VOLUME 1, ISSUE 6

FOCUS on Field Epidemiology

CONTRIBUTORS

Author:

Michelle Torok, MPH

Reviewers:

Amy Nelson, PhD, MPH

FOCUS Workgroup*

Production Editors:

Lorraine Alexander, DrPH

Gloria C. Mejia, DDS, 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 Center for Public Health Preparedness is funded by a cooperative agreement between the Centers for Disease Control and Prevention and the Association of Schools of Public Health, Project # A1011-21/21

Hypothesis Generation During Outbreaks

What is hypothesis generation and why is it important?

You are cruising down the Internet superhighway, with the breeze whipping about. Your fingers are flying at top speed; you are one with the digital age. It's a good thing, too, because your department is in serious trouble. People everywhere are coming down with a gastrointestinal illness accompanied by other symptoms, and none of the usual suspects in outbreaks seem to be related to the problem. But on the Internet, you can find the answer to almost anything! With amazing dexterity, you navigate through online medical literature, state health department postings and the CDC Web site. Unfortunately, your search is fruitless; you do not find the answer to your problem. "What are we going to do?" you cry to a coworker. "How are we going to figure out what is making these people sick?"

"Well," your coworker replies, "we could begin by generating a hypothesis."

In an outbreak investigation, a hypothesis is an educated "guess" about the source of the outbreak. Generating hypotheses about potential causative exposure(s) that may be responsible for the illness allows investigators to then test these hypotheses with an analytic study (future FOCUS issues will address analytic studies). Hypothesis generation is critical because the success of an outbreak investigation largely depends upon the quality of the hypotheses.

Overview of hypothesis generation

When an outbreak has been identified, demographic, clinical and/or laboratory data are usually obtained from the health department, clinicians, or laboratories, and these data are organized in a line listing (see FOCUS Issue 4 for more information about line listings). The next step in the investigation involves generating hypotheses about the cause of the outbreak. This step involves reviewing the information available in the literature and administering an open-ended hypothesis-generating questionnaire to some of the case-patients to attempt to learn about potential exposures. Then, based on the information contained in the line listing and the results of the literature review and/or open-ended hypothesis-generating questionnaire, specific hypotheses about the cause of the outbreak may be developed and a second structured hypothesis-testing questionnaire may be administered. The results from this second questionnaire can be used to test the hypotheses in an analytic epidemiological study.

Here we describe several approaches to generating hypotheses.

Reviewing the medical, epidemiology, microbiology, and veterinary literature (if relevant) and talking to other experts in the field to learn about previous similar outbreaks can provide valuable insight into the potential causative agent(s) and/or exposure(s):

- Example: If a causative organism has not yet been identified in an outbreak of diarrheal disease, a literature review may reveal that organisms such as *Cyclospora, Shigella, Salmonella,* Norwalk virus, *Giardia*, and *E coli* have been associated with outbreaks of diarrhea in the past. This information may help investigators focus their list of possible etiologic agents and generate hypotheses about the source and/or mode of transmission in the outbreak they are investigating.

Literature reviews are helpful in identifying organisms, risk factors and sources of exposure that have been observed in the past. However, it is important to recognize that some outbreaks are caused by previously unrecognized causative agents or through previously unrecognized modes of transmission.

- Example: In 1993 an outbreak of acute respiratory illness in the southwestern United States was found to be caused by a hantavirus. Hantavirus had not previously been associated with acute illness in the western hemisphere (1).
- Example: In 2002, several organ transplant recipients from a common donor developed unexplained fevers and encephalitis. Testing of the organ donor's blood revealed that the serum was infected with West Nile virus. Transmission of West Nile virus through organ donation had not been previously observed (2).

If the causative agent has been identified, familiarity with the microbiology, natural history and ecological niche of the organism may also help guide the investigation.

- Example: if the etiologic agent has been identified as *Blastomyces dermatitidis*, a literature review would reveal past outbreaks associated with activities that occurred near recreational water, in nitrogen-rich soil (3). This information may help identify potential exposure sites that are epidemiologically consistent with the current outbreak and are conducive to *B dermatitidis* growth.
- Example: Although hantavirus pulmonary syndrome (HPS) had not been seen in the United States before the 1993 outbreak, the investigators were able to identify rodents as a likely source of infection in the US Southwest based on HPS experience in Asia.

Existing hypothesis-generating resources may be helpful. For example, the Foodborne Outbreak Surveillance and Response Unit of the CDC has a standard questionnaire intended for use as a template for conducting initial interviews and generating hypotheses in investigations of foodborne disease outbreaks. It can be found at http://www.cdc.gov/foodborneoutbreaks/standard_questionnaire.htm.

Once the illness onset date has been determined, investigators can work backwards to estimate the maximum incubation period (for more details, see FOCUS Issue 5 on epidemic curves). Using interviews to elicit information from case-patients about exposure to known risk factors for the disease during the incubation period can help generate hypotheses about the exposure (eg, travel, food, drink, or activities). Memory aids such as calendars and receipts can help the case-patients remember details for the interview. One point to consider, however, is that these kinds of interviews can be time consuming. Also, case-patients who are interviewed may need to be re-interviewed for the analytic study.

Using the Internet to find outbreak resources and to conduct a literature review for hypothesis generation

There are several sources of electronic full-text information that can be obtained using a personal computer with an Internet connection. The following Web sites may be particularly helpful:

Centers for Disease Control and Prevention (CDC): http://www.cdc.gov

The CDC Web site includes a wealth of information on infectious diseases and outbreaks. It contains basic information on diseases in the "Diseases and Conditions" section and timely articles on health-related topics in the "Press Room" (http://www.cdc.gov/od/oc/media/), as well as surveillance, lab, and health statistics information in the "Data and Statistics" section (http://www.cdc.gov/scientific.htm). There is also a link to health-related hoaxes and rumors (http://www.cdc.gov/hoax_rumors.htm). For example, This section features an article that critiqued a 2001 *Weekly World News* story in which the authors claimed that CDC had discovered a variant of HIV transmissible through the air.

The CDC Web site also has a "Search" feature that allows the user to search the entire Web site for specific organisms, diseases, or concepts (eg, "Outbreak Investigations"). The standard hypothesis-generating foodborne disease questionnaire that can be used as a template in a foodborne disease outbreak investigation is also on the CDC Web site.

Morbidity and Mortality Weekly Report (MMWR):
http://www.cdc.gov/mmwr

The MMWR is a weekly CDC publication that contains articles on diseases reported by state and territorial health departments. While some reports originally published in the MMWR are later published in other journals, some are published only in the MMWR. Unlike most other journals, the MMWR publishes reports quickly, making it a valuable resource for gaining information in a timely manner. The Web site allows for full-text access and also has a "Search" function.

Internet search engines:

General searches on the Internet can also be helpful, using search engines such as http://www.google.com.

- Example: A search for "syphilis" will retrieve fact sheets produced by state health departments and microbiology and pathology information from universities that have coursework or research related to syphilis.

Of course, it is important to recognize that information found on the Internet is not always accurate. It is important therefore to be mindful of the credibility of different Internet-based sources (eg, a university Web site is potentially more credible than an individual's home page).

PubMed:
http://www.ncbi.nlm.nih.gov/PubMed

PubMed allows individuals to examine abstracts from biomedical journals for free. Publishers participating in PubMed electronically submit their citations prior to or at the time of publication. If the publisher has a web site that offers full-text journal access, PubMed provides a link to that site. There may be a charge for accessing the full text, but accessing abstracts through PubMed is free.

Some states or academic institutions within states have agreements with companies such as Ovid (http://www.ovid.com/site/index.jsp) to provide access to journal abstracts or full texts. If your state or affiliated institution has such an agreement, you may be able to access more information through the contracting company than through PubMed. If your state or institution does not have any such agreement, however, PubMed will allow access to all available journal abstracts.

While abstracts provide useful information, it may be necessary to read articles in their entirety. Some articles identified by a PubMed search may be available free online (for example, articles published in the CDC's MMWR). If a medical library is convenient, articles may be copied from library journals. However, if the article cannot be found online or in a library, the National Library of Medicine has a service called "Loansome Doc," that allows users to order full-text copies of articles for a fee. (Fees and service charges may vary by location.) The Loansome Doc Web site provides further details on this service (http://www.nlm.nih.gov/loansomedoc/loansome_home.html).

How to conduct a PubMed search: Conducting a literature search is not difficult, but some simple guidelines make the process even easier.

Basic rules of searching PubMed include using the Boolean operators "AND," "OR," and "NOT." These operators should be capitalized and are processed from left to right. Parentheses can also be used to group terms together so they are processed as a unit.

- Example: A search for "*Salmonella* AND eggs OR chicken" will retrieve all articles about *Salmonella* and eggs as well as all articles about chicken in general. In contrast, a search for "*Salmonella* AND (eggs OR chicken)" will retrieve all articles about *Salmonella* and eggs as well as all articles about *Salmonella* and chicken.

There are also helpful ways to broaden or limit a search.

- Example: A search for "*Salmonella*" will return approximately 50,000 related articles, which is too many to review. However, a search for

Useful resource on outbreak reports for hypothesis generation:

- ProMED-mail is an electronic network for reporting outbreaks of infectious human, plant and animal diseases that sends free email reports. More information on ProMED-mail can be found at http://fas.org/promed/.

"*Salmonella* AND outbreak AND food" will limit the number of articles to about 500. If the type of food in question (such as eggs) is known, a search on "*Salmonella* AND outbreak AND eggs" will reduce the number of articles still further, to approximately 100.

Further limitations can be placed on the search by clicking "Limits" below the search bar. For example, the search can be limited to review articles, to articles published in English or to articles concerned with humans (as opposed to animals).

Another useful tool, which can be found under "PubMed Services" is the "Cubby." The Cubby allows users to store and update searches. It requires a user-defined log-in and password, but is free. A user can run a search (eg, "*Pseudomonas* AND outbreak AND ventilator); then, once the search has been processed, the user can click the "Cubby" link to store the search. Any time after this, the user can go to the PubMed Web site, sign into the Cubby, and click the "What's new for selected" button for the desired search to retrieve any new articles for the search topic since the last search.

The "History" button can also be helpful when using PubMed. This button is found on the "Features" bar and is only available after having conducted a search. Clicking on the "History" button allows one to view a list of searches in the order in which they were run. The search number, query, time of search, and number of citations are displayed.

A final tip may help with printing citations or abstracts from PubMed. After running a search, select "Summary" from the "Display" drop down bar and "Text" from the "Send to" drop down bar at the top or bottom of the page. Then click on "Send to" for a printable version of the citations. If the entire abstract is desired, select the "Abstract" option from the "Display" drop down bar.

The PubMed Web site offers a tutorial for further tips: http://www.nlm.nih.gov/bsd/pubmed_tutorial/m1001.html.

Case study

The importance of hypothesis generation may be illustrated by a 1997 outbreak of *E coli* O157:H7 infections in Michigan and Virginia (4). This bacterial strain of *E coli* lives mainly in cattle (the reservoir), although it has been found in deer, sheep, and caribou. The bacterium passes from the reservoir through feces. Meat from the animals can be contaminated with bacteria during slaughter, and when the meat is ground, the bacteria are distributed thoughout the meat. Thus many *E coli O157:H7* outbreaks have been associated with undercooked hamburger. Feces can also contaminate fruits and vegetables, water, milk, and humans directly. Outbreaks have been associated with consumption of unpasteurized apple juice or cider, consumption of lettuce, swimming in water contaminated by cattle or humans, and direct contact with animals such as in a petting zoo.

Knowing that any of these sources might be related to the outbreaks in Michigan and Virginia, the investigators conducted hypothesis-generating interviews. They wanted to see if any of the risk factors in previous outbreaks might be related to the current outbreak. They interviewed 7 case-patients and asked them about their diet and activities in the 7 days (maximum incubation period) before their illness onset. The only common risk factor found among the case-patients was consumption of lettuce and alfalfa sprouts. Very few reported eating any meat or hamburgers.

Another factor that supported the hypothesis of a vegetable source was the demographic profile of the cases. From national surveillance data, the investigators knew that most cases of *E coli O157:H7* infection occur among children, and men are more likely than women to be infected among adults. In the Michigan outbreak, nearly 70% of the cases were women, and the median age was 31 years. This demographic profile led investigators to hypothesize that produce might be responsible for the outbreak rather than hamburgers.

Prior to 1997, alfalfa sprouts had caused outbreaks of salmonellosis, but not *E coli O157* infection. In Montana in 1995, lettuce had been implicated in an outbreak and was the leading hypothesis. However, to be thorough, the investigators included a variety of toppings used in salads, including alfalfa sprouts, in their questionnaire. To test their hypothesis, they used a case-control study method (this method will be covered in a future FOCUS issue on choosing a study design). In the analysis, alfalfa sprouts were found to be the strongest risk factor for infection in separate studies in Michigan and Virginia.

If the investigators had relied only on previous known risk factors rather than taking the time to generate a hypothesis through interviews with patients, they would have missed the cause of this outbreak.

Conclusion

In general, hypothesis generation should be guided by descriptive information gained from the investigation combined with information learned from past outbreaks. However, it is important to keep an open mind. Even when an investigation seems straightforward, it is important to be receptive to alternative explanations.

- Example: The source of a 1985 outbreak of Legionnaires' disease in a Rhode Island hospital was first thought to be contaminated drinking water (*Legionella pneumophila* had been detected in the potable water supply and had been implicated in other nosocomial Legionnaires' disease outbreaks). However, further investigation revealed that the hospital cooling towers were the source of exposure (5).

Hypothesis generation is a critical step in an outbreak investigation. Once preliminary data have been collected and a line listing has been created, investigators should have enough information to begin to think about how and why the outbreak occurred and what actions are needed to stop the outbreak (if it is ongoing) and prevent similar outbreaks in the future.

Depending on the situation, a literature review and/or open-ended hypothesis generating interviews may be necessary. The Internet is a useful resource for epidemiologic and microbiologic journal articles as well as for general information on pathogenic organisms and diseases. Preliminary data collected through hypothesis-generating questionnaires may underscore aspects of the investigation that warrant additional or more detailed data collection. Open-ended interviews can be a useful way to obtain this information.

Generating hypotheses is a creative process that requires a careful balance between maintaining an open mind and following scientifically and epidemiologically sound leads to minimize the use of limited human and financial resources. This is not always an easy task, but being aware of the importance of hypothesis generation and the resources available for it can help immensely.

<u>Glossary:</u>

Hypothesis: 1. A supposition, arrived at from observation or reflection, that leads to refutable predictions. 2. Any conjecture cast in a form that will allow it to be tested and refuted. (6)

THE UNIVERSITY OF NORTH CAROLINA

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
- Jill Koshiol, MSPH
- Pia D.M. MacDonald, PhD, MPH
- Gloria C. Mejia, DDS, MPH
- Sally B. Mountcastle, PhD, MSPH
- Amy Nelson, PhD, MPH
- E. Danielle Rentz, MPH
- Tara P. Rybka, MPH
- Cheryl R. Stein, MSPH
- Michelle Torok, MPH
- Nicole Tucker, MPH
- Drew Voetsch, MPH

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 email any of your colleagues? If so, please include their email 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/

REFERENCES:

1 CDC. Outbreak of acute illness — Southwestern United States, 1993. *MMWR Morb Mortal Wkly Rep.* 1993; 44(22):421-424.

2. CDC. Public health dispatch: West Nile Virus infection in organ donor and transplant recipients — Georgia and Florida, 2002. *MMWR Morb Mortal Wkly Rep.* 2002; 51(35):790.

3. Davies SF, Sarosi GA. Epidemiological and clinical features of pulmonary Blastomycosis. *Semin Respir Infect.* 1997; 12(3):206-218.

4. Breuer T, Benkel DH, Shapiro RL, et al. A multistate outbreak of *Escherichia coli O157:H7* infections linked to alfalfa sprouts grown from contaminated seeds. *Emerg Infect Dis*. [serial online] 2001;7 (6):977-82. Available from: http://www.cdc.gov/ncidod/eid/vol7no6/breuer.htm.

5. Garbe PL, Davis BJ, Weisfeld JS, et al. Nosocomial Legionnaires' disease. Epidemiologic demonstration of cooling towers as a source. *JAMA.* 1985;254:521-524.

6. Last JM, ed. *A dictionary of epidemiology*. New York, NY: Oxford University Press, Inc; 2001.

UPCOMING TOPICS!

- Hypothesis Generating Interviews
- Designing Questionnaires for Outbreaks
- Interviewing Techniques for Epi Studies
- Introduction to Forensic Epidemiology
- Differences between Public Health and Law Enforcement Investigations

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