NMCP Infection Control
Outbreak Investigation Outline

Steps of an Outbreak Investigation

In investigating an outbreak, speed is essential, but getting the right answer is essential, too. To satisfy both requirements, outbreak investigations must be approached 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 in this policy 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, one should:

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

Step 2: Establish the Existence of an Outbreak

One of the first tasks 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 it can decided whether an outbreak exists (i.e., whether the observed number of cases exceeds the expected number), one must first determine the expected number of cases for the area in the given time frame.

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), use health department surveillance records.
- For other diseases and conditions, 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, make estimates using data from neighboring states or national data, or consider conducting a telephone survey of physicians to determine whether they have seen more cases of the disease than usual. It may be necessary to 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 an apparent problem is further investigated is not strictly tied to 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.

Step 3: Verify the Diagnosis

In addition to verifying the existence of an outbreak early in the investigation, identity of the specific nature of the disease must be established as accurately as possible. Goals in verifying the diagnosis are two-fold. First, 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, be certain that the increase in diagnosed cases is not the result of a mistake in the laboratory.

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

Finally, visitation of several of the people who became ill is necessary. A doctor or other qualified clinician with a clinical background should perform the visitations to verify the diagnosis. Seeing and listening to as many of these people as possible is necessary to gain a better understanding of the disease and those affected by it. In addition, much critical information can be obtained by asking such questions as, “What were your exposures before becoming ill”? “What do you think caused your illness"? "Do you know anyone else with the disease"? "Do you 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.

Step 4: Define and Identify Cases

Establish a case definition. The next task 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.

The clinical criteria should be based on simple and objective measures. For example, one 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, one 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 the criteria, apply them consistently and without bias to all of the people included in the investigation.

Ideally, the 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 the capture of as many cases as possible. Later on, when hypotheses have come into sharper focus, the case definition will be tightened by dropping the "possible" category. This strategy is particularly useful when travel to different hospitals, homes, or other places to gather information is necessary, because it keeps 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. "Cast the net wide" to determine the true size and geographic extent of the problem.

When identifying cases, use as many sources as possible. Initially, direct case finding at health care facilities where the diagnosis is likely to be made; these facilities include physicians' offices, clinics, hospitals, and laboratories. It could be decided to send out a letter describing the situation and asking for reports (passive surveillance); or it may be decided 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.

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), a survey may be conducted of the entire population. In such settings, a questionnaire may be administered to determine the true occurrence of clinical symptoms, or collection of laboratory specimens to determine the number of asymptomatic cases may need to be performed. Finally, ask people who are affected if they know anyone else with the same condition.

Regardless of the particular disease being investigated, collect the following types of information about every person affected:

- **Identifying information:** This may include name, address, and telephone number and allows the ability to contact patients for additional questions and to notify them of laboratory results and the outcome of the investigation. Addresses also assist in mapping the geographic extent of the problem.
- **Demographic information:** This may include age, sex, race, and occupation and provides the details needed to characterize the population at risk.
- **Clinical information:** This information allows verification that the case definition has been met. Date of onset allows the creation of a graph of the outbreak. Supplementary clinical information may include whether the person was hospitalized or died and will help describe the spectrum of illness.
- **Risk factor information:** Information about risk factors will allow tailoring the investigation to the specific disease in question.

Traditionally, the information described above is collected on a standard case report form, questionnaire, or data abstraction form. Then selected critical items are abstracted 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 for scanning key information on every case and to update it easily.

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

Once some data are collected, begin to characterize an outbreak by time, place, and person. In fact, this step may performed several times during the course of an outbreak. Characterizing an outbreak by these variables is called **descriptive epidemiology**, because one describes what has occurred in the population under study. This step is critical for several reasons. First, by becoming familiar with the data, one 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, it provides 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 allows for the beginning of assessment of 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. In turn, these hypotheses can be tested using the techniques of analytic epidemiology described later in **Step 7: Evaluate Hypotheses.**

Begin descriptive epidemiology early and update it as additional data are collected. To keep an investigation moving quickly and in the right direction, discover both errors and clues in the data as early as possible.

Characterizing by time

Traditionally, 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.

Insert EPI Curve

An epidemic curve provides a great deal of information. First, it is usually possible to tell where in the course of the epidemic one is, and possibly to project its future course. Second, if the disease has been identified and its usual incubation period is known, an estimate of a probable time period of exposure can be obtained and then a questionnaire focusing on that time period can be developed. Finally, inferences can be drawn 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, one 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. 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 onefourth to one-third as long as the incubation period. Unfortunately, there will be times when the specific disease and/or its incubation period is not known. 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 the 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 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, 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, 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 being compared, a spot map, because it shows numbers of cases, can be misleading. This is a weakness of spot maps. In such instances, 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
Determine what populations are at risk for the disease by characterizing an outbreak by person. 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 one were investigating an outbreak of hepatitis B, one 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, summarize what is known to see whether the initial hypotheses are on track. A new hypothesis may need to be developed to explain the outbreak.

Step 6: Develop Hypotheses

In real life, hypotheses are generated to explain why and how the outbreak occurred when a problem is first discovered. But at this point in an investigation, after some affected people have been interviewed, other health officials in the community have been contacted, and the outbreak has be characterized by time, place, and person, the 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.

Hypotheses are developed in a variety of ways. First, consider what is known 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, one 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**. Conversations about possible exposures should be open-ended and wide-ranging and not confined to the known sources and vehicles. Sometimes it may be necessary to meet with a group of the affected people as a way to search for common exposures. It may be 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.

Step 7: Evaluate Hypotheses

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

The first method would be used when evidence is so strong that the hypothesis does not need to be tested.

The second method, analytic epidemiology, is used when the cause is less clear. With this method, test the 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 will be used.

Cohort studies
A cohort study is the best technique for analyzing an outbreak in a small, well-defined population. For example, a cohort study would be used 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, 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, 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.

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*.

Case-control studies
In most outbreaks the population is not well defined, and so cohort studies are not feasible. In these instances, use the case-control study design. In a case-control study, ask both case-patients and controls about their exposures. Then can a simple mathematical measure of association—called an **odds ratio**—can be calculated 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 designing a case-control study, the 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, the easier it will be to find an association. Often, however, these 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 needed. In an outbreak of 50 or more cases, 1 control per case-patient will usually suffice. In smaller outbreaks, 2, 3, or 4 controls per case-patient might be used. More than 4 controls per case-patient will rarely be worth the effort.

In a case-control study, attack rates cannot be calculated because the total number of people in the community who were and were not exposed to the source of the disease under study is not known. Without attack rates, relative risk cannot be calculated; instead, the measure of association used in a case study is an odds ratio. When preparing to calculate an odds ratio, it is helpful to look at the data in a 2×2 table. For instance, suppose one were investigating an outbreak of hepatitis A in a small town, and the source was a favorite restaurant of the townspeople was suspected.

Testing statistical significance
The final step in testing the hypothesis is to determine how likely it is that the study results could have occurred by chance alone. In other words, how likely is it that the exposure the 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.

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, 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 a value for chi-square is established, look up its corresponding p-value (or probability value) in a table of chi-squares.

In interpreting p-values, set in advance a cutoff point beyond which chance is a factor to be considered. A common cutoff point is .05. When a p-value is below the predetermined cutoff point, the finding is considered "statistically significant," and the null hypothesis can be rejected in favor of the **alternative hypothesis. T**hat is, it can be concluded that the exposure is associated with disease. The smaller the p-value, the stronger the evidence that the finding is statistically significant.

Step 8: Refine Hypotheses and Carry Out Additional Studies

Additional epidemiological studies
When analytic epidemiological studies do not confirm the hypotheses, reconsider the 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.

Even when an analytic study identifies an association between an exposure and a disease, refining the hypotheses may be necessary. Sometimes more specific exposure histories or a more specific control group may need to be obtained. When an outbreak occurs, whether it is routine or unusual, consider what questions remain unanswered about the disease and what kind of study might be used in the particular setting to answer some of these questions. The circumstances may allow more learning 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. Environmental studies often help explain why an outbreak occurred and may be very important in some settings

Step 9: Implementing Control and Prevention Measures

Even though implementing control and prevention measures is listed as Step 9, in a real investigation one should do this as soon as possible. Control measures, which can be implemented early if the source of an outbreak is known, 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, 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, use the method of "cohorting" by putting infected people together in a separate area to prevent exposure to others. Another example would be to 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, direct control measures at reducing susceptibility. Two such examples are immunization against rubella and malaria chemoprophylaxis (prevention by taking antimalarial medications) for travelers.

Step 10: Communicate Findings

The final task in an investigation is to communicate the 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.

The oral briefing should be attended by the local health authorities and people responsible for implementing control and prevention measures. This presentation is an opportunity to describe what was done, what was found, and what should be done about it. Present the findings in a scientifically objective fashion, and be able to defend the conclusions and recommendations.

In addition, a written report that follows the usual scientific format of introduction, background, methods, results, discussion, and recommendations must be provided. 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.

Written: 07 Jan 2010