Case-control studies

In many outbreak settings, the population is not well defined. Therefore, cohort studies are not feasible. However, since cases have been identified in an earlier step of the investigation, the case-control study is ideal. Indeed, case-control studies are more common than cohort studies in the investigation of an outbreak.

As we discussed in Lesson 1, in a case-control study you ask both case-patients and a comparison group of persons without disease ("controls") about their exposures. You then compute a measure of association–an **odds ratio**–to quantify the relationship between exposure and disease. Finally, as in a cohort study, you can compute a chi-square or other test of statistical significance to determine your likelihood of finding this relationship by chance alone.

This method, while not *proving* that a particular exposure caused disease, certainly has served epidemiologists well over time in implicating sources and vehicles associated with disease, and leading them to appropriate control and prevention measures.

Choosing controls. When you design a case-control study, your first, and perhaps most important, decision is who the controls should be. Conceptually, the controls must not have the disease in question, but should represent the population that the cases come from. In other words, they should be similar to the cases except that they don't have the disease. If the null hypothesis were true, the controls would provide us with the level of exposure that you should expect to find among the cases. If exposure is much higher among the cases than the controls, you might choose to reject the null hypothesis in favor of a hypothesis that says exposure is associated with disease.

In practice, it is sometimes difficult to know who the controls should be. Precisely what is the population that the cases came from? In addition, we must consider practical matters, such as how to contact potential controls, gain their cooperation, ensure that they are free of disease, and get appropriate exposure data from them. In a community outbreak, a random sample of the healthy population may, in theory, be the best control group. In practice, however, persons in a random sample may be difficult to contact and enroll. Nonetheless, many investigators attempt to enroll such "population-based" controls through dialing of random telephone numbers in the community or through a household survey.

Other common control groups consist of:

- neighbors of cases
- patients from the same physician practice or hospital who do not have the disease in question
- friends of cases

While controls from these groups may be more likely to participate in the study than randomly identified population-based controls, they may not be as representative of the population. These **biases** in the control group can distort the data in either direction, masking an association between the exposure and disease, or producing a spurious association between an innocent exposure and disease.

In designing a case-control study, you must consider a variety of other issues about controls, including how many to use. Sample size formulas are widely available to help you make this decision. In general, the more subjects (cases and controls) you use in a study, the easier it will be to find an association.

Often, the number of cases you can use will be limited by the size of the outbreak. For example, in a hospital, 4 or 5 cases may constitute an outbreak. Fortunately, the number of potential controls will usually be more than you need. In an outbreak of 50 or more cases, 1 control per case will usually suffice. In smaller outbreaks, you might use 2, 3, or 4 controls per case. More than 4 controls per case will rarely be worth your effort.

As an example, consider again the outbreak of Legionnaires' disease which occurred in Louisiana. Twenty-seven cases were enrolled in a case-control study. The investigators enrolled 2 controls per case, or a total of 54 controls. Using descriptive epidemiology, the investigators did not see any connection with the town's various cooling towers. Using analytic epidemiology, the investigators determined quantitatively that cases and controls were about equally exposed to cooling towers. However, cases were far more likely to shop at Grocery Store A, as shown in the following two-by-two table (6).

Table 6.7
Exposure to Grocery Store A among cases and controls, Legionellosis outbreak, Louisiana, 1990

		Cases	Controls	Total
Shopped at Grocery Store A?	Yes	25	28	53
	No	2	26	28
	Total	27	54	81

In a case-control study, we are unable to calculate attack rates, since we do not know the total number of people in the community who did and did not shop at Grocery Store A. Since we cannot calculate attack rates, we cannot calculate a relative risk. The measure of association of choice in a case-control study is the **odds ratio**. Fortunately, for a rare disease such as legionellosis or most other diseases which cause occasional outbreaks, the odds ratio approximately equals the relative risk we would have found if we had been able to conduct a cohort study.

The odds ratio is calculated as ad / bc. The odds ratio for Grocery Store A is thus 25 x 26 / 28 x 2, or 11.6. These data indicate that persons exposed to Grocery Store A were 11.6 times more likely to develop Legionnaires' disease than persons not exposed to that store!

To test the statistical significance of this finding, we can compute a chi-square test using the following formula:

$$\text{Chi-square} = \frac{T\left[|ad - bc| - (T/2)\right]^2}{V1 \times V2 \times H1 \times H2}$$

For Grocery Store A, the chi-square becomes:

$$= \frac{81 \times \left[|25 \times 26 - 28 \times 2| - 81/2\right]^2}{27 \times 54 \times 53 \times 28}$$

$$= 24{,}815{,}342.25 / 2{,}163{,}672$$

$$= 11.47$$

Referring to Table 6.6, a chi-square of 11.47 corresponds to a p-value less than 0.001. A p-value this small indicates that the null hypothesis is highly improbable, and the investigators rejected the null hypothesis.

Exercise 6.5

You are called to help investigate a cluster of 17 men who developed leukemia in a community. Some of them worked as electrical repair men, and others were ham radio operators. Which study design would you choose to investigate a possible association between exposure to electromagnetic fields and leukemia?

Answers on page 401.

Exercise 6.6

To study rash illness among grocery store workers, investigators conducted a cohort study. The following table shows the data for exposure to celery. What is the appropriate measure of association? Calculate this measure and a chi-square test of statistical significance.

		Rash	No rash	Total	Attack Rate (%)
Exposed to celery?	Yes	25	31	56	44.64
	No	5	65	70	7.14
	Total	30	96	126	23.81

How would you interpret your results?

Answer on page 401.

Step 8: Refining Hypotheses and Executing Additional Studies

Epidemiologic studies

Unfortunately, analytic studies sometimes are unrevealing. This is particularly true if the hypotheses were not well founded at the outset. It is an axiom of field epidemiology that if you cannot generate good hypotheses (by talking to some cases or local staff and examining the descriptive epidemiology and outliers), then proceeding to analytic epidemiology, such as a case-control study, is likely to be a waste of time.

When analytic epidemiology is unrevealing, you need to reconsider your hypotheses. This is the time to convene a meeting of the case-patients to look for common links and to visit their homes to look at the products on their shelves. Consider new vehicles or modes of transmission.

An investigation of an outbreak of *Salmonella muenchen* in Ohio illustrates how a reexamination of hypotheses can be productive. In that investigation, a case-control study failed to implicate any plausible food source as a common vehicle. Interestingly, *all* case-households, but only 41% of control households, included persons 15 to 35 years. The investigators thus began to consider vehicles of transmission to which young adults were commonly exposed. By asking about drug use in a second case-control study, the investigators implicated marijuana as the likely vehicle. Laboratory analysts subsequently isolated the outbreak strain of *S. muenchen* from several samples of marijuana provided by case-patients (24).

Even when your analytic study identifies an association between an exposure and disease, you often will need to refine your hypotheses. Sometimes you will need to obtain more specific exposure histories. For example, in the investigation of Legionnaires' disease (page 380), what about Grocery Store A linked it to disease? The investigators asked cases and controls how much time they spent in the store, and where they went in the store. Using the epidemiologic data, the investigators were able to implicate the ultrasonic mist machine that sprayed the fruits and vegetables. This association was confirmed in the laboratory, where the outbreak subtype of the Legionnaires' disease bacillus was isolated from the water in the mist machine's reservoir (6).

Sometimes you will need a more specific control group to test a more specific hypothesis. For example, in many hospital outbreaks, investigators use an initial study to narrow their focus. They then conduct a second study, with more closely matched controls, to identify a more specific exposure or vehicle. In a large community outbreak of botulism in Illinois, investigators used three sequential case-control studies to identify the vehicle. In the first study, investigators compared exposures of cases and controls from the general public to implicate a restaurant. In a second study they compared restaurant exposures of cases and healthy restaurant patrons to identify a specific menu item, a meat and cheese sandwich. In a third study, investigators used radio broadcast appeals to identify healthy restaurant patrons who had eaten the implicated sandwich. Compared to cases who had also eaten the sandwich, controls were more likely to have avoided the onions that came with the sandwich. Type A Clostridium botulinum was then identified from a pan of leftover sauteed onions used only to make that particular sandwich (17).

Finally, recall that one reason to investigate outbreaks is research, that is, to expand our knowledge. An outbreak may provide an "experiment of nature," which would be unethical for us to set up deliberately, but which we can learn from when it occurs naturally. For example, in the previously described outbreak of hypervitaminosis D in Massachusetts, investigators quickly traced the source to a dairy that was adding too much vitamin D to its milk. After they had instituted the appropriate control measures, the investigators used the "experiment of nature" to characterize the spectrum of health effects caused by overexposure to vitamin D (CDC, unpublished data, 1991). Thus the investigation led to increased knowledge about an unusual problem as well as to prompt action to remove the source.

When an outbreak occurs, whether it is routine or unusual, consider what questions remain unanswered about that particular disease and what kind of study you might do in this setting to answer some of those questions. The circumstances may allow you to learn more about the disease, its modes of transmission, the characteristics of the agent, host factors, and the like. For example, an outbreak of mumps in a highly immunized population may be an opportunity to study vaccine efficacy and duration of protection.

Laboratory and environmental studies

While epidemiology can implicate vehicles and guide appropriate public health action, laboratory evidence can clinch the findings. The laboratory was essential in both the outbreak of salmonellosis linked to marijuana and in the Legionellosis outbreak traced to the grocery store mist machine. You may recall that the investigation of Legionnaires' disease in Philadelphia in 1976 was not considered complete until the new organism was isolated in the laboratory some 6 months later (10).

Environmental studies are equally important in some settings. They are often helpful in explaining **why** an outbreak occurred. For example, in the investigation of the outbreak of shigellosis among swimmers in the Mississippi (Figure 6.7), the local sewage plant was identified as the cause of the outbreak (20). In the study of thyrotoxicosis described earlier, a review of the procedures used in a slaughterhouse near Luverne, Minnesota, identified a practice that caused pieces of the animals' thyroid gland to be included with beef (13). Use a camera to photograph working conditions or environmental conditions. Bring back physical evidence to be analyzed in the laboratory, such as the slabs of beef from the slaughterhouse in the thyrotoxicosis study or the mist machine from the grocery store in the Legionellosis outbreak investigation.

Step 9: Implementing Control and Prevention Measures

In most outbreak investigations, your primary goal will be control and prevention. Indeed, although we are discussing them as Step 9, you should implement control measures as soon as possible. You can usually implement control measures early if you know the source of an outbreak. In general, you aim control measures at the weak link or links in the chain of infection. You might aim control measures at the specific agent, source, or reservoir. For example, an outbreak might be controlled by destroying contaminated foods, sterilizing contaminated water,