



# FOCUS on Field Epidemiology

## Developing a Questionnaire

### CONTRIBUTORS

#### Authors:

**Paul A. Stehr-Green, DrPH, MPH \***

**Jeanette K. Stehr-Green, MD \***

**Amy Nelson, PhD, MPH**

#### Reviewers:

**FOCUS Workgroup\*\***

#### Production Editors:

**Lorraine Alexander, DrPH**

**Gloria C. Mejia, DDS, MPH**

#### Editor in chief:

**Pia D.M. MacDonald, PhD, MPH**

\* NCCPHP Consultants

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

Consider trying to find the word “psychology” in Webster’s Dictionary if no one has tipped you off that it starts with a “p.” You will have a hard time finding it if you do not already have some idea about how the word is spelled (the first letter or two). Without this knowledge, it might take a long time to find “psychology”. In fact, you might never find it.

The same is true for epidemiologic studies. If you do not know what you are looking for, you are unlikely to find it. To conduct a successful epidemiologic study, you need a solid understanding (or a well-educated guess!) about what is behind the problem you are investigating. That is, before you design the study, create a questionnaire, or begin data collection, you should have a good idea about who is being affected by the problem and how widespread it is. You should have narrowed down the possible sources of the problem and have some sense of the contributing factors. If you ask the right questions in the right way, you will get answers that are meaningful and useful. And isn’t that the point of doing the study in the first place?

*If you don't ask the right questions, you don't get the right answers ... Asking questions is the ABC of diagnosis. Only the inquiring mind solves problems.*

- Edward Hodnett

In this issue of FOCUS we discuss asking the right questions in the right way as part of an epidemiologic study. First we cover the steps for creating a questionnaire and the categories of information typically collected. We then talk about the questions themselves and about organizing them into a questionnaire.

In creating a questionnaire for an epidemiologic study, investigators typically follow the steps shown in the box below. One of the most common errors in questionnaire development is to start by writing the questions. The first step in creating a questionnaire is actually to identify the leading hypotheses about the source of the problem. You can then identify the information needed to test your hypotheses and identify the information needed for the

#### Steps in creating a questionnaire

1. Identify the leading hypotheses about the source of the problem.
2. Identify the information needed to test the hypotheses.
3. Identify the information needed for the logistics of the study and to examine confounding factors.
4. Write the questions to collect this information.
5. Organize the questions into questionnaire format.
6. Test the questionnaire.
7. Revise the questionnaire.
8. Train interviewers to administer the questionnaire.



**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.

logistics of the study and to examine confounding factors. Only after you have identified the information you need are you ready to write the questions and organize them into a questionnaire.

So what information do you need to collect? An epidemiologic questionnaire typically includes five categories of information: identifying information, demographic information, clinical information, exposure or risk factor information, and the source of the information.

#### Categories of information for an epidemiologic questionnaire

- Identifying information
- Demographic information
- Clinical information
- Exposure or risk factor information
- Source of information

**Identifying information** is important for the logistics of the study and includes the respondent's name or other identifier, address, and telephone number. This information identifies the subject, allows updating of the questionnaire as more information becomes available, and can be used to link the questionnaire to other records (e.g., laboratory results or other questionnaires). It can also prevent duplicate entry of records.

**Demographic information** includes items like age, sex, educational level, and location. These items are used to characterize the population at risk and to explore the problem under investigation. The information is also important in the search for possible confounders (factors that distort the apparent exposure-disease relationship). Demographic characteristics need to be evaluated to determine whether they are affecting the relationship you are seeing (or think you are seeing) between an exposure and a disease.

**Clinical information** includes signs and symptoms of the disease, date of onset of illness, and results of laboratory testing. This information allows you to characterize the illness, decide who has the outcome of interest, and chart the time course of the problem.

**Exposure or risk factor information** is used to test the hypotheses under investigation. It will probably be the major focus of your questionnaire. It should be specific

to the problem under investigation and reflect your hypotheses about the source of the problem. Exposure or risk factor information often includes:

- the respondent's exposure to the factor of interest
- the route of exposure
- the amount of exposure
- the timing of exposure
- other details of exposure (e.g., brand, distributor).

For example, an *E. coli* O157:H7 outbreak was detected in Michigan in June 1997. Through the hypothesis-generation process, alfalfa sprouts and lettuce were identified as the most probable vehicles of transmission. Thus they were the leading hypotheses.

#### *E. coli* O157:H7 Infection in Michigan

- **Leading hypotheses on vehicle of transmission:**
  - consumption of alfalfa sprouts
  - consumption of lettuce
- **Information to test hypotheses:**
  - Did you eat either of these?
  - When did you eat them?
  - How much did you eat?
  - Where did you buy them?
  - Were they from a particular producer or distributor?

What information do we need to test those hypotheses? Obviously, we want to know if the respondent ate either of these items. What else might be important? We also want to know when the alfalfa sprouts or lettuce was eaten (eating these items before or after the period of interest is not of concern). Since there could be a dosage effect, we usually want to know how much was eaten. (With *E. coli* O157:H7, a small number of organisms can lead to infection. Therefore, exploring the amount eaten is unlikely to be of great interest in this outbreak. But it might be of concern with other pathogens.) Finally, if the outbreak is due to alfalfa sprouts or lettuce, it might be limited to alfalfa sprouts or lettuce from a particular source. Therefore, we want to

know where the respondent bought or ate these items and whether they were from a particular producer or distributor.

**The source of the information** includes the individual supplying the information as well as the person receiving it (i.e., the interviewer). Identification of the person supplying the information provides some insight into its validity. Is this person a study subject or a surrogate such as the spouse or parent? (In most situations, the actual subject of the study will provide better information than others who speak on behalf of the subject, although this is certainly not the case when dealing with infants or deceased persons!) Identification of the interviewer can also be important. If some questions are left unanswered, the handwriting is illegible, or responses are nonsensical, the interviewer might be able to correct these problems. Tracking the interviewer can also help identify a more serious problem: if inadequately trained, an interviewer may consistently misread a question, inappropriately interpret a question for respondents, or misquote respondents' answers. Of course, if interviewers are properly trained, this problem will be greatly minimized (a future FOCUS issue will deal with interviewing techniques).

### Writing questions

Now that you have identified the information you need, you can begin writing the questions.

The exact wording of questions depends on your knowledge of the problem, the hypotheses being tested, the kind of information you wish to collect, and the affected population. Therefore, it will be different with each epidemiologic investigation. But some generalizations can be made about the types of questions and their construction.

Three types of questions are used in questionnaires for epidemiologic studies:

- Open-ended questions
- Fill-in-the-blank questions
- Closed-ended questions
  - with categorical response options
  - with ordinal response options

### Open-Ended Questions

As discussed in the last issue of FOCUS (Hypothesis-Generating Interviews), open-ended questions allow the

respondent to say whatever he or she desires, rather than requiring a response based on a specific list of choices. Thus, the possible responses are limitless. Open-ended questions give the respondent an opportunity to express his/her own perspectives about a subject and are useful in characterizing attitudes, beliefs, and behaviors. However, investigators tend to limit their use of this question type to hypothesis-generating activities.

### Fill-in-the-Blank Questions

The second type of question is "fill-in-the-blank." Fill-in-the-blank questions, like open-ended questions, do not provide any response choices. But with fill-in-the-blanks, it is expected that the response will be relatively short, a word or two. Fill-in-the-blank questions should be used when the possible response categories are numerous and the question is measuring a simple respondent attribute (such as age, educational level, place of residence), collecting a date (birth date, date of onset of illness, date of exposure), or quantifying something specific.

- Examples of fill-in-the blank questions are:

In what county do you live? \_\_\_\_\_  
(DK=don't know R=refused)  
What is your date of birth? \_\_\_\_\_  
(mm/dd/yyyy)  
(DK=don't know R=refused)

Fill-in-the-blank questions often require that the investigator categorize and code the responses.

### Useful resources: Online questionnaires

- **Centers for Disease Control and Prevention:**  
[http://www.cdc.gov/foodborneoutbreaks/standard\\_questionnaire.htm](http://www.cdc.gov/foodborneoutbreaks/standard_questionnaire.htm)
- **Food Risk Clearinghouse:**  
[http://www.foodriskclearinghouse.umd.edu/dose\\_r\\_esp.htm](http://www.foodriskclearinghouse.umd.edu/dose_r_esp.htm)
- **Minnesota Department of Health:**  
[http://www.health.state.mn.us/divs/hpcd/chp/hpkit/pdf/hcheck\\_main\\_ws1.pdf](http://www.health.state.mn.us/divs/hpcd/chp/hpkit/pdf/hcheck_main_ws1.pdf)
- **Oregon Department of Human Services:**  
<http://www.ohd.hr.state.or.us/acd/foodrpt.cfm>

*Closed-Ended Questions*

A final type of question is the closed-ended question. These questions have response choices from which the respondent must choose. Closed-ended questions are used when possible responses are known and the range of responses is narrow enough to be provided in a short list to respondents. Unlike open-ended and fill-in-the-blank questions, closed-ended questions do not require the investigator to categorize and code the responses, but they do require the investigator to:

- anticipate the most likely responses to the question
- present the responses as a list of mutually exclusive choices (choices that do not overlap)
- state the responses in a clear and concise manner that can easily be understood by respondents

There are two types of closed-ended questions: categorical and ordinal. In categorical closed-ended questions, the available responses fit into categories that have no particular order or inherent numerical value with respect to each another.

- For example, to explore why Medicare beneficiaries did not get vaccinated against influenza, the following categorical closed-ended question could be posed:
  - What was the main reason you did not get the flu shot last winter?
    1. Did not know it was needed
    2. Could cause influenza
    3. Could cause side effects
    4. Vaccine not available
    5. Did not think it would prevent influenza
    6. Did not think of it
    7. Not at risk for getting influenza
    8. Do not like shots or needles
    9. Doctor recommended against it
    10. Doctor did not recommend it
    11. Unable to get to location
    12. Do not know
    13. Refuse to answer

In this example, you can see that the responses do not have any numerical relationship to each other. “Did not know it was needed” is not inherently higher or lower in value than “Could cause influenza,” and this in turn bears no quantitative relationship to “Could cause side effects,” and so on.

In ordinal closed-ended questions, the available responses tend to describe a range of choices and have a quantitative value with respect to each other. For example, the value of the response choice indicated by a code of “1” could be more than the value indicated by a code of “2.” Ordinal closed-ended questions are useful for determining frequency of participation, degree of involvement, or intensity of feelings.

- For example, in seeking information on how much “could cause side effects” influences a person’s decision to get vaccinated against influenza, the following ordinal closed-ended question may be posed:
  - Describe your level of concern about the side effects caused by the flu shot (circle one):
    1. Not concerned
    2. A little concerned
    3. Somewhat concerned
    4. Moderately concerned
    5. Very concerned
 DK = Do not know  
 R = Refuse to answer

Here, the lower numbered responses suggest less concern and the higher numbered responses suggest more concern. The exact relationship between the choices, however, may be difficult to determine (and open to interpretation).

All three question types – open-ended, fill-in-the-blanks, and closed-ended – have their place. When do you use which types? You select the type based on the kind of information you need and your expectations about that information. In the early exploration of a problem, when you know little about it, you will be more likely to use open-ended questions. As information about the problem grows, you will be able to anticipate possible responses and will be more likely to use closed-ended questions.

*Question Wording*

Choosing the type of question is just the first issue. Care needs to be given to the wording of the question and the response categories (if applicable). Here are a few guidelines for writing questions for an epidemiologic study:

- Use language respondents can understand. Avoid technical jargon, slang, abbreviations, and stuffy-sounding bureaucratic words. Depending on the level of education of your respondents, you might need to simplify the language.

- If the questionnaire is translated from one language (say, English) to another (say, Chinese), test it with native speakers. A good test is to have the translated questionnaire “backtranslated” (that is, translated from Chinese back to English). Backtranslation allows investigators to see or hear what the respondents will be seeing or hearing and may reveal problems.
- Limit each question to a single idea. Combining questions can result in responses that are conflicting and may not answer the question you think you asked.

For example:

Use: 1) *Have you had vomiting?*  
2) *Have you had diarrhea?*

Not: *Have you had vomiting and diarrhea?*

- Word each question as precisely as possible. Leave little to the respondent’s interpretation. Use caution with general adjectives and adverbs that may have different meanings for different people (e.g., big, bad, nice, etc.).

For example:

Use: *Have you had three or more loose bowel movements in any 24-hour period between April 25 and May 1?*

Not: *Did you have severe diarrhea?*

- Do not phrase questions in a way that suggests a response (i.e., a leading question) or implies a value judgment. The interviewer’s attitude, as perceived by respondents, can influence their response.
- Avoid double negatives. They are rarely not confusing!
- Always include a “Do not know” or “Refuse to answer” category with potential responses. This will help you distinguish between respondents who do not answer a question because they do not remember the answer and those who choose not to answer or just skipped the question.
- In creating the responses for a closed-ended question, be sure that the categories cover all potential responses and are mutually exclusive (i.e., categories do not overlap.) Every response should fit into one and only one category.

For example, when asking respondents how old they are, the following choices may be given: A. 1-20 yrs., B. 20-40 yrs., C. 40-60 yrs., etc. Where does an infant less than 1 year of age fit? If respondents are 20 years old, do they fit into category A or B? The following categories would be much better: A. Less than 20 yrs., B. 20-39 yrs., C. 40-59 yrs., etc.

### Putting it all together

You now know the categories of information collected in an epidemiologic questionnaire and the general types of questions. Here some tips to help a questionnaire flow smoothly.

#### Introduction

An introduction allows you to identify the organization sponsoring the study and explain the study purpose. It lends credence to the undertaking and increases the likelihood that respondents will participate and answer honestly. However, you do not want to provide too much information about the investigation (such as the suspected source of the problem). The purpose of the study should be explained in general terms to help respondents understand the importance of the interview and their part in the process. The introduction should state how long the interview is likely to take and reassure participants that their answers will be strictly confidential.

*Quot Homines Tot Sententiae*

*(So many men so many questions.)*

Terence (185 BC - 159 BC)

*It is not every question that deserves an answer.*

Publilius Syrus (~100 BC), Maxims

#### Length

In general, a questionnaire should be as short as possible and should focus on the hypotheses being tested in the study. A lengthy questionnaire uses a lot of resources and tires respondents. Focus on your leading hypotheses and the particular pieces of information you need to test these hypotheses. Then try to strike a balance between testing the hypotheses and taking advantage of opportunities to gather ‘extra’ information that is of interest to the investigation.

#### Logic

The questions should be organized to promote the development of rapport between the respondent and the interviewer. The best way to do this depends on the outbreak and the population. Commonly used methods of organization include grouping together similar types

of information or topics; asking general questions first, followed by more specific questions; asking the least sensitive questions first, leaving questions about sexual habits, religious beliefs, political orientation, or income for later when rapport has been developed; and asking the most important questions first (relating to your leading hypotheses), followed by less critical questions in case the respondent loses interest. Regardless of which approach you take, the questionnaire should appear logically organized and not skip from topic to topic.

*Layout*

The organization of the questionnaire should make it easy to read and complete. Instructions should be clearly stated. Questions and pages should be numbered. Each page should include an identifying code for the respondent. Possible responses to questions and the place for recording responses should be clearly separated from the questions.

You should include skip patterns to avoid asking irrelevant questions. A skip pattern usually begins with a “screening question” that tells the interviewer to know whether a set of subsequent questions pertain to a particular respondent. If they do, the subsequent questions are read. If not, the interviewer skips over the questions and continues with the next set.

Here is an example of a skip pattern. Note that Question 1 is a screening question

1. In the past 7 days, did you eat any mixed lettuce? This lettuce can be pre-bagged. Or it can be picked from a bin, and you use tongs to put it into a bag yourself. It is sometimes called Spring or Mesclun mix.

Yes    No    Don't know    Refused

If no, skip to question 3.

2. Was the type of mix called:
  - a. Italian?
 

Yes    No    Don't know    Refused
  - b. Caesar?
 

Yes    No    Don't know    Refused

3. In the last 7 days, did you eat sprouts, such as alfalfa or bean sprouts? These may have been eaten as part of a salad or as part of any other food item such as sandwiches, tacos, etc.
- Yes    No    Don't know    Refused

Using skip patterns can prevent a respondent from hearing and having to answer questions that do not pertain to him or her. Thus, they save a lot of time and are less wearing on the patience of respondents. If you use skip patterns, however, make sure that they are clearly marked and easy to follow.

*Wrap it up*

Finally, an ending statement is also important. You should thank respondents for their input and their time. You should also provide them with a means to contact the study investigators if they have questions or remember additional information pertinent to the investigation. Providing a contact number will also reassure respondents that the interview was part of a legitimate undertaking.

**Conclusion**

In an epidemiologic investigation, it is sometimes difficult to know how to form the right questions, let alone understand the answers. The first step is to carefully frame and refine the hypotheses under investigation and decide what information is needed to test those hypotheses. With this information in hand, you can then develop a series of open-ended, fill-in-the-blank, and closed-ended questions to obtain the information, paying careful attention to the specific wording and organization. By following this time-tested approach, you are more likely to get the right information from the right people at the right time.

## 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](mailto:nccphp@unc.edu)

## REFERENCES:

University of North Carolina at Chapel Hill School of Public Health, Department of Maternal and Child Health (2003). Data Skills Online: A Maternal and Child Health Toolbox. "Designing Questionnaires." <http://www.sph.unc.edu/toolbox/>

### FOCUS Workgroup:

- Lorraine Alexander, DrPH
- 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

## UPCOMING TOPICS!

- Interviewing Techniques for Epi Studies
- Selecting a Study Design
- Introduction to Forensic Epidemiology
- Differences between Public Health and Law Enforcement Investigations
- Environmental Health Assessment

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/>

We are on the web!

<http://www.sph.unc.edu/nccphp>