VOLUME 3, ISSUE 2

FOCUS on Field Epidemiology

Case-Control Studies for Outbreak Investigations

CONTRIBUTORS

Authors:

Amy Nelson, PhD, MPH

Kim Brunette, MPH

FOCUS Workgroup*

Reviewers:

FOCUS Workgroup*

Production Editors:

Tara P. Rybka, MPH

Lorraine Alexander, DrPH

Rachel A. Wilfert, MD, MPH

Editor in chief:

Pia D.M. MacDonald, PhD, MPH

*** All members of the FOCUS Workgroup are named on the last page of this issue.**

UNC SCHOOL OF PUBLIC HEALTH

NORTH CAROLINA CENTER FOR PUBLIC HEALTH PREPAREDNESS

The North Carolina Center for Public Health Preparedness is funded by Grant/Cooperative Agreement Number U90/CCU424255 from the Centers for Disease Control and Prevention. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the views of the CDC.

"If you have never been to Western Michigan in the winter, the odds are you have never seen how lake-effect snow can keep falling and falling…." "And if you have never been rollerblading, the odds are you do not know the joy of rolling freely down a path or the terror of not being able to stop at an intersection."

People make statements like these all the time. They use odds as a way of describing the likelihood of a particular event in relation to some more general experience. Maybe you do rollerblade; then you probably have had the experience of yelling "How do I STOP?!" But if you have never been rollerblading, you probably have not experienced the terror of having wheels on your feet and not being able to stop.

Let's think about this example in terms of exposure and outcome. The exposure is equivalent to the experience of rollerblading, while the outcome is the terror of not being able to stop. And the odds reflect the likelihood of the exposure-outcome relationship. In this issue of FOCUS, we discuss the case-control study design, which uses the idea of odds to describe exposure-disease relationships.

As noted in earlier FOCUS issues on study design, analytic studies are conducted to answer the question, "What is the relationship between exposure and disease?" The case-control design is an efficient method of exploring this relationship. It is often conducted with relatively few diseased individuals (so it is efficient), and it is particularly useful for studying a rare disease or investigating an outbreak.

Case Selection

Determining who will be a case in the study depends on how a case is defined. A case definition is "a set of standard criteria for deciding whether an individual should be classified as having the health condition of interest." (1) The definition usually includes clinical criteria and is restricted to a certain time and place, and certain personal characteristics. The case definition must be clear, objective, and consistently applied.

After the case definition is decided on, case-patients need to be recruited into the study. There are many ways to identify case-patients, including medical records, laboratory results, surveillance systems, registries, and mass screening programs. Case-patients can also be asked to identify other persons they know who have a similar illness.

- For example, in August 2001, the Illinois Department of Health was notified of a cluster of cases "of diarrheal illness associated with exposure to a recreational waterpark in central Illinois." To identify additional case-patients, local media and community networks were asked to encourage ill persons to contact the local health department. In addition, case-patients were asked "if there were any other ill persons in their household or if anyone attending the waterpark with them was ill." (2)

Control Selection

The most difficult part of a case-control study is choosing the control group. It has even been said that, "This is one of the most difficult problems in epidemiology. The challenge is this: If we conduct a case-control study and find more exposure in the cases than in the controls, we would like to be able to conclude that there is an association between the exposure and the disease in question. The way the controls are selected is a major determinant of whether such a conclusion is valid." (3)

Controls are persons who do not have the disease in question. "Controls should be representative of the population from which the cases arose [known as the source population], so that if a control had developed the disease, he or she would have been included as a case in the study." Controls should also provide a good estimate of the level of exposure one would expect in that population. (1)

There are several sources of controls for case-control studies. They can be selected from the same health-care institutions or providers as the cases, the same institution or organization as the cases (e.g., schools, workplaces), from relatives, friends, or neighbors of the cases, or randomly from the source population. (1) Investigators may even use multiple methods of control selection. Additionally, investigators may choose to select multiple controls per case to increase the likelihood of identifying significant associations (usually no more than 3 controls per case). Keep in mind, however, that the sources for your controls will depend on the scope of the outbreak.

- *Persons served by the same health-care institution or providers as the cases*

  "In August 2001, a cluster of *Ralstonia pickettii* bacteremia occurred among neonatal intensive care unit (NICU) infants at a California hospital....A case-control study was conducted to identify risk factors for infection....Controls were NICU infants who: (1) had blood cultures taken during either cluster period (July 30-August 3 and August 19-30); (2) had blood cultures that did not yield *R. pickettii*; and (3) had been in the hospital for at least 72 hours." Investigators attempted to recruit 2 controls per case-patient. (4)

- *Members of the same institution or organization*

  In a 2004 outbreak of varicella in a primary school in a suburb of Beijing, China, a case-control study was conducted to identify factors contributing to the high rate of transmission and to "assess the effectiveness of control measures." Controls included randomly selected students in grades K—2 of the primary school with "no history of current or previous varicella." One control was recruited for each case-patient. (5)

- *Relatives, friends, or neighbors*

  In August 2000, an increase was noted in *Salmonella* serotype Thompson isolates from Southern California patients with onset of illness in July. "The preliminary interviews confirmed that many of the patients with *S.* Thompson infection had eaten at a Chain A restaurant in the 5 days before illness onset. Therefore a case-control study was conducted to evaluate specific food and drink exposures at Chain A restaurants....Controls were well friends or family members who shared meals with cases at Chain A during the exposure period." (6)

- *Random sample of the source population*

  "During January - June 2004, an aflatoxicosis outbreak in eastern Kenya resulted in 317 cases and 125 deaths." A case-control study was conducted to "identify risk factors for contamination of implicated maize....[Investigators] randomly selected 2 controls from each case-patient's village....To choose each control, [investigators] spun a bottle in front of the village elder's home and walked to the fifth house in the direction indicated by the bottle (or to the third house in sparsely populated areas)." At the selected household, a random number list was used to select one household member. (7)

- *Multiple methods of control selection*

  In the waterpark outbreak of diarrheal illness in Illinois mentioned above, investigators recruited 1 control per case using 3 methods. First, case-patients were asked to identify a healthy person. Second, investigators used "the local reverse-telephone directory based on residential address of case-patients." Lastly, investigators canvassed local schools and community groups for controls. (2)

Selection Bias

Bias is a distortion of the relationship between exposure and disease. If there are systematic differences in the way you select your controls and the way you select your cases, you could introduce bias. In epidemiology, we refer to bias related to the way cases or controls are chosen as "selection bias."

- For example, suppose that most of your case-patients work on lower floors of an office building and employees on the lower floors are more likely to leave the building to go out for lunch. If your control population is mostly employees from upper floors, the comparison may lead you to conclude that there is a real difference between cases and controls associated with eating at a local deli. But in fact the difference is due to where they worked in the building, which resulted in how often they ate in restaurants.

- Let's say that you are investigating an outbreak at a gym and a majority of the case-patients are females. A majority of the controls you recruit are male. You conclude that there is an association between the illness and participation in an aerobics class. In reality, however, the outbreak was caused by the steam in the sauna in the women's locker room. The appearance of a relationship between the illness and the aerobics class was simply due to the fact that women are more likely to take an aerobics class than men.

Matching

Since the validity of case-control studies is dependent on the similarity of cases and controls in all respects except exposure, investigators frequently "match" cases and controls on characteristics like age and gender. Matching factors should be important in the development of the disease, but not in the exposure under investigation. Since the matching variable will not be associated with either case or control status, it cannot confound, or distort the exposure-disease association.

There are two ways to match cases and controls: individual matching and group matching. In individual matching, or the use of matched pairs, each case is matched with a control that has specific characteristics in common with the case. In group matching, also known as frequency matching and category matching, the proportion of controls with certain characteristics must be identical to the proportion of cases with these characteristics. This method requires that all cases be selected first so that the investigator knows the proportions to which the controls should be matched. For example, if investigators wanted to match on gender and 30% of the cases were male, then investigators would select controls so that 30% of controls would be male.

- To illustrate individual matching, in an outbreak of tularemia in Sweden in 2000, investigators conducted a case-control study to identify risk factors for the illness. Investigators selected 2 controls for each case, matched for age, sex, and place of residence, and identified through the computerized Swedish National Population Register, which stores the name, date of birth, personal identifying number, and address of all citizens and residents. (8)
- To illustrate group matching, in an outbreak of *Escherichia coli* associated with a petting zoo at the 2004 North Carolina State Fair, investigators recruited 3 controls for each case. Controls were group-matched by age groups (1-5 years, 6-17 years, and 18 years and older), meaning that the proportion of controls in each age group was identical to the proportion of cases in each age group. The controls were identified "from a list provided by fair officials of 23,972 persons who purchased tickets to the fair online, at kiosks, or in malls." (9)

Matching can be time efficient and cost effective, and improve statistical power. However, the more variables chosen as matching characteristics, the more difficult it is to find a suitable control to match to the case. It is important to remember that once a variable is used for matching, there can be no discernible relationship between this variable and the disease. So do not match on anything you think might be a risk factor for disease. Remember, when a matched study is carried out, data must be analyzed using methods consistent with matched data.

Conducting the Investigation

The next step is to gather demographic information and exposure histories from cases and controls. (Information on questionnaire development and interviewing techniques is available in past issues of FOCUS.) After you have collected the data you need, you can begin the analysis and calculate measures of association.

Analyzing the Data

In a case-control study, an odds ratio is calculated to measure the association between an exposure and the occurrence of a disease.

Calculating Odds and Odds Ratios

What are odds? How are odds different from probability or

Useful resources for case-control studies:

- Case-Control Studies. ERIC Notebook. September 1999, Issue 5.
- Rothman KJ. Epidemiology: An Introduction. New York, Oxford University Press; 2002.

risk? Let's say there is a bag containing 20 poker chips: 4 red chips and 16 blue chips.

Probability is the number of times something occurs divided by the total number of possible occurrences. The probability of getting a red chip is 4/20 (or 1/5 or 20%). The probability of getting a blue chip is 16/20 (or 4/5 or 80%).

Odds are the number of times something occurs divided by the number of times something does not occur. The odds of getting red are 4/16 (or 1/4) and the odds of picking a blue chip are 16/4 (or 4/1). People may refer to the odds of getting a blue chip as 4 to 1 against getting a red chip.

Odds = probability/(1-probability)

If the probability of picking a red chip is 20%, then the odds are 0.20/(1-0.20) or 1/4.

Probability = odds/(1+odds)

If the odds of picking a red chip are 1/4, then the probability of getting a red chip is 0.25/(1+0.25)=0.20.

Thus, while odds are a measure related to probability, they measure the occurrence of an event as compared to the non-occurrence of the same event. Variables with two levels, called *binary* variables, are used to calculate odds. Binary variables are those with yes/no responses, such as disease/no disease, or exposed/not exposed.

The odds of exposure among cases are calculated by dividing the number of exposed cases by the number of unexposed cases. Similarly, the odds of exposure among controls are calculated by dividing the number of exposed controls by the number of unexposed controls.

$$\text{Odds for cases} = \frac{\text{Exposed cases}}{\text{Unexposed cases}}$$

$$\text{Odds for controls} = \frac{\text{Exposed controls}}{\text{Unexposed controls}}$$

A 2x2 table shows the distribution of cases and controls:

Odds: 2x2 table

	Case	Control
Exposed	a	b
Not exposed	c	d
Odds of exposure	a/c	b/d

An odds ratio (OR) is the odds of exposure among cases (a/c) divided by the odds of exposure among controls (b/d). Exposure among cases is compared to exposure among controls to assess whether and how exposure levels differ between cases and controls. This OR is numerically the same as that obtained by multiplying diagonally across the 2x2 table and dividing the products (ad/bc); hence the name "cross-products ratio."

Odds Ratio = (a/c) ÷ (b/d) = ad/bc

To interpret the odds ratio, we compare the value of the OR to 1:

If the odds ratio = 1; the odds of exposure are *the same* for cases and controls (no association between disease and exposure).

If the odds ratio > 1; the odds of exposure are *greater* among cases than among controls (a positive association between disease and exposure).

If the odds ratio < 1; the odds of exposure are *less* among cases than among controls (a negative, or protective, association between disease and exposure).

Note: Accurate interpretation of the odds ratio requires the use of confidence intervals and other statistical methods that will be discussed in an upcoming issue of FOCUS.

- For example, in an investigation of an outbreak of Hepatitis A among patrons of a Pennsylvania restaurant, a case-control study was conducted to identify food associated with the illness. A total of 240 case-patients and 134 controls were identified. Data obtained from the case-patients and controls found that 218 case-patients and 45 controls had consumed mild salsa, as shown in the table below. (10)

2x2 table showing exposure to mild salsa among case-patients and controls

	CASE	CONTROL
Ate mild salsa (exposed)	218	45
Did not eat mild salsa (not exposed)	22	89

$$OR = \frac{(218/22)}{(45/89)} = \frac{(218x89)}{(45x22)} = 19.6$$

- In this example, the odds ratio of 19.6 means that among patrons who became ill the odds of having eaten the mild salsa were 19.6 times the odds of patrons not who did not get ill. To put it simply, cases were 19.6 times more likely than controls to have eaten mild salsa. (10)

On occasion, you may have an exposure variable that has more than two levels (e.g., age group, race, or serving size). You can calculate an odds ratio for each level relative to a reference group. For example, suppose you have several serving sizes: 0 servings, <1, 1, and 2 or more servings. The reference group is those who had 0 servings. You calculate the odds ratios as follows:

	Case	Control	Odds Ratio
≥2 servings	a_3	b_3	a_3d/b_3c
1 serving	a_2	b_2	a_2d/b_2c
<1 serving	a_1	b_1	a_1d/b_1c
0 servings	c	d	reference

Matched Analysis

If you decide to match controls to cases using individual matching (rather than group matching), the 2x2 table needs to be set up differently: you examine *pairs* in the table, so you have cases along one side and controls along the other, and each cell in the table contains pairs. The generic set-up of the table is:

		Controls		
		Exposed	Not Exposed	Total
Cases	Exposed	e	f	e + f
	Not Exposed	g	h	g + h
	Total	e + g	f + h	

Cell e contains the number of matched case-control pairs in which both the case and the control were exposed. This is a concordant cell because the case and the control have the same exposure status. Cell h is also a concordant cell.

Cell f contains the number of matched case-control pairs in which the cases were exposed but the controls were not exposed. This is a discordant cell (as is cell g) because the case and the control have a different exposure status.

Because you want to contrast the exposure between cases and controls, only the discordant cells (f and g) provide useful data. In individually matched analysis, the matched odds ratio is calculated as cell f divided by cell g.

Matched Odds Ratio = f/g

With group matching, stratified analysis should be used. This type of analysis will be discussed in an upcoming issue of FOCUS.

Odds vs. Risk

Odds differ qualitatively from the risk calculated in a cohort study, and it is important to understand this distinction. Case-control studies select participants based on disease status and then measure exposure among the participants; therefore they can only approximate the risk of disease given exposure. The values needed to calculate risk are not available from a case-control study because the study does not include the entire population at risk. Although you might be able to include all or most of the cases in the study, finding all of those who did not get sick would be difficult or impossible. A case-control study thus uses only a subset of many potential controls, and you calculate the odds ratio as an estimate of the risk.

Examples of Case-Control Studies

E. coli associated with a fast-food restaurant chain

In November 1999, a children's hospital notified the Fresno County (California) Health Department of 5 cases of *E. coli* O157 infections during a 2-week period (6). Initial interviews "revealed that all case-patients had eaten at popular fast-food restaurant chain (chain A) in the 7-day period before the onset of illness." Local health officials and clinicians throughout California were asked to enhance surveillance for *E. coli* O157 infections. Additionally, states bordering California were asked to "review medical histories of persons with recent *E. coli* O157 infections and arrange for PFGE [pulse-field gel electrophoresis] subtyping of recent *E. coli* O157 isolates. To identify risk factors for infection, 2 sequential case-control studies were conducted in early December 1999." (11)

The first case-control study was conducted "to determine the restaurant associated with the outbreak. For this study, a 'case' was defined as a patient with (1) an infection with the PFGE-defined 'outbreak strain' of *E. coli* O157:H7..., a diarrheal illness with ≥ 3 loose stools during a 24-hour period, and/or an HUS [hemolytic uremic syndrome] during the first 2 weeks of November 1999; or (2) an illness clinically compatible with *E. coli* O157:H7 infection, without laboratory confirmation but with epidemiologic connection to the outbreak. A 'control' was defined as a person without a diarrheal illness or HUS during the first 2 weeks of November 1999. Controls were age-matched and systematically identified using computer-assisted telephone interviewing of residents in the same telephone exchange area as case-patients. [Investigators attempted] to obtain 2 controls per case. Case-patients and controls were queried using a standardized questionnaire to determine whether they had eaten at a number of national fast-food restaurant chains in the week before illness onset." Investigators enrolled 10 cases and 19 matched controls. "Of the 9 restaurants, only chain A showed a statistically significant association with illness among cases and controls." (11)

"[Based on these results], a second case-control study involving patrons of chain A restaurants was conducted to determine the specific menu item or ingredient associated with illness. For this study, a case was defined as above but restricted to those who had eaten at chain A, and only those who could be matched with 'meal companion-controls.' Cases and controls were asked about consumption of specific foods and beverages that appeared on the chain A restaurant menu....Eight cases and 16 meal companion-controls were enrolled in this study." By calculating the matched odds ratio, consumption of a beef taco was found to be significantly associated with illness. A traceback investigation implicated an upstream supplier of beef, but a farm investigation was not possible. (11)

Listeriosis associated with deli meat

"[In] July and August 2002, there were 22 cases of listeriosis in Pennsylvania, a nearly 3-fold increase over baseline." PFGE subtyping "identified a cluster of cases caused by a single *Liseteria monocytogenes* strain. The CDC [Centers for Disease Control and Prevention] asked health departments in the northeast United States to conduct active case finding, prompt reporting of listeriosis cases, and retrieval of clinical isolates for rapid PFGE testing...." Investigators conducted a case-control study to identify the source of the outbreak. (12)

"A case-patient was defined as a person with culture-confirmed listeriosis between 1 July and 30 November 2002, whose infection was caused by the outbreak strain. A control-patient was defined as a person with culture-confirmed listeriosis between 1 July and 30 November 2002, whose infection was caused by any other non-outbreak strain of *L. monocytogenes*, and who was from a state with at least 1 case-patient. Case-patients and control-patients were interviewed with a standard questionnaire [including more than 70 specific food items]...to gather medical and food histories during the 4 weeks preceding culture for *L. monocytogenes*." (12)

"[The study obtained data] from 38 case-patients and 53 controls." By calculating the odds ratio, investigators found that "infection with the outbreak strain was strongly associated with consumption of precooked turkey breast products sliced at the deli counter of groceries and restaurants....No other single food item was significantly associated with outbreak strain infection, except for lettuce, which was protective." (12)

Based on the results of a traceback investigation, 4 turkey processing plants were investigated. The outbreak strain of *L. monocytogenes* was found in the environment of plant A and in turkey breast products from plant B. Both plants suspended production and together recalled more than 30 million pounds of products, resulting in one of the largest meat recalls in US history. (12)

Conclusion

In using the case-control study method, considering underlying characteristics of the population that gave rise to the cases will help to select appropriate controls. Improper selection of controls can introduce bias and result in a spurious association between an exposure and illness. If controls are representative of the source population, case-control studies are an efficient way to conduct an analytic study to determine the relationship between exposure and disease.

Glossary

Matching: The process of making cases and controls comparable with respect to extraneous factors.

Odds: The ratio of the probability of occurrence of an event to that of nonoccurrence, or the ratio of the probability that something is so to the probability that it is not so.

Odds Ratio: The ratio of the odds of exposure among cases to the odds of exposure among controls.

From: Last, JM. *A Dictionary of Epidemiology*. 4th ed. New York, NY: Oxford University Press; 2001.

CONTACT US:

The North Carolina Center for Public Health Preparedness

The University of North Carolina at Chapel Hill

Campus Box 8165

Chapel Hill, NC 27599-8165

Phone: 919-843-5561

Fax: 919-843-5563

Email: nccphp@unc.edu

FOCUS Workgroup:

- Lorraine Alexander, DrPH
- Kim Brunette, MPH
- Anjum Hajat, MPH
- Pia D.M. MacDonald, PhD, MPH
- Gloria C. Mejia, DDS, MPH
- Sandi McCoy, MPH
- Amy Nelson, PhD, MPH
- Tara P. Rybka, MPH
- Michelle Torok, MPH
- Rachel A. Wilfert, MD, MPH

REFERENCES:

1. Gregg MB. *Field Epidemiology*. 2nd ed. New York, NY: Oxford University Press; 2002.
2. Causer LM, Handzel T, Welch P, et al. An outbreak of *Cryptosporidium hominis* infection at an Illinois recreational waterpark. *Epidemiol Infect.* 2006;134(1):147-156.
3. Gordis L. *Epidemiology*. 2nd ed. Philadelphia, PA: WB Saunders Company; 2000.
4. Kimura AC, Calvet H, Higa JI, et al. Outbreak of *Ralstonia pickettii* bacteremia in a neonatal intensive care unit. *Pediatr Infect Dis J*. 2005;24:1099-1103.
5. Ma H, Fontaine R. Varicella outbreak among primary school students--Beijing, China, 2004. *MMWR Morb Mortal Wkly Rep*. 2006;55(suppl):39-43.
6. Kimura AC, Palumbo MS, Meyers H, Abbott S, Rodriguez R, Werner SB. A multi-state outbreak of *Salmonella* serotype Thompson infection from commercially distributed bread contaminated by an ill food handler. *Epidemiol Infect*. 2005;133(5):823-828.
7. Azziz-Baumgartner E, Lindblade K, Gieseker K, et al and the Aflatoxin Investigative Group. Case-control study of an acute aflatoxicosis outbreak, Kenya, 2004. *Environ Health Perspect*. 2005;113:1779-1783.
8. Eliasson H, Lindbäck J, Nuorti JP, et al. The 2000 tularemia outbreak: a case-control study of risk factors in disease-endemic and emergent areas, Sweden. *Emerg Infect Dis.* 2002;8:956-960.
9. Goode B, O'Reilly C. Outbreak of Shiga toxin producing *E. coli* (STEC) infections associated with a petting zoo at the North Carolina State Fair – Raleigh, North Carolina, November 2004. Raleigh, NC: NC Dept of Health and Human Services; June 29, 2005. Available at: www.epi.state.nc.us/epi/gcdc/ecoli/EColiReportFinal062905.pdf. Accessed September 6, 2006.
10. Wheeler C, Vogt TM, Armstrong GL, et al. An outbreak of hepatitis A associated with green onions. *N Engl J Med.* 2005;353:890-897.
11. Jay MT, Garrett V, Mohle-Boetani JC, et al. A multistate outbreak of *Escherichia coli* O157:H7 infection linked to consumption of beef tacos at a fast-food restaurant chain. *Clin Infect Dis.* 2004;39:1-7.
12. Gottlieb SL, Newbern EC, Griffin PM, et al and the Listeriosis Working Group. Multistate outbreak of listeriosis linked to turkey deli meat and subsequent changes in US regulatory policy. *Clin Infect Dis*. 2006;42:29-36.

If you would like to receive electronic copies of FOCUS on Field Epidemiology, please fill out the form below:

- NAME: ________________________________
- DEGREE (S): ____________________________
- AFFILIATION: ___________________________
- E-MAIL ADDRESS: _______________________
- May we e-mail any of your colleagues? If so, please include their e-mail addresses here:

  ____

  ____

  ____

Please fax to: (919) 919-843-5563

or mail to: North Carolina Center for Public Health Preparedness
The University of North Carolina at Chapel Hill
Campus Box 8165
Chapel Hill, NC 27599-8165

Or go online: http://www.sph.unc.edu/nccphp/focus/

UPCOMING TOPICS!

- **Conducting Traceback Investigations**
- **Conducting Environmental Health Assessments**
- **Basics of Data Analysis**
- **Advanced Data Analysis**

We are on the web!
http://www.sph.unc.edu/nccphp