Steps of an Outbreak Investigation

In investigating an outbreak, speed is essential, but getting the right answer is essential, too. To satisfy both requirements, epidemiologists approach investigations systematically, using the following 10 steps:

1. Prepare for field work
2. Establish the existence of an outbreak
3. Verify the diagnosis
4. Define and identify cases
5. Describe and orient the data in terms of time, place, and person
6. Develop hypotheses
7. Evaluate hypotheses
8. Refine hypotheses and carry out additional studies
9. Implement control and prevention measures
10. Communicate findings

The steps are presented here in conceptual order. In practice, however, several may be done at the same time, or they may be done in a different order. For example, control measures should be implemented as soon as the source and mode of transmission are known, which may be early or late in any particular outbreak investigation.

Step 1: Prepare for Field Work

Before leaving for the field, you should:

- Research the disease and gather the supplies and equipment you will need
- Make necessary administrative and personal arrangements for such things as travel, and
- Consult with all parties to determine your role in the investigation and who your local contacts will be once you arrive on the scene.

Back to Top

Step 2: Establish the Existence of an Outbreak

One of your first tasks as a field investigator, or disease detective, is to verify that a suspected outbreak is indeed a real outbreak. Some will turn out to be true outbreaks with a common cause, some will be unrelated cases of the same disease, and others will turn out to be unrelated cases of similar but unrelated diseases. Before you can decide whether an outbreak exists (i.e., whether the observed number of cases exceeds the expected number), you must first determine the expected number of cases for the area in the given time frame.

How, then, do you determine what is expected? Usually you can compare the current number of cases with the number from the previous few weeks or months, or from a

comparable period during the previous few years. The sources of these data vary:

- For a notifiable disease (one that, by law, must be reported), you can use health department surveillance records.
- For other diseases and conditions, you can usually find data from local sources such as hospital discharge records, death (mortality) records, and cancer or birth defect registries.
- If local data are not available, you can make estimates using data from neighboring states or national data, or you might consider conducting a telephone survey of physicians to determine whether they have seen more cases of the disease than usual. You could even conduct a survey of people in the community to establish the background level of disease.

Even if the current number of reported cases exceeds the expected number, the excess may not necessarily indicate an outbreak. Reporting may rise because of changes in local reporting procedures, changes in the case definition, increased interest because of local or national awareness, or improvements in diagnostic procedures. For example, if a new physician, infection control nurse, or health care facility is reporting cases more consistently than they were reported in the past, the numbers would go up even though there might be no change in the actual occurrence of the disease. Finally, particularly in areas with sudden changes in population size, such as resort areas, college towns, and migrant farming areas, changes in the number of reported cases may simply reflect changes in the size of the population.

Whether or not you should investigate an apparent problem further is not strictly tied to your verifying that an epidemic exists (that is, that the observed number is greater than the number expected). As noted earlier, other factors may come into play, including, for example, the severity of the illness, the potential for spread, political considerations, public relations, and the availability of resources.

Back to Top

Step 3: Verify the Diagnosis

In addition to verifying the existence of an outbreak early in the investigation, you must also identify as accurately as possible the specific nature of the disease. Your goals in verifying the diagnosis are two-fold. First, you must ensure that the problem has been properly diagnosed—that it really is what it has been reported to be. Second, for outbreaks involving infectious or toxic-chemical agents, you must to be certain that the increase in diagnosed cases is not the result of a mistake in the laboratory.

Verifying the diagnosis requires that you review the clinical findings (the symptoms and features of illness) and laboratory results for the people who are affected. If you are at all uncertainty about the laboratory findings (e.g., if they are inconsistent with the clinical findings), you should have a laboratory technician review the techniques being used. If you expect a need for specialized laboratory work (e.g., special culturing or DNA analysis), you should begin obtaining the appropriate specimens, isolates, and other laboratory material from a sufficient number of patients as soon as possible.

Finally, you should visit several of the people who became ill. If you do not have the clinical background to verify the diagnosis, a doctor or other qualified clinician should do so. Regardless of your background, though, you should see and talk to some of these people to gain a better understanding of the disease and those affected by it. In addition, you may be able to gather critical information by asking such questions as, What were their exposures before becoming ill? What do they think caused their illness? Do they know anyone else with the disease? Do they have anything in common with others who have the disease? Conversations with patients are very helpful in generating hypotheses about the cause, source, and spread of disease.

Back to Top

Step 4: Define and Identify Cases

Establish a case definition. Your next task as an investigator is to establish a case definition, or a standard set of criteria for deciding whether, in this investigation, a person should be classified as having the disease or health condition under study. A case definition usually includes four components:

1. clinical information about the disease,
2. characteristics about the people who are affected,
3. information about the location or place, and
4. a specification of time during which the outbreak occurred.

You should base the clinical criteria on simple and objective measures. For example, you might require the presence of an elevated level of antibody to the disease agent, the presence of a fever of at least 101°F, three or more loose bowel movements per day, or muscle aching severe enough to limit the patient's activities. Regarding the characteristics of people, you might restrict the definition to those who attended a wedding banquet, or ate at a certain restaurant, or swam in the same lake. By time, the criterion might be onset of illness within the past 2 months; by place, it might be living in a nine-county area or working at a particular plant. Whatever your criteria, you must apply them consistently and without bias to all of the people included in the investigation.

Ideally, your case definition should be broad enough to include most, if not all, of the actual cases, without capturing what are called "false-positive" cases (when the case definition is met, but the person actually does not have the disease in question). Recognizing the uncertainty of some diagnoses, investigators often classify cases as "confirmed," " probable," or "possible."

To be classified as confirmed, a case usually must have laboratory verification. A case classified as probable usually has the typical clinical features of the disease without laboratory confirmation. A possible case usually has fewer of the typical clinical features. For example, in an outbreak of bloody diarrhea and severe kidney disease (hemolytic-uremic syndrome) caused by infection with the bacterium *E. coli* O157:H7, investigators defined cases in the following three classes:

- **Confirmed case:** *E. coli* O157:H7 isolated from a stool culture or development of hemolytic-uremic syndrome in a school-aged child resident of the county and who had gastrointestinal symptoms beginning between Nov. 3 and Nov. 8, 1990;
- **Probable case:** Bloody diarrhea (but no culture), with the same person, place, and time restrictions;
- **Possible case:** Abdominal cramps and diarrhea (at least three stools in a 24-hour period) in a school-age child resident of the county with onset during the same period (CDC, unpublished data, 1991).

Early in an investigation, a loose case definition that includes confirmed, probable, and even possible cases is often used to allow investigators to capture as many cases as possible. Later on, when hypotheses have come into sharper focus, the investigator may tighten the case definition by dropping the "possible" category. This strategy is particularly useful when you have to travel to different hospitals, homes, or other places to gather information, because it keeps you from having to go back for additional data. This illustrates an important axiom of field epidemiology: "Get it while you can."

Identify and count cases

As noted above, many outbreaks are first recognized and reported by concerned health care providers or citizens. However, the first cases to be recognized usually are only a small proportion of the total number. As a Disease Detective investigating an outbreak, you must therefore "cast the net wide" to determine the true size and geographic extent of the problem.

When identifying cases, you should use as many sources as you can, and you may need to be creative and aggressive in identifying these sources. Initially, you may want to direct your case finding at health care facilities where the diagnosis is likely to be made; these facilities include physicians' offices, clinics, hospitals, and laboratories. You also may decide to send out a letter describing the situation and asking for reports (passive surveillance); or you may decide to telephone or visit the facilities to collect information (active surveillance).

In some outbreaks, public health officials may decide to alert the public directly, usually through the local media. For example, in outbreaks caused by a contaminated food product such as salmonellosis caused by contaminated milk (7) or L-tryptophan-induced EMS (8), announcements in the media have alerted the public to avoid the implicated product and to see a physician if they had symptoms of the disease.

If an outbreak affects a population in a restricted setting, such as a cruise ship, school, or worksite, and if a high proportion of cases are unlikely to be diagnosed (if, for example, many cases are mild or asymptomatic), you may want to conduct a survey of the entire population. In such settings, you could administer a questionnaire to determine the true occurrence of clinical symptoms, or you could collect laboratory specimens to determine the number of asymptomatic cases. Finally, you can ask people who are affected if they know anyone else with the same condition.

Regardless of the particular disease you are investigating, you should collect the following types of information about every person affected:

- **Identifying information:** This may include name, address, and telephone number and allows you and other investigators to contact patients for additional questions and to notify them of laboratory results and the outcome of the investigation. Addresses also allow you to map the geographic extent of the problem.
- **Demographic information:** This may include age, sex, race, and occupation and provides the details that you need to characterize the population at risk.
- **Clinical information:** This information allows you to verify that the case definition has been met. Date of onset allows you to create a graph of the outbreak. Supplementary clinical information may include whether the person was hospitalized or died and will help you describe the spectrum of illness.
- **Risk factor information:** Information about risk factors will allow you to tailor your investigation to the specific disease in question. For example, in an investigation of hepatitis A, you would look at exposure to food and water sources.

Traditionally, we collect the information described above on a standard case report form, questionnaire, or data abstraction form. We then abstract selected critical items in a table called a "line listing." In a line listing, each column represents an important variable, such as name or identification number, age, sex, and case classification, while each row represents a different case, by number. New cases are added to a line listing as they are identified. This simple format allows the investigator to scan key information on every case and update it easily. Even in the era of microcomputers, many epidemiologists still maintain a hand-written line listing of key data items and turn to their computers for more complex manipulations of data. Here is a portion of a line listing that might have been created for an outbreak of hepatitis A.

					Diagnostic							Lab		
					Signs and Symptoms									
Case#	Initials	Date of Report	Date of Onset	Physician Diagnosis	N	V	A	F	DU	J	HAIgM	Other	Age	Sex
1	JG	10/12	12/6	Hep A	+	+	+	+	+	+	+	SGOT ↓	37	M

2	BC	10/12	10/5	Hep A	+	-	+	+	+	+	+	Alt↓	62	F
3	HP	10/13	10/4	Hep A	±	-	+	+	+	S*	+	SGOT ↓	30	F
4	MC	10/15	10/4	Hep A	-	-	+	+	?	-	+	Hbs/ Ag-	17	F
5	NG	10/15	10/9	NA	-	-	+	-	+	+	NA	NA	32	F
6	RD	10/15	10/8	Hep A	+	+	+	+	+	+	+		38	M
7	KR	10/16	10/13	Hep A	±	-	+	+	+	+	+	SGOT = 240	43	M

S*=Sclera;, N=Nausea; V=Vomiting; A=Anorexia; F=Fever; DU=Dark urine; J=Jaundice; HAIgm=Hepatitis AIgM antibody test

Back to Top

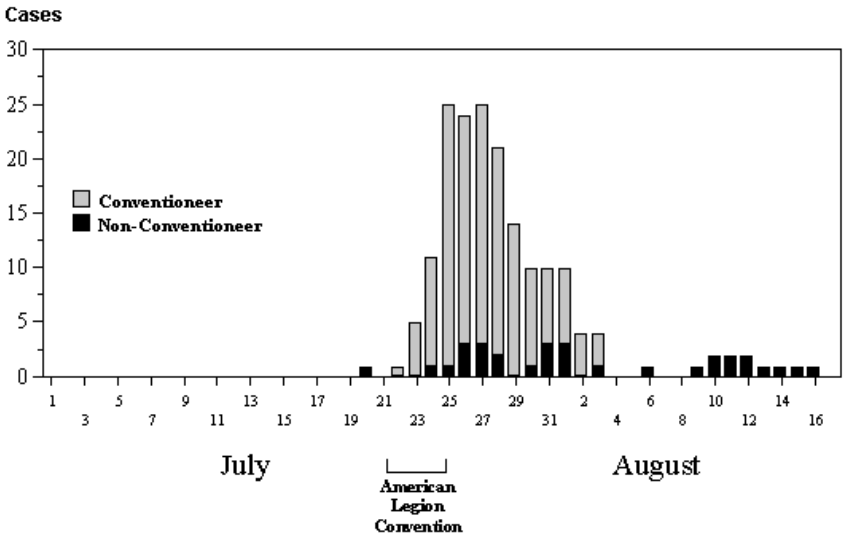
Step 5: Describe and Orient the Data in Terms of Time, Place, and Person

Once you have collected some data, you can begin to characterize an outbreak by time, place, and person. In fact, you may perform this step several times during the course of an outbreak. Characterizing an outbreak by these variables is called **descriptive epidemiology**, because you describe what has occurred in the population under study. This step is critical for several reasons. First, by becoming familiar with the data, you can learn what information is reliable and informative (e.g., the same unusual exposure reported by many of the people affected) and what may not be as reliable (e.g., many missing or "don't know" responses to a particular question). Second, you provide a comprehensive description of an outbreak by showing its trend over time, its geographic extent (place), and the populations (people) affected by the disease. This description lets you begin to assess the outbreak in light of what is known about the disease (e.g., the usual source, mode of transmission, risk factors, and populations affected) and to develop causal hypotheses. You can, in turn, test these hypotheses using the techniques of analytic epidemiology described later in **Step 7: Evaluate Hypotheses.**

Note that you should begin descriptive epidemiology early and should update it as you collect additional data. To keep an investigation moving quickly and in the right direction, you must discover both errors and clues in the data as early as possible.

Characterizing by time
Traditionally, we show the time course of an epidemic by drawing a graph of the number of cases by their date of onset. This graph, called an **epidemic curve**, or "epi curve" for short, gives a simple visual display of the outbreak's magnitude and time trend. The following example depicts the first outbreak of Legionnaires' disease, in Philadelphia, Pennsylvania, in 1976.

Insert EPI Curve
An epidemic curve provides a great deal of information. First, you will usually be able to tell where you are in the course of the epidemic, and possibly to project its future course. Second, if you have identified the disease and know its usual incubation period, you may be able to estimate a probable time period of exposure and can then develop a questionnaire focusing on that time period. Finally, you may be able to draw inferences about the epidemic pattern—for example, whether it is an outbreak resulting from a common source exposure, from person-to-person spread, or both.



How to draw an epidemic curve

To draw an epidemic curve, you first must know the time of onset of illness for each person. For most diseases, date of onset is sufficient; however, for a disease with a very short incubation period, hours of onset may be more suitable. The number of cases is plotted on the *y*-axis of an epi curve; the unit of time, on the *x*-axis. We usually base the units of time on the incubation period of the disease (if known) and the length of time over which cases are distributed. As a rule of thumb, select a unit that is one-fourth to one-third as long as the incubation period. Thus, for an outbreak of *Clostridium perfringens* food poisoning (usual incubation period 10-12 hours), with cases during a period of only a few days, you could use an *x*-axis unit of 2 or 3 hours. Unfortunately, there will be times when you do not know the specific disease and/or its incubation period. In that circumstance, it is useful to draw several epidemic curves, using different units on the *x*-axes, to find one that seems to show the data best. Finally, show the pre- and post-epidemic period on your graph to illustrate the activity of the disease during those periods.

Interpreting an epidemic curve

The first step in interpreting an epidemic curve is to consider its overall shape, which will be determined by the pattern of the epidemic (e.g., whether it has a common source or person-to-person transmission), the period of time over which susceptible people are exposed, and the minimum, average, and maximum incubation periods for the disease.

An epidemic curve with a steep up slope and a gradual down slope, such as the illustration above on the first outbreak of Legionnaires'disease, indicates a single source (or "point source") epidemic in which people are exposed to the same source over a relatively brief period. In fact, any sudden rise in the number of cases suggests sudden exposure to a common source. In a point source epidemic, all the cases occur within one incubation period. If the duration of exposure is prolonged, the epidemic is called a "continuous common source epidemic," and the epidemic curve will have a plateau instead of a peak. Person-to-person spread (a "propagated" epidemic) should have a series of progressively taller peaks one incubation period apart.

Cases that stand apart (called "outliers") may be just as informative as the overall pattern. An early case may represent a background (unrelated) case, a source of the epidemic, or a person who was exposed earlier than most of the people affected (e.g., the cook who tasted her dish hours before bringing it to the big picnic). Similarly, late cases may be unrelated to the outbreak, may have especially long incubation periods, may indicate exposure later than most of the people affected, or may be secondary cases (that is, the person may have become ill after being exposed to someone who was part of the initial outbreak). All outliers are worth examining carefully because if they are part of the outbreak, their unusual exposures may point directly to the source. For a disease with a human host such as hepatitis A, for instance, one of the

early cases may be in a food handler who is the source of the epidemic.

In a point-source epidemic of a known disease with a known incubation period, you can use the epidemic curve to identify a likely period of exposure. This is critical to asking the right questions to identify the source of the epidemic.

Characterizing by place
Assessment of an outbreak by place provides information on the geographic extent of a problem and may also show clusters or patterns that provide clues to the identity and origins of the problem. A simple and useful technique for looking at geographic patterns is to plot, on a "spot map" of the area, where the affected people live, work, or may have been exposed.

A spot map of cases in a community may show clusters or patterns that reflect water supplies, wind currents, or proximity to a restaurant or grocery store. On a spot map of a hospital, nursing home, or other such facility, clustering usually indicates either a focal source or person-to-person spread, while the scattering of cases throughout a facility is more consistent with a common source such as a dining hall. In studying an outbreak of surgical wound infections in a hospital, we might plot cases by operating room, recovery room, and ward room to look for clustering.

If the size of the overall population varies between the areas you are comparing, a spot map, because it shows numbers of cases, can be misleading. This is a weakness of spot maps. In such instances, you should show the proportion of people affected in each area (which would also represent the rate of disease or, in the setting of an outbreak, the "attack rate").

Characterizing by person
You determine what populations are at risk for the disease by characterizing an outbreak by person. We usually define such populations by personal characteristics (e.g., age, race, sex, or medical status) or by exposures (e.g., occupation, leisure activities, use of medications, tobacco, drugs). These factors are important because they may be related to susceptibility to the disease and to opportunities for exposure.

Age and sex are usually assessed first, because they are often the characteristics most strongly related to exposure and to the risk of disease. Other characteristics will be more specific to the disease under investigation and the setting of the outbreak. For example, if you were investigating an outbreak of hepatitis B, you should consider the usual high-risk exposures for that infection, such as intravenous drug use, sexual contacts, and health care employment.

Summarizing by time, place, and person
After characterizing an outbreak by time, place, and person, you need to summarize what you know to see whether your initial hypotheses are on track. You may find that you need to develop new hypotheses to explain the outbreak.

Back to Top

Step 6: Develop Hypotheses

In real life, we usually begin to generate hypotheses to explain why and how the outbreak occurred when we first learn about the problem. But at this point in an investigation, after you have interviewed some affected people, spoken with other health officials in the community, and characterized the outbreak by time, place, and person, your hypotheses will be sharpened and more accurately focused. The hypotheses should address the source of the agent, the mode (vehicle or vector) of transmission, and the exposures that caused the disease. Also, the hypotheses should be proposed in a way that can be tested.

You can develop hypotheses in a variety of ways. First, consider what you know about the disease itself: What is the agent's usual reservoir? How is it usually transmitted?

What vehicles are commonly implicated? What are the known risk factors? In other words, simply by becoming familiar with the disease, you can, at the very least, "round up the usual suspects."

Another useful way to generate hypotheses is to talk to a few of the people who are ill, as discussed under **Step 3: Verifying the Diagnosis**. Your conversations about possible exposures should be open-ended and wide-ranging and not confined to the known sources and vehicles. Sometimes investigators meet with a group of the affected people as a way to search for common exposures. Investigators have even found it useful to visit the homes of people who became ill and look through their refrigerators and shelves for clues.

Descriptive epidemiology often provides some hypotheses. If the epidemic curve points to a narrow period of exposure, ask what events occurred around that time. If people living in a particular area have the highest attack rates, or if some groups with particular age, sex, or other personal characteristics are at greatest risk, ask why. Such questions about the data should lead to hypotheses that can be tested.

Back to Top

Step 7: Evaluate Hypotheses

The next step is to evaluate the credibility of your hypotheses. There are two approaches you can use, depending on the nature of your data: 1) comparison of the hypotheses with the established facts and 2) **analytic epidemiology**, which allows you to test your hypotheses.

You would use the first method when your evidence is so strong that the hypothesis does not need to be tested. A 1991 investigation of an outbreak of vitamin D intoxication in Massachusetts is a good example. All of the people affected drank milk delivered to their homes by a local dairy. Investigators hypothesized that the dairy was the source, and the milk was the vehicle of excess vitamin D. When they visited the dairy, they quickly recognized that far more than the recommended dose of vitamin D was inadvertently being adding to the milk. No further analysis was necessary.

The second method, analytic epidemiology, is used when the cause is less clear. With this method, you test your hypotheses by using a comparison group to quantify relationships between various exposures and the disease. There are two types of analytic studies: **cohort studies** and **case-control studies**. Cohort studies compare groups of people who have been exposed to suspected risk factors with groups who have not been exposed. Case-control studies compare people with a disease (case-patients) with a group of people without the disease (controls). The nature of the outbreak determines which of these studies you will use.

Cohort studies

A cohort study is the best technique for analyzing an outbreak in a small, well-defined population. For example, you would use a cohort study if an outbreak of gastroenteritis occurred among people who attended a social function, such as a wedding, and a complete list of wedding guests was available. In this situation, you would ask each attendee the same set of questions about potential exposures (e.g., what foods and beverages he or she had consumed at the wedding) and whether he or she had become ill with gastroenteritis.

After collecting this information from each guests, you would be able to calculate an attack rate for people who ate a particular item (were exposed) and an attack rate for those who did not eat that item (were not exposed). For the exposed group, the attack rate is found by dividing the number of people who ate the item and became ill by the total number of people who ate that item. For those who were not exposed, the attack rate is found by dividing the number of people who did not eat the item but still became ill by the total number of people who did not eat that item.

To identify the source of the outbreak from this information, you would look for an item with:

- a high attack rate among those exposed *and*
- a low attack rate among those not exposed (so the difference or ratio between attack rates for the two exposure groups is high); *in addition*
- most of the people who became ill should have consumed the item, so that the exposure could explain most, if not all, of the cases.

Usually, you would also calculate the mathematical association between exposure (consuming the food or beverage item) and illness for each food and beverage. This is called the **relative risk** and is produced by dividing the attack rate for people who *were exposed* to the item by the attack rate for those who *were not exposed*.

The table on the next page is based on a famous outbreak of gastroenteritis following a church supper in Oswego, New York, in 1940 and illustrates the use of a cohort study (9). Of the 80 people who attended the supper, 75 were interviewed. Forty-six people met the case definition. Attack rates for those who did and did not eat each of 14 items are presented in the table. Scan the column of attack rates among those who ate the specified items. Which item shows the highest attack rate? Did most of the 46 people who met the case definition eat that food item? Is the attack rate low among people who did not eat that item? You should have identified vanilla ice cream as the implicated vehicle, or source. The relative risk is calculated as 80 / 14, or 5.7. This relative risk indicates that people who ate the vanilla ice cream were 5.7 times more likely to become ill than were those who did not eat the vanilla ice cream.

**Attack Rates by Items Served at a Church Supper,
Oswego, New York, April 1940**

	Number of people who **ate** specified item				Number of people who **did not eat** specified item			
Food	Ill	Well	Total	Attack Rate %	Ill	Well	Total	Attack Rate %
Baked Ham	29	17	46	63	17	12	29	59
Spinach	26	17	43	60	20	12	32	62
Mashed potatoes*	23	14	37	62	23	14	37	62
Cabbage salad	18	10	28	64	28	19	47	60
Jell-O	16	7	23	70	30	22	52	58
Rolls	21	16	37	57	25	13	38	66
Brown bread	18	9	27	67	28	20	48	58
Milk	2	2	4	50	44	27	71	62
Coffee	19	12	31	61	27	17	44	61
Water	13	11	24	54	33	18	51	65
Cakes	27	13	40	67	19	16	35	54
Ice Cream (van)	43	11	54	80	3	18	21	14
Ice Cream (choc)*	25	22	47	53	20	7	27	74

Fruit Salad	4	2	6	67	42	27	69	61

*Excludes 1 person with indefinite history of consumption of that food. *Source:* 9

Case-control studies

In most outbreaks the population is not well defined, and so cohort studies are not feasible. In these instances, you would use the case-control study design. In a case-control study, you ask both case-patients and controls about their exposures. You then can calculate a simple mathematical measure of association—called an **odds ratio**—to quantify the relationship between exposure and disease. This method does not prove that a particular exposure caused a disease, but it is very helpful and effective in evaluating possible vehicles of disease.

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 be from the same population as the case-patients. In other words, they should be similar to the case-patients except that they do not have the disease. Common control groups consist of neighbors and friends of case-patients and people from the same physician practice or hospital as case-patients.

In general, the more case-patients and controls you have, the easier it will be to find an association. Often, however, you are limited because the outbreak is small. 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-patient will usually suffice. In smaller outbreaks, you might use 2, 3, or 4 controls per case-patient. More than 4 controls per case-patient will rarely be worth your effort.

In a case-control study, you cannot calculate attack rates because you do not know the total number of people in the community who were and were not exposed to the source of the disease under study. Without attack rates, you cannot calculate relative risk; instead, the measure of association you use in a case study is an odds ratio. When preparing to calculate an odds ratio, it is helpful to look at your data in a 2×2 table. For instance, suppose you were investigating an outbreak of hepatitis A in a small town, and you suspected that the source was a favorite restaurant of the townspeople. After questioning case-patients and controls about whether they had eaten at that restaurant, your data might look like this:

		Case Patients	Controls	Total
Ate at Restaurant A?	Yes	a = 30	b = 36	66
	No	c = 10	d = 70	80
Total:		40	106	146

The odds ratio is calculated as ad/bc. The odds ratio for Restaurant A is thus 30 × 70 / 36 × 10, or 5.8. This means that people who ate at Restaurant A were 5.8 times more likely to develop hepatitis A than were people who did not eat there. Even so, you could not conclude that Restaurant A was the source without comparing its odds ratio with the odds ratios for other possible sources. It could be that the source is elsewhere and that it just so happens that many of the people who were exposed also ate at Restaurant A.

Testing statistical significance

The final step in testing your hypothesis is to determine how likely it is that your study results could have occurred by chance alone. In other words, how likely is it that the exposure your study results point to as the source of the outbreak was not related to the disease after all? A test of statistical significance is used to evaluate this likelihood.

Statistical significance is a broad area of study, and we will include only a brief overview here.

The first step in testing for statistical significance is to assume that the exposure is not related to disease. This assumption is known as the **null hypothesis**. Next, you compute a measure of association, such as a relative risk or an odds ratio. These measures are then used in calculating a chi-square test (the statistical test most commonly used in studying an outbreak) or other statistical test. Once you have a value for chi-square, you look up its corresponding p-value (or probability value) in a table of chi-squares.

In interpreting p-values, you set in advance a cutoff point beyond which you will consider that chance is a factor. A common cutoff point is .05. When a p-value is below the predetermined cutoff point, the finding is considered "statistically significant," and you may reject the null hypothesis in favor of the **alternative hypothesis**, that is you may conclude that the exposure is associated with disease. The smaller the p-value, the stronger the evidence that your finding is statistically significant.

Back to Top

Step 8: Refine Hypotheses and Carry Out Additional Studies

Additional epidemiological studies
When analytic epidemiological studies do not confirm your hypotheses, you need to reconsider your hypotheses and look for new vehicles or modes of transmission. This is the time to meet with case-patients to look for common links and to visit their homes to look at the products on their shelves.

An investigation of an outbreak of *Salmonella muenchen* in Ohio during 1981 illustrates this point. A case-control study failed to turn up a food source as a common vehicle. Interestingly, people 15 to 35 years of age lived in all of the households with cases, but in only 41% of control households. This difference caused the investigators to consider vehicles of transmission to which young adults might be exposed. By asking about drug use in a second case-control study, the investigators found that illegal use of marijuana was the likely vehicle. Laboratory analysts subsequently isolated the outbreak strain of *S. muenchen* from several samples of marijuana provided by case-patients (10).

Even when your analytic study identifies an association between an exposure and a disease, you often will need to refine your hypotheses. Sometimes you will need to obtain more specific exposure histories or a more specific control group. For example, 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 case-patients and controls from the general public and implicated a restaurant. In a second study, they compared the menu items eaten by the case-patients with those eaten by healthy restaurant patrons and identified a specific menu item, a meat and cheese sandwich. In a third study, appeals were broadcast over radio to identify healthy restaurant patrons who had eaten the sandwich. It turned out that controls were less likely than case-patients to have eaten the onions that came with the sandwich. Type A *Clostridium botulinum* was then identified from a pan of leftover sautéed onions used only to make that particular sandwich (11).

When an outbreak occurs, whether it is routine or unusual, you should consider what questions remain unanswered about the disease and what kind of study you might use in the particular setting to answer some of these questions. The circumstances may allow you to learn more about the disease, its modes of transmission, the characteristics of the agent, and host factors.

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 the outbreak of salmonellosis linked to use of contaminated marijuana. The investigation of the outbreak of Legionnaires' disease in Philadelphia mentioned earlier was not considered complete until the new organism was isolated in the laboratory over 6 months after the outbreak actually had occurred (12). Environmental studies often help explain why an outbreak occurred and may be very important in some settings. For example, in an investigation of an outbreak of shigellosis among swimmers in the Mississippi River, a local sewage plant was identified as the cause of the outbreak (13).

Back to Top

Step 9: Implementing Control and Prevention Measures

Even though implementing control and prevention measures is listed as Step 9, in a real investigation you should do this as soon as possible. Control measures, which can be implemented early if you know the source of an outbreak, should be aimed at specific links in the chain of infection, the agent, the source, or the reservoir. For example, an outbreak might be controlled by destroying contaminated foods, sterilizing contaminated water, destroying mosquito breeding sites, or requiring an infectious food handler to stay away from work until he or she is well.

In other situations, you might direct control measures at interrupting transmission or exposure. For example, to limit the airborne spread of an infectious agent among residents of a nursing home, you could use the method of "cohorting" by putting infected people together in a separate area to prevent exposure to others. You could instruct people wishing to reduce their risk of acquiring Lyme disease to avoid wooded areas or to wear insect repellent and protective clothing. Finally, in some outbreaks, you would direct control measures at reducing susceptibility. Two such examples are immunization against rubella and malaria chemoprophylaxis (prevention by taking antimalarial medications) for travelers.

Back to Top

Step 10: Communicate Findings

Your final task in an investigation is to communicate your findings to others who need to know. This communication usually takes two forms: 1) an oral briefing for local health authorities and 2) a written report.

Your oral briefing should be attended by the local health authorities and people responsible for implementing control and prevention measures. This presentation is an opportunity for you to describe what you did, what you found, and what you think should be done about it. You should present your findings in scientifically objective fashion, and you should be able to defend your conclusions and recommendations.

You should also provide a written report that follows the usual scientific format of introduction, background, methods, results, discussion, and recommendations. By formally presenting recommendations, the report provides a blueprint for action. It also serves as a record of performance, a document for potential legal issues, and a reference if the health department encounters a similar situation in the future. Finally, a report that finds its way into the public health literature serves the broader purpose of contributing to the scientific knowledge base of epidemiology and public health.

Back to Top

Back to How to Investigate an Outbreak

This page last reviewed November 17, 2004