Module Objectives

By the end of this module, participants will be able to evaluate the epidemiologic approaches used in an effective foodborne outbreak response.

- Recognize the various methods used to gather information during an investigation.
- Describe the measures of association calculated during an epidemiologic investigation.
- Explain the concept of statistical significance as it relates to measures of association.

Performance Objectives

By the end of this module, participants will be able to evaluate the epidemiologic approaches used in an effective foodborne outbreak response.
Enabling Learning Objectives
By the end of this module, the instructor shall accomplish the following learning objectives in support of the performance objective:

- Recognize the various methods used to gather information during an investigation.
- Describe the measures of association calculated during an epidemiologic investigation.
- Explain the concept of statistical significance as it relates to measures of association.

Activities of the Epidemiologic Investigator

The Epidemiologic Investigator

- Conducts Surveillance Activities
- Identifies and Interviews Cases
- Generates Hypothesis and Case Definitions
- Identifies Comparison Groups (Controls)
- Conducts Analytic Studies

Epidemiologic investigators collect, analyze, and interpret information that may be associated with a foodborne disease outbreak. Information collected may come from routine surveillance activities as described in Module 3, identifying and interviewing cases, finding additional cases and establishing appropriate comparison groups necessary to conduct analytic studies. The epidemiologic investigator analyzes information gathered throughout the investigation process and develops hypotheses and case definitions to share with the outbreak investigation team. You will recall that surveillance activities and hypothesis generation was covered in the initial outbreak activities discussed in Module 3. This Module will focus how the epidemiologic investigator collects and arranges data for further analysis, up to and including evaluating whether the data are statistically significant.
Cross-Disciplinary Activities

The epidemiologic investigator will work with the environmental investigator to focus in on possible foods and their sources to be investigated during the environmental assessment. They can also be vital partners with product tracing activities, which can link cases to food sources. The epidemiologic investigator works with the laboratory investigator to identify appropriate clinical samples to be collected and analyzed in support of the investigation.

Initial Data Collection Activities

The epidemiologic investigator acts on information. The investigator receives information constantly and filters this information to determine if an illness may be foodborne in origin, leading to an outbreak investigation. As the investigator sorts through surveillance information from clinicians, laboratories and other partners, they first must understand if further investigation is warranted.

Initial information may come in the form of a laboratory-confirmed case of disease that generally is associated with a pathogen that may cause foodborne illness. This initial information from the laboratory report provides important information regarding the illness – specifically the name of the ill person and the pathogen causing disease. At this point, the case is contacted and the appropriate pathogen-specific investigation form is employed to gain further information. Most, if not all states will utilize a standardized form with specific questions that take into consideration the characteristics of the pathogen causing illness. It is important at the earliest stages of the investigation to gather information regarding illness onset, symptoms, common exposures and if others are known to be ill with similar signs and symptoms of illness.

Information received directly from a clinician without laboratory confirmation may be more challenging that a laboratory-confirmed case but may be no less important. The clinician’s workup of the case may lead to actionable information if there appears to be multiple people ill from the same event or if they notice an increase in number of ill presenting with similar illness.
In these cases, the clinician provides a name, symptoms and some other notable event that causes the report. The perceptive clinician with a strong relationship with their public health partners may be more likely to provide reports of suspected foodborne illness than clinicians without a relationship. Without the information regarding a pathogen, the epidemiologic investigator will approach the investigation differently, similar to a complaint investigation. The investigator will contact the ill person and develop information using a general food history form of sufficient number of days to cover the incubation period generally associated with the reported symptoms. The number of days of food consumption data collection will range between two and three days for what appears to be a toxin-mediated or norovirus illness up to seven days for diarrheal illness with other complications such as bloody stool and fever. In certain cases, the use of a hypothesis-generating questionnaire may be useful especially if indicated by exposure information and symptoms that may be linked to other illness under investigation. In all cases reported in this fashion, the investigator must ask the ill person to provide a sample for laboratory analysis.

**Case Definitions**

<table>
<thead>
<tr>
<th>Case Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform criteria employed to define a case</td>
</tr>
<tr>
<td>Based on the best available information at the time</td>
</tr>
<tr>
<td>Subject to change</td>
</tr>
<tr>
<td>Will include:</td>
</tr>
<tr>
<td>Person</td>
</tr>
<tr>
<td>Place</td>
</tr>
<tr>
<td>Time</td>
</tr>
<tr>
<td>Clinical Criteria</td>
</tr>
</tbody>
</table>

When searching for additional cases it is important to understand what constitutes a case of illness that may be associated with other cases in an outbreak. The epidemiologic investigator establishes a case definition to classify ill individuals as cases by person, place and time as well as clinical or laboratory information.

A “case definition” is a set of uniform criteria used in making decisions whether a person with an illness should be classified as a case associated with a foodborne illness outbreak investigation.

A case definition is based on the best information available for that time during the response. A case definition should never be considered static and can change based upon new information.
gained throughout the investigation. Criteria may be completely abandoned and reestablished or the case definition may become more granular as more information becomes available.

The following slide is used as an exercise to establish case definitions at the start of an investigation. Participants are provided information on the slide and, working in their table groups, should develop a case definition. The slide will be advanced to reveal a case definition constructed with the information.

### Building a Case Definition

Person: Attended a bar-b-que on July 4th  
Place: Great Bear Wilderness Club  
Time: Illness onset July 6th or later  
Clinical Criteria: Anyone with diarrheal illness (3 or more loose stools in the past 24 hours)

A case associated with this outbreak is a person attending a Bar-B-Que on July 4th at the Great Bear Wilderness Club with diarrheal illness with onset July 6th or later.

Information Available:

- Person: attended a Bar-B-Que on July 4th
- Place: is the Great Bear Wilderness Club
- Time: is Onset of Illness July 6th or later
- Clinical Criteria: Anyone with diarrheal illness (3 or more loose stools in the past 24 hours)

**Case Definition:**

_____________________________________________________________________________________
_____________________________________________________________________________________
_____________________________________________________________________________________
_____________________________________________________________________________________
_____________________________________________________________________________________
The case definition provides the opportunity to search for additional cases using a variety of methods. Having additional cases in the investigation will result in a more powerful analytic investigation as the investigation matures. Not only will the case definition assist with finding cases, it will allow the investigator to find individuals that are not cases to match when conducting analytic studies to determine association. This concept of matching cases with non-cases will be explored later in this Module.

To find additional cases, the investigation team may use a variety of methods. The following are a few common examples where additional cases may be obtained:

- **Guest Lists.** A guest list, if available, may provide a thorough list of attendees to an event. Be sure when contacting attendees on a list to determine whether others attended with them as children are often not listed.

- **Credit Card Receipts.** If the investigation team is focusing on a specific time period of illness at a retail food establishment, credit and debit card receipts may provide a sufficient number of additional people to contact to see if they meet the case definition or if they may be a well person that can be used in the comparison group for further study. As with the guest list, ask the person being interviewed if others attended in their party.

- **Reservation lists or other apps that are used to make reservations.**

- **Coordination between response team member may provide additional cases.** Searching complaint data bases or notifying laboratory partners may yield additional cases. Providing the case definition to healthcare providers with some additional qualifying information may produce additional cases.

- **Media Releases.** Oftentimes used as a last resort, the investigation team may use various forms of media channels to find cases and inform the public. This should be used with caution as it may wrongly implicate an establishment or food. The
investigation team must weigh the likelihood of people contracting a severe illness associated with the outbreak against the use of this type of notifications.

**Hypothesis Generation**

A suggested explanation of what is making people sick  
Ongoing process subject to change  
Testing hypothesis may require a multidisciplinary approach  
- Food and environmental surface laboratory analysis  
- Product tracing activities by environmental health  
- Additional epidemiologic investigation activities  
- Hypothesis-generating questionnaires

To start to refine the outbreak investigation, it is important to generate a hypothesis. A hypothesis is a suggested explanation of what may be making people sick. As with cases definitions, hypothesis generation is an ongoing process of suggesting a source of illness, testing the suggestion and refining the hypothesis based upon the testing, and testing the new hypothesis.

Hypothesis generation can be started as soon as information is collected that is suggestive of a source of illness. Hypotheses can be tested by various methods by response team partners. If a suggested source is related to a food, it may be possible to test the food or environmental surfaces by laboratory analysis. It may be possible to conduct product tracing to identify if a food is or was available as stated by the case during interview. Testing a hypothesis may require additional epidemiologic investigation methods past the initial information gathering steps. If the first step of the investigation does not reveal a suspected source, the epidemiologic investigator may conduct a hypothesis-generating interview. These interviews attempt to identify where and what people had to eat days or weeks before they became ill. If there is an indication of certain food consumption or other activity associated with illness, this questioning may focus on certain foods or events. If the pathogen is known or if the symptoms provide guidance of the potential pathogen, investigators use the incubation period of the pathogen to identify the most important time period to focus questioning of the case’s activities.
Hypothesis-Generating Questionnaires

- Consumption of foods from a standard list
- Meals consumed during the time of likely exposure
- Food shopping habits
- Travel during the time of likely exposure
- Consumption of food away from the home
  - Restaurant
  - Events

Another method used by the epidemiologic investigator is the hypothesis-generating questionnaire (CDC 2018). Questionnaires are standardized and include questions about:

- Consumption of foods from a standard list
- Meals consumed during the time of likely exposure (date of illness onset subtracting the shortest incubation period from the range provided)
- Food shopping habits
- Travel during the time of likely exposure
- Consumption of food away from the home (restaurant, events)

Based upon the responses to this questionnaire, the epidemiologic investigator can create a list of foods consumed by cases to focus other response activities, compare consumption of foods noted with other demographic information, and establish a hypothesis regarding the possible source of the outbreak.
The challenges of hypothesis generation are consistently observed throughout the interviewing process. Recall will always be a problem as time passes. Investigators must use all the tools at their disposal to assist the case with recall. The use of a calendar, credit card statements, and reviewing itineraries may prompt the case-patient to remember what they were doing on a specific day and that, in turn, may assist with food consumption recall. Ingredient-level contamination may pose a significant challenge as some ingredients will go unnoticed. This challenge may be overcome by the judicious work of the environmental investigator to obtain menus and recipes and conduct product tracing. Foods commonly consumed can pose a challenge to the investigation as they may have many sources and be difficult to trace and make it difficult to find comparative groups for analytic studies. Some foods may not be disclosed due to their sensitive nature. An example of this may be the consumption of raw milk. In some states, it is illegal to sell raw milk but when a young child becomes ill from consuming raw milk it may be difficult for a parent to disclose due to the possible ramifications of their action. Hypothesis generation and testing, especially when it entails interviewing cases multiple times can create interview fatigue. Make every effort to minimize this process and use other tools at your disposal to prevent this fatigue as it may result in the loss of the case to the investigation.
The FoodNet population survey, also known as the Atlas of Exposures is published by the CDC and is a compilation of data of FoodNet. Population-based surveys were conducted where participants answered questions on various exposures associated with diarrheal illness, episodes of diarrhea or vomiting in the past month, and basic demographics. FoodNet uses the data to estimate the prevalence and severity of diarrheal illness in the community, describe common symptoms associated with diarrhea, and determine the proportion of persons with diarrhea who seek medical care. They assess exposures that might be risk factors for foodborne illness, such as consumption of risky foods or recent travel outside the United States. As an example, the likelihood of becoming ill from the consumption of raw oysters is greater than consuming cooked oysters and travel to countries that do not have comparable sanitation standards to the U. S. is a risk factor for contracting foodborne illness. If a cluster of cases has a high rate of exposure to a certain food well above what would be considered a normal rate of consumption within the population, it may be a risk factor for foodborne illness and considered for hypothesis testing.

Another method that may lead to understanding of how likely it would be to get the observed number of cases reporting consumption of food given a known or estimated consumption rate would be to use a cumulative binomial probability model. This model, also referred to as Bernoulli trials, calculates binomial probabilities based upon two parameters. The first parameter is the number of interviews where the case was asked about a given food exposure. This is the binomial nature of the calculation as a case either ate the suspected food or did not. The other parameter is the background rate of exposure which may be a known or be a reasonable guess. A calculation table and explanation of its use can be found on the Oregon Food Safety Center of Excellence web site at: https://www.oregon.gov/oha/ph/DiseasesConditions/CommunicableDisease/Outbreaks/Gastroenteritis/Pages/Outbreak-Investigation-Tools.aspx
The use of cumulative binomial probabilities may be beneficial to test a hypothesis or assess the potential utility of a case-control study to determine association of illness with the consumption of a given food.

**Interpretation of Gathered Information**

<table>
<thead>
<tr>
<th>Interpretation of Gathered Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line Lists</td>
</tr>
<tr>
<td>Epi Curves</td>
</tr>
<tr>
<td>Spot Maps</td>
</tr>
</tbody>
</table>

The epidemiologic investigator will often use a variety of methods to inspect the information gathered prior to employing analytic methods. Generally referred to as observational interpretations the following interpretive methods can tell the epidemiologic investigator much about the outbreak, leading to a quick resolution or a more effective analytic study:

- **Line List.** A line list is a means of collating information for rapid examination. By examination, the investigator can look for associations relating ill persons such as symptoms associated with illness, food(s) consumed and incubation period.
- **Epi Curves.** An epidemic “epi” curve is a graphical method of looking at the temporal aspects of a foodborne illness. An epi curve is a type of graph called a histogram and provides a visual interpretation of numerical data (cases) by indicating the number of data points that lie within a range of values (time). Examples of epi curves will be presented later in the Module.
- **Mapping.** The placing of cases of foodborne illness on a map can lead the epidemiologic investigator to better understand the spatial aspects of disease. This mapping of illness is referred to as a spot map.
The following exercise will demonstrate how the line list can be used to better understand aspects of the foodborne illness. The following line list is a compilation of information provided through initial interviews:

<table>
<thead>
<tr>
<th>#</th>
<th>Sex</th>
<th>Age</th>
<th>Onset</th>
<th>Vomiting</th>
<th>Diarrhea*</th>
<th>Fever</th>
<th>Headache</th>
<th>Salmonella Positive Stool Culture</th>
<th>Salad</th>
<th>Burger</th>
<th>Ice Cream</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>14</td>
<td>6/12</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>5</td>
<td>6/13</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>NA</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>16</td>
<td>6/12</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>NA</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>18</td>
<td>6/10</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>5</td>
<td>M</td>
<td>15</td>
<td>6/14</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>NA</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>17</td>
<td>6/12</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>NA</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>16</td>
<td>6/11</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>NA</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>15</td>
<td>6/13</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>NA</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>17</td>
<td>6/20</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>10</td>
<td>F</td>
<td>14</td>
<td>6/12</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>NA</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>
Working in pairs or in table groups answer the following questions regarding interpreting information:

What is the gender and age breakdown of respondent?

______% Female  ______% Male

Age Range _______ to _______

What percentage of each symptom were provided by respondents?

Vomiting: _____ out of 10 or _____%
Diarrhea: _____ out of 10 or _____%
Fever: ______ out of 10 or ______%

There are data presented that may be relevant to the investigation. With the understanding that you may not have all of the information from this event, name at least two things that you would follow up on?

1. ___________________________________________________________________
2. ___________________________________________________________________
3. ___________________________________________________________________

EPI Curves

Epi Curves

- Visually depicts the time element associated with illness onset
- May provide a pattern of onsets that can lead to better understanding of how the illness was contracted
  - Point Source
  - Continuous Common Source
  - Propagated
The epi curve can tell you a number of things about an outbreak. It can give you a sense of the magnitude of the outbreak and provide an understanding of the progression of disease (cases still increasing vs. cases have peaked). It can be used to identify outliers (cases that stand apart from the overall pattern of disease), thereby providing important clues about the source of the outbreak. The shape of an epidemic curve can also suggest the pattern of spread for an outbreak, especially when linked to the incubation period of the pathogen.

The next three slides will show you three different patterns of onset of disease that are commonly associated with outbreaks and what the epi curves typically look like with each.

**Point Source Outbreak**

- Exposure to same source over brief time
- Cases rise rapidly to a peak and fall off gradually
- Majority of cases within one incubation period

A point source outbreak is an outbreak in which persons are exposed to the same source of the outbreak over a brief period of time, such as through a single meal or an event attended by all cases. In a point source outbreak, the number of cases usually rises rapidly to a peak and falls off gradually. The majority of cases occur within one incubation period of the disease.
Continuous Common Source Outbreak

- Exposure to same source over a prolonged time
- Epidemic curve usually rises gradually
- Curve may flatten

A continuous common source outbreak is an outbreak in which persons are exposed to the same source but exposure is prolonged over a period of days, weeks, or longer (e.g., a shelf-stable food that is available to consumers over weeks and months). In a continuous common source outbreak, the epidemic curve rises gradually and may plateau for a time. Cases do not fall until the source of the outbreak is removed or all susceptible persons in the population have developed the illness.
A propagated outbreak is an outbreak that does not have a common source, but instead spreads from person to person. In a propagated outbreak, the epidemic curve classically consists of a series of progressively taller peaks of cases, each peak being one incubation period apart. The peaks grow in size because more and more cases become a source of infection. An ill food handler would be a point source of infection but may also be an index case for diseases generally transmitted person-to-person.
Spot Maps

Distribution of Cases Reveals Clues Regarding the Source of the Outbreak
- Over broad area: commercial product with wide distribution
- Clustering: locally sold product, point source, or person-to-person spread
- Concentrated areas with outliers: travel to affected area or importation of product

**Bars represent cases**

Identifying the places where a patient may have spent time prior to illness, taking into the general or specific incubation period, can lead to identifying important clues about pathogen exposure. Many exposures occur at home or in the places near residences (restaurants, grocery stores, work, school). Exposures occur during travel outside the area of residence and illness associated with travel is commonly solved with this understanding.

Assessment of “place” may demonstrates the geographic extent of the outbreak and may provide important clues about the source of the outbreak. If you see cases:

- Spread over a broad area, it suggests a commercial product with a wide distribution
- Clustering or confined to a more limited area, it suggests a locally sold product, point source (local restaurant, water supply, social event), or person-to-person spread.
- Concentrated with a few outliers, look at the outliers to see if they traveled to the affected area or possibly imported food products from the infected area.

During outbreak investigation, a spot map is prepared in which we indicate the “location” of each case with symbol showing where a case lives, works, or may have been exposed. The use of a spot map, with location of the patient’s residence and food establishment identified by other cases may prompt cases to remember places that were visited at the time of possible exposure.
Epidemiologic Studies – Understanding Association

Epidemiologic Studies - Understanding Association

- Case Series
- Cohort Study
- Case-Control Study
- Case-Case Comparison

The epidemiologic investigator will use various methods to determine if illness is associated with a possible exposure. These methods range from the intensive inspection of available information only from cases associated with a possible outbreak to detailed statistical analysis comparing cases to a comparison group.

Case Series
In a case series, detailed information on exposures that occurred during the incubation period of the disease (including foods eaten) is collected from cases. Commonalities among cases are then used to suggest possible sources of the outbreak.

Since only cases are examined, there is no way to set up a study design to measure association using a statistical approach. Because no comparison group is available in a case series, investigators must be careful in making inferences of association. But occasionally, what is learned from just cases is so remarkable that inferences about the outbreak can be made and control measures implemented. Use of comparison groups such as the results from the FoodNet Atlas for Exposures may assist with the interpretation of results of a case series investigation. Using our response partners from environmental health and the laboratory can strengthen the findings of the epidemiologic investigation. Examples of this cross-disciplinary approach is matching a pathogen identified by clinical laboratory results with a pathogen in a food or environmental sample.

**Case Series Example**

### Multistate Outbreak of *Salmonella* Enteritidis

- Using shopper card information, it was determined that 7 of 9 cases bought Turkish pine nuts from chain store the week before onset of illness
- Background rate: <1% of all shoppers bought Turkish pine nuts at store in previous six months
- Lab testing identified outbreak strain of *S. Enteritidis* in pine nuts/pesto from store. Strain was indistinguishable to the clinical isolates
- Store and producer voluntarily recalled pine nuts

The following is an example of a case series used to identify the source of an outbreak.

A multistate outbreak of *Salmonella* Enteritidis infections was detected through PulseNet. CDC collaborated with public health and agriculture officials in New York and other states and the U.S. Food and Drug Administration (FDA) to investigate.

Early in the investigation, information was collected from cases on foods eaten prior to becoming ill. A review of shopper card records from 7 of 9 cases identified that ill persons had purchased the same type of Turkish pine nuts from bulk bins at different locations of a chain grocery store before becoming ill. The store reported that less than 1% of all shoppers bought Turkish pine nuts at the chain in the previous 6 months.
Subsequent laboratory testing conducted by public health laboratories in several states identified the outbreak strain of *Salmonella* Enteritidis in 14 samples of Turkish pine nuts or pesto containing Turkish pine nuts.

The chain grocery store recalled approximately 5,000 lbs. of Turkish pine nuts sold in the bulk foods department in New York, Pennsylvania, New Jersey, Virginia, and Maryland. The repacker and producer recalled approximately 3,800 lbs. of pine nuts and four lots of the bulk Turkish pine nuts, totaling more than 21,000 pounds.

Epidemiologic evidence about the source of this outbreak was based on a case series. Shopper card information provided background rates of exposure and laboratory testing on pine nut samples confirmed the source of illness.

**Cohort Study**

One of the more common analytic studies used in localized foodborne outbreaks is the cohort study. As the term implies, the study is based on a well-defined group of people (cohort) and the analysis compares attack rates among people exposed to attack rates among people who were not exposed to a certain food. Examples of well-defined groups in which an outbreak occurred would be guests at a wedding banquet or employees of a hospital. It is assumed that members of the group will include both people exposed to the (unknown) source of the outbreak and people not exposed. The cohort study design can be used when all members of the group are included in the analytic study, or when group members are enrolled randomly (i.e., not based on illness status).

In the analysis of a cohort study, people are categorized as to whether they ate particular foods or not. The epidemiologist then determines the proportion of people becoming ill in each group.

If the attack rate of disease (as defined by your case definition) is statistically significantly higher among people who ate a particular food than people who did not eat the food, the food is...
potentially associated with the outbreak; however, sometimes multiple foods are statistically associated with illness and additional analyses are necessary. In addition, other factors (e.g., biologic plausibility) must be considered before concluding that there is clear association between ill persons and consumption of a specific food.

**Attack Rates**

**Is the Incidence of Disease in a Group of People (Cohort)**

**Food-specific Attack Rates are calculated in the following manner:**

\[
\text{Attack Rate} = \frac{\text{Number of ill that consumed the food item}}{\text{Total number that consumed the food item}} \times 100
\]

\[
? = \frac{10 \text{ ill persons consumed the salad}}{20 \text{ persons consumed the salad}} \times 100
\]

The first step in calculating associations with a cohort study is to determine attack rates. An attack rate is the incidence of disease in a group of people. There is generally an overall attack rate determined for the event. If 75 out of 100 attendees at an event became ill, the overall attack rate for the event would be 75%. Overall attack rates are a good way to open a conversation about the epidemiologic study but they do not lead to determining associations between illness and food that cause foodborne illness. Generating specific attack rates is a process of identifying cohorts; determining the number of ill persons that ate a specific food and dividing that number by the total number of persons eating the food.
Attack rates are generally calculated and expressed in a table. By inspection of the attack rate percentages, it is possible to predict what may have been the source of illness. The attack rate associated with consumption of hamburger is double that of the attack rate of those who did not eat hamburger. The calculation of attack rates is the first step of finding an association with illness. The next step is to calculate relative risk.

Relative Risk

- Measure of association for a cohort study
- Answers: “How much more likely is it for people who ate the food to become ill than people who did not eat the food?”

\[
\text{Relative Risk} = \frac{\text{attack rate among exposed}}{\text{attack rate among unexposed}}
\]
The relative risk is the measure of association for a cohort study. The relative risk is the ratio of the disease attack rate among people who ate a particular food to the attack rate among people who did not eat the food.

A relative risk answers the question “How much more likely is it for people who ate the food to become ill than people not eating the food?”

CDC defines “Relative Risk” as the risk of a health event (disease, injury, risk factor, or death) among one group with the risk among another group. It does so by dividing the risk (incidence proportion, attack rate) in group 1 by the risk (incidence proportion, attack rate) in group 2. The two groups are typically differentiated by such demographic factors as sex (e.g., males versus females) or by exposure to a suspected risk factor (e.g., did or did not eat potato salad). Often, the group of primary interest is labeled the exposed group, and the comparison group is labeled the unexposed group.

Relative Risk is calculated in a two by two Table. The two by two table is set up in a specific way to calculate Relative Risk. The formula follows:

\[
\text{Relative Risk} = \frac{a/(a+b)}{c/(c+d)}
\]

The following tables will provide an example of how to calculate the relative risk associated with eating the chicken salad from the attack rate table.

<table>
<thead>
<tr>
<th>Foods Served</th>
<th>Ate the Food</th>
<th>Did Not Eat the Food</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ill (a)</td>
<td>Not ill (b)</td>
</tr>
<tr>
<td>Chicken Salad</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Attack Rate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40%</td>
<td></td>
</tr>
</tbody>
</table>

Ate the Chicken Salad | Total |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>a</td>
</tr>
<tr>
<td>No</td>
<td>c</td>
</tr>
<tr>
<td></td>
<td>a+c</td>
</tr>
</tbody>
</table>

AR Exposed = a/a+b
AR Unexposed = c/c+d
Relative Risk = a/a+b

6-23
<table>
<thead>
<tr>
<th>Ate the Chicken Salad</th>
<th>Disease</th>
<th>AR Exposed</th>
<th>AR Unexposed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>12</td>
<td>18</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>10</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
<td>28</td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

Relative Risk = \( \frac{12/30}{10/20} \)

Relative Risk = \( \frac{0.4}{0.5} \)

Relative Risk = \( 0.8 \)

### Interpreting Relative Risk

**Interpreting Relative Risk**

- **Close to 1.0**: risk of disease is similar among people eating and not eating the food → food not associated with illness
- **Greater than 1.0**: risk of disease is higher among people eating the food than people not eating the food → food could be risk factor
- **Less than 1.0**: risk of disease is lower among people eating the food than people not eating the food → food could be “inversely associated with”
- **Magnitude** of Relative Risk reflects strength of association between eating food and illness.

Relative risk is the measure of association calculated for a cohort study. It answers the question “How much more likely is it for people who ate the food to become ill than people not eating the food?” A relative risk of:

- Close to 1.0 means the risk of disease is similar among people eating and not eating the food. Eating the food is not associated with the outbreak.
- Greater than 1.0 means that the risk of disease is higher among people eating the food than people not eating the food. Eating the food could be a risk factor for the illness.
- Less than 1.0 means that the risk of disease is lower among people eating the food than people not eating the food. Eating the food could be inversely associated with the illness.

“Inversely associated with” can in some instances mean that the factor actually prevents the person from developing the illness. Such might be the case for use of a particular antibiotic or vaccination against a particular infectious disease. In foodborne outbreaks, inversely
associated tends to reflect that persons who ate a particular food were less likely to eat the food that was the cause of the outbreak. For example, if an outbreak of salmonellosis occurs at a Thanksgiving dinner due to contaminated turkey, ham served at the same meal might be inversely associated because people who ate the ham were less likely to also eat the turkey and become ill. It doesn’t protect the person from developing the illness but prevents their exposure to the food that did.

The magnitude of the relative risk is called the “strength of association.” The further away a relative risk is from 1.0, the more likely that the relationship between eating the food and illness is causal. For example, a relative risk of 1.5 is above 1.0, but is not as strong of an association than a relative risk of 10.

### Attack Rate Table with Relative Risk

<table>
<thead>
<tr>
<th>Foods Served</th>
<th>Ate the Food</th>
<th>Did Not Eat the Food</th>
<th>Relative Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ill (a)</td>
<td>Not ill (b)</td>
<td>ill (c)</td>
</tr>
<tr>
<td>Chicken Salad</td>
<td>12</td>
<td>18</td>
<td>40%</td>
</tr>
<tr>
<td>Hamburger</td>
<td>20</td>
<td>7</td>
<td>74%</td>
</tr>
<tr>
<td>Baked Beans</td>
<td>12</td>
<td>10</td>
<td>55%</td>
</tr>
<tr>
<td>Veggie Burger</td>
<td>2</td>
<td>8</td>
<td>20%</td>
</tr>
<tr>
<td>Iced Tea</td>
<td>31</td>
<td>20</td>
<td>61%</td>
</tr>
</tbody>
</table>

By looking at the attack rates, it appears that there may be an association between eating hamburger and becoming ill as the Relative Risk is 3.21 – meaning that if you ate a hamburger at the event, you were 3.2 times more likely to become ill compared to those who did not eat a hamburger. There are slightly elevated relative risks with consuming baked beans and iced tea but without further analysis there is no way to understand if the risks are statistically associated with consumption.
A case-control study is another analytic study design used to determine association. There are two instances where a case-control study may be utilized. There are times that unlike the cohort study, a well-defined group associated with an event is not known. Therefore, a comparison group must be identified to conduct the study. In this use of the case-control study, subjects are enrolled based on whether or not they have (or had) the illness associated with the outbreak. In these studies, persons with the disease of interest are called “cases” or “case-patients” and persons without the disease are called “controls”. Prior exposures, such as eating a particular food item, are compared between cases and controls to see if there is a statistically significant relationship between the disease and the exposure.

Another use of the case-control is when a well-defined group is known, but it is more expedient to conduct a case-control study because you don’t have to interview the entire cohort. The sample interview to determine cases and controls is non-random sample.

In a cohort study we look at people who ate a food or did not eat a food and determine if they became ill or not. With a case-control study, we look at people who were ill or not and look back to see if they ate the food or not.
Odds Ratio

- Measure of association for a case-control study
- Answers: “How much higher is the odds of eating the food among cases than controls?”

\[
\text{odds ratio} = \frac{\text{odds of eating food among cases}}{\text{odds of eating food among controls}}
\]

Generally, a case-control study is used to determine association when the exposed group is large and the extent of exposure is unknown. The study design yields the odds of an exposure to a given food resulting in illness. It answers the question “How much higher is the odds of eating the food among cases as compared to controls?”. The measure of association for a case-control study is called an odds ratio. The odds ratio compares the odds of eating a food among cases to the odds of eating the food among controls.
Two sisters and their mother from Vancouver, British Columbia, developed signs and symptoms suggestive of botulism. After these cases were publicized, 34 additional cases of botulism were identified in the area. All case-patients had eaten at a single, family-style restaurant.

A case-control study was undertaken to determine the source of the outbreak at the restaurant. Cases were persons who had eaten at the Vancouver restaurant who had neurologic signs and symptoms suggestive of botulism. Controls were persons who ate at the restaurant with case-patients but developed no gastrointestinal or neurologic symptoms in the following 2 weeks. Twenty-two case-patients and 22 controls were interviewed. It was determined that 20 (91%) of 22 case-patients but only 3 (14%) of 22 controls ate a beef dip sandwich at the restaurant.

By inspection, it seems that eating beef dip sandwiches was related to becoming ill. However, epidemiologists typically don’t just “eyeball” the numbers. To the untrained epidemiologist, one might assume that chances of cases ingesting beef dip sandwiches was 91%, but without information about illness in the control group you do not see the entire picture.
Determining the Odds Ratio

- 20 of 22 cases ate beef dip sandwich (2 didn’t)
- 3 of 22 controls ate beef dip sandwich (19 didn’t)

\[
\text{odds ratio} = \frac{\text{odds of eating food (cases)}}{\text{odds of eating food (controls)}} = \frac{20/2}{3/19} = 63
\]

An odds ratio of 63 means the odds that cases ate the beef dip sandwich was 63 times higher than the odds among controls. Eating the beef dip sandwich might be a risk factor for botulism in this outbreak.

The odds of eating the beef dip sandwich among cases is \( \frac{20}{2} \). The odds of eating beef dip sandwiches among control was \( \frac{3}{19} \). The odds ratio is \( \frac{20}{2} \) divided by \( \frac{3}{19} \) or 63.

Interpreting Odds Ratios

- **Close to 1.0:** odds of eating food is similar among cases and controls → no association between food and illness
- **Greater than 1.0:** odds of eating food among cases is higher than among controls → food could be risk factor
- **Less than 1.0:** odds of eating food among cases is lower than among controls → food could be “inversely associated ”

**Magnitude** reflects strength of association between illness and eating the food.

The odds of eating beef dip sandwiches was 63 times higher among cases than controls. Because the odds of eating the food is higher among cases than controls (greater than 1.0),
eating the food could be a risk factor for the outbreak. The magnitude of the odds ratio suggests a strong association.

**Case-Case Studies**

---

**Case-Case Study Design**

- Yields odds ratio similar to case-control studies depending on risk factors under investigation
- May eliminate the need to match controls
  - Reduces selection and recall bias
  - Saves valuable resources as case-patients are interviewed as part of an effective surveillance program

---

As an alternative to the case-control studies to determine association with odds ratios, an effective alternative may be the use of the case-case study. Identifying control groups for case-control studies has limitations. Selection bias, recall bias and intensive resources are necessary to recruit and interview non-ill controls. In the case-case study design, interviews are conducted on case-patients and based on disease and risk factors, so they may act as a comparison cases (controls) for the disease under investigation.

Depending upon the risk factor under study, if the risk factors are similar but not necessarily tied to the outbreak under investigation, the odds ratios may need thoughtful interpretation as case patients with similar risk factors may yield an odds ratio close to 1.0. When the case-patient population used as a comparison group does not report the specific risk factor under investigation, the case-case study is strikingly similar to the case-control study as a means to measure association.

Since routine surveillance systems collect data on cases associated or not with an outbreak under investigation, a case-case study design may eliminate the need to match a control group to cases, reducing selection and recall bias, and saving valuable resources.

Note: Presenting the case-case study is intended to demonstrate an alternative study design that yields association by odds ratios. It will be up to the epidemiologic investigator to present the most appropriate study design for the information available. From a cross-disciplinary perspective, the outbreak response team members must understand this is a variation in study design that yields plausible associations for action.
Choosing a Study Design

The three common types of epidemiologic studies used in foodborne disease outbreaks are case series, case-control studies, and cohort studies. The study type depends largely on the circumstances of the particular outbreak and the resources available. The following guidance will be helpful:

Case Series – The number of cases is small (less than five) and no controls are available. The epidemiologic investigator attempts to demonstrate an association utilizing descriptive statistics.

Cohort Study – Investigators can easily identify the population at risk (i.e., outbreak has occurred in a well-defined group) and the population at risk can be enumerated. Examples of when you might use a cohort study is when an outbreak occurs among students at a school, residents of a particular nursing home, or the people attending a certain event.

Case-control Study – The population at risk (people potentially exposed to source of outbreak) is unknown or cannot be enumerated or the illness is rare. An example of when you might use a case-control study is a multijurisdictional outbreak of a particular disease in which cases are distributed across many different states with no obvious association. Another example when a population is known but it is not necessary to interview the entire cohort. This may occur with illness outbreaks in restaurants. The table below summarizes when each of the studies may be used:
<table>
<thead>
<tr>
<th>Case Series</th>
<th>Cohort Study</th>
<th>Case-Control Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>People with Disease</td>
<td>People in a well-defined group who ate or did not eat certain foods</td>
</tr>
<tr>
<td>Purpose</td>
<td>Collect information or those ill</td>
<td>Developed illness or not</td>
</tr>
<tr>
<td>Measure of Association</td>
<td>Descriptive Statistics</td>
<td>Relative Risk</td>
</tr>
<tr>
<td>When to Use</td>
<td>Small Groups with no comparisons</td>
<td>Easily Identified, Discrete population</td>
</tr>
<tr>
<td>Example</td>
<td>3 people ill at a common meal</td>
<td>20 people attending a picnic that became ill and the entire number of attendees is known.</td>
</tr>
</tbody>
</table>

**Statistical Significance**

- Odds ratios and relative risks are estimates
- Observed results could be due to chance alone
- Role of chance explored through
  - Confidence interval (CI)
  - p-value

The odds ratio or relative risk are estimates of association for a food item or disease occurrence. Tests of statistical significance must be used to determine the likelihood (probability) of finding an odds ratio or relative risk as high as the one observed due to chance alone. Chance alone is the concept that there was no true association between the food item and disease.
Two commonly used tests to explore the role of chance are the confidence interval and p-value.

### Confidence Intervals

<table>
<thead>
<tr>
<th>Confidence Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range of values for the measure of association that are consistent with study findings</td>
</tr>
<tr>
<td>Has specified probability of including “true value” for the measure of association. Generally a probability of 95% is used.</td>
</tr>
<tr>
<td>Positive association when range of values are greater than 1.0</td>
</tr>
<tr>
<td>Inverse association when range of values between 0 and 1.0</td>
</tr>
<tr>
<td>Ranges that include 1.0 are not significant!</td>
</tr>
</tbody>
</table>

The confidence interval is the range of values for a particular measure of association consistent with the study findings. It is the range of plausible values and is described using lower and upper ranges of values. The measure of association should lie somewhere within the lower and upper bounds.

The confidence interval is calculated so that the range has a specified probability of including the “true value” of the measure of association, which, of course, is only an estimation. The 95% confidence interval is the most commonly used confidence interval, hence it has a 95% probability of including the true value of the measure of association.

A confidence interval that includes 1.0 is not statistically significant. Recall that a Relative Risk of 1.0 means there is no association between the consuming the food and contracting illness. Hence, if the range of values includes 1.0, the confidence interval is not significant and the food is not statistically associated with illness (please note, though, that lack of statistical significance, even when a food is the vehicle, could be due to any number of factors such as small numbers of cases and/or controls in the study).

For example, if the odds ratio for a case-control study is 5.2, the best single estimate for quantifying the relationship between illness and eating the food is 5.2. A confidence interval of 4.0-6.1 suggests that study results are consistent with an odds ratio anywhere between 4.0 and 6.1 and that there is a 95% probability that the true value for the odds ratio lies in that range. Another way to look at these results is that there is only a 5% chance that the odds ratio lies outside of the calculated confidence interval.
Confidence intervals of the Cohort Study

Referring back to the Cohort Study, the upper and lower limits of the 95% confidence interval have been calculated. Based upon this information, what can be inferred about each of the foods listed from the attack rate table:

<table>
<thead>
<tr>
<th>Foods Served</th>
<th>Ate the Food</th>
<th>Did Not Eat the Food</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ill (a)</td>
<td>Not ill (b)</td>
<td>Attack Rate</td>
</tr>
<tr>
<td>Chicken Salad</td>
<td>12</td>
<td>18</td>
<td>40%</td>
</tr>
<tr>
<td>Hamburger</td>
<td>20</td>
<td>7</td>
<td>74%</td>
</tr>
<tr>
<td>Baked Beans</td>
<td>12</td>
<td>10</td>
<td>55%</td>
</tr>
<tr>
<td>Veggie Burger</td>
<td>2</td>
<td>8</td>
<td>20%</td>
</tr>
<tr>
<td>Iced Tea</td>
<td>31</td>
<td>20</td>
<td>61%</td>
</tr>
</tbody>
</table>

Chicken Salad:

Hamburger:

Baked Beans:

Veggie Burger:

Iced Tea:
P-Value

The p-value is the probability of finding an association between an exposure (eating a certain food) and a disease as strong as the one observed, if the exposure is not actually related to disease. Stated another way, the p-value is the likelihood that the finding occurred simply through chance.

Probabilities range from 0.0 (0%) to 1.0 (100%). A p-value close to 1.0 (100%) indicates a high probability that the finding is due to chance alone. A p-value close to 0.0 (0%) indicates a low probability that the finding is due to chance.

As an example, if we find that the measure of association between a particular food and becoming ill has a p-value of 0.02, that means there is a 2 in 100 chance (probability) that the observed measure of association is due to chance alone. A 2 in 100 chance of it occurring by chance alone is fairly low.

In public health investigations, p-Value targets (cutoffs) are lower than 0.05. Measure of association is statistically significant when the p-value is lower than 0.05. Statistical software packages are relied upon to calculate p-value.
### P-Values Calculated for Attack Rate Table

<table>
<thead>
<tr>
<th>Foods Served</th>
<th>Ate the Food</th>
<th>Did Not Eat the Food</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ill (a)</td>
<td>Not ill (b)</td>
<td>Attack Rate</td>
<td>ill (c)</td>
<td>Not ill (d)</td>
<td>Attack Rate</td>
<td>Relative Risk</td>
</tr>
<tr>
<td>Chicken Salad</td>
<td>12</td>
<td>18</td>
<td>40%</td>
<td>10</td>
<td>10</td>
<td>50%</td>
<td>0.80</td>
</tr>
<tr>
<td>Hamburger</td>
<td>20</td>
<td>7</td>
<td>74%</td>
<td>3</td>
<td>10</td>
<td>23%</td>
<td>3.21</td>
</tr>
<tr>
<td>Baked Beans</td>
<td>12</td>
<td>10</td>
<td>55%</td>
<td>8</td>
<td>10</td>
<td>44%</td>
<td>1.23</td>
</tr>
<tr>
<td>Veggie Burger</td>
<td>2</td>
<td>8</td>
<td>20%</td>
<td>22</td>
<td>5</td>
<td>81%</td>
<td>0.25</td>
</tr>
<tr>
<td>Iced Tea</td>
<td>31</td>
<td>20</td>
<td>61%</td>
<td>7</td>
<td>6</td>
<td>54%</td>
<td>1.13</td>
</tr>
</tbody>
</table>

The slide above uses the attack rate table and generates the associated p-value to determine significance. How would you interpret the table above?

**Responses**

- 
- 
- 
- 
- 

**Note to Participants**

There are on-line resources that will assist with the calculation of confidence intervals and p-values. The CDC has a tool at: [https://www.cdc.gov/epiinfo/user-guide/statcalc/statcalcintro.html](https://www.cdc.gov/epiinfo/user-guide/statcalc/statcalcintro.html).
Summary

- Recognize the various methods used to gather information during an investigation
- Describe the measures of association calculated during an investigation
- Explain the concept of statistical significance as it relates to measures of association

Coming Up Next

Environmental Investigation