio of four controls for each case is generally considered the maximum ratio before reaching the "point of diminishing returns".
- Second, public health urgency: identification and recruitment of controls takes time; a balance must be struck between your ability to do this and the need to rapidly identify the source of the outbreak.