VOLUME 5, ISSUE 5

FOCUS on Field Epidemiology

Introduction to Public Health Surveillance

CONTRIBUTORS

Author:

Michelle Torok, MPH

Meredith Anderson, MPH, CPH

Reviewers:

FOCUS Workgroup*

Production Editors:

Tara P. Rybka, MPH

Lorraine Alexander, DrPH

Rachel A. Wilfert, MD, MPH, CPH

Editor in chief:

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

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

UNC GILLINGS SCHOOL OF GLOBAL 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.

The stakeout is on-going. The devoted epidemiologist stays up all night waiting, watching, hoping the dread disease will poke its nose out of its apartment and dart down the street. If this happens, the epidemiologist will be hot on its tracks, ready to catch the disease in the act of taking out another victim. Is that how your health department conducts surveillance?

Surveillance is one of the most important tasks of an epidemiologist. But while most public health professionals know what surveillance is, few know how to actually conduct surveillance. Moreover, very little has been written for those who wish to learn more about surveillance. This issue of *FOCUS* is therefore designed for persons who need to carry out surveillance activities but have little prior experience or training. It should also be helpful for people who may not perform surveillance, but would like to better understand the process and reasoning behind surveillance methods and interpretation.

What Is Surveillance and Why Is It Important?

Before discussing how to conduct surveillance, let's briefly review what it is and what its uses are.

According the Centers for Disease Control and Prevention (CDC), epidemiologic surveillance is "the ongoing systematic collection, analysis, and interpretation of health data essential to the planning, implementation, and evaluation of public health practice, closely integrated with the timely dissemination of these data to those who need to know." (1)

It is important to note that collecting data is merely one step in carrying out surveillance. A critical goal of surveillance is to control and/or prevent diseases. Therefore, any data collected must be organized and carefully examined, and any results need to be communicated to the public health and medical communities. It is vital to communicate results during a potential outbreak so that the public health and medical communities can help with disease prevention and control efforts, but it is also important during non-outbreak times to provide information about baseline levels of disease. These baseline measurements provide important information to public health officials in monitoring health at a community level and serve as important references in any future outbreaks.

Surveillance systems are classified as passive or active. Passive surveillance occurs when local and state health departments rely on health care providers or laboratories to report cases of disease. The primary advantage of passive surveillance is its efficiency: it is simple and requires relatively few resources. The disadvantage is the possibility of incomplete data due to underreporting. The majority of public health surveillance systems are passive, but in some situations it is preferable to conduct active surveillance.

Active surveillance occurs when the health department contacts health care providers or laboratories requesting information about conditions or diseases to identify possible cases. This method requires more resources than passive surveillance, but is especially useful when it is important to identify all cases. For example, between 2002 and 2005, active surveillance was used to detect adverse events associated with the smallpox vaccine. (2)

Surveillance information has many uses, including monitoring disease trends, describing the natural history of diseases, identifying epidemics or new syndromes, monitoring changes in infectious agents, identifying areas for research, evaluating hypotheses, planning public health policy, and evaluating public health policy and interventions.

Examples of important ways that surveillance data has been used include:

- Evaluating the impact of national vaccination campaigns;
- Identifying AIDS when it was a previously unknown syndrome;
- Estimating the impact of AIDS on the US health care system in the 1990s (by using mathematical models based on surveillance data);
- Identifying outbreaks of rubella and congenital rubella among Amish and Mennonite communities in 6 states in 1990 and 1991; (3) and
- Monitoring obesity, physical activity, and other factors that are important indicators for chronic diseases such as diabetes.

How to Conduct Surveillance

Surveillance data allow the description and comparison of patterns of disease by person, place, and time. There are several ways to describe and compare these patterns, ranging from straightforward presentations to statistically complex analyses. In this *FOCUS* issue, we will concentrate on simple techniques.

Person

When available, demographic characteristics such as gender, age, race/ethnicity, occupation, education level, socio-economic status, sexual orientation, or immunization status can reveal important disease trends. For example, in looking at *Streptococcus pneumoniae*, a common cause of community-acquired pneumonia and bacterial meningitis, examining the distribution of cases by race provides important information about the burden of disease in different populations.

Table 1 shows data collected on *Streptococcus pneumoniae* from the CDC Emerging Infections Program Network, a surveillance program that collects data from multiple counties in 10 different US states. (4)

These data show that the majority of the cases reported were among whites. However, we can draw only limited conclusions from these data because race was not recorded for 684 of the cases (15%). Furthermore, inferences about the incidence of *S. pneumoniae* in different racial groups could not be made based on this table even if there were race information for all cases, because the table shows only the *number* of reported cases, not the *rate* of reported cases. The total number of individuals by race would be needed to determine whether or not there is a disproportionate burden of disease among whites or blacks.

Table 1. Reported cases of *Streptococcus pneumoniae* by race, 2006 (4)

Race	Number
White	2,614
Black	1,095
Other	213

Unknown race (n=684) distributed among knowns

Now let's look at the same data (Table 2) when 2006 population estimates of the total number of persons in each racial category were used to calculate disease rates. (4)

Table 2. Rates of invasive pneumococcal disease by race, 2006

Race	Number	Rate*
White	2,614	11.8
Black	1,095	25.1
Other	213	12.8

*Cases per 100,000 population of surveillance areas

While Table 1 showed that whites had the highest *number* of cases, Table 2 indicates that the *rate* of disease was highest among blacks. Using rates and stratifying by race thus provides important information about disease burden in different populations that would not be apparent by just looking at total case numbers.

Rates— A rate is "an expression of the frequency with which an event occurs in a defined population." In epidemiology, using rates rather than raw numbers is essential to compare different classes of persons or populations at different times or places. (5)

$$\text{Rate} = \frac{\text{number of events in a specified period}}{\text{average population during the period}}$$

Place

When examining surveillance data by place, it is best to characterize cases by place of exposure rather than by the place at which cases are reported, since the two may differ and the place of exposure is more relevant to the epidemiology of a disease.

For example, travelers on a cruise ship may have been exposed to a disease just prior to disembarking but may not become symptomatic and be diagnosed until they return to their various home locations. Or a person may have been exposed to disease in his small rural town but is referred to a tertiary care center 100 miles away, where the disease is diagnosed and reported. In both of these examples, the place of exposure, rather than the place of diagnosis and reporting, is the important factor for monitoring and tracking disease events.

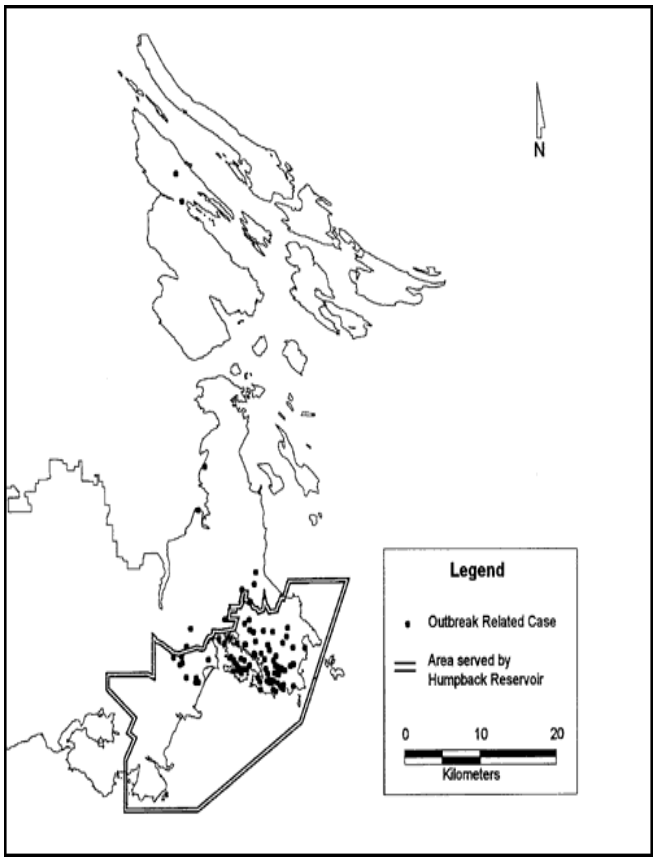
Data by geographic location can be presented in a table, but it may also be helpful to use maps to facilitate recognition of spatial associations in the data. (See *FOCUS* Volume 5, Issue 2: Mapping for Surveillance and Outbreak Investigation for a discussion of maps and visual presentation of information.) Inferential analysis can also be done using multilevel modeling and other statistical methods. Modeling of surveillance data by place is beyond the scope of this *FOCUS* issue, but resources for obtaining further information on this topic are listed in the "Resources" section below.

Maps on which a dot or symbol marks a case of disease are called spot maps. These maps can be made by indicating the exposure locations of reported cases of disease on a hard copy of a map with pins or a colored pen or with geographic information systems (GIS). GIS are computer programs designed for storing, manipulating, analyzing, and displaying data in a geographic context. GIS can be very useful for mapping surveillance data by place. Epi Map, which is part of Epi Info™, is one GIS program that can be downloaded for free at http://www.cdc.gov/epiinfo to assist with map making.

Figure 1 is an example of a spot map used to show the geographic spread of cases in a 1995 outbreak of toxoplasmosis thought to be associated with a municipal water system in British Columbia, Canada. (5)

Spot maps are helpful in showing the geographic distribution of cases, but since population size at each location on the map is not taken into account, this method should not be used to assess disease risk.

Figure 1. Computer-generated spot map showing outbreak-related cases of toxoplasmosis in Vancouver Island, British Columbia, Canada, 1995 (6)



Resources:

Centers for Disease Control and Prevention. Resources for creating public health maps. http://www.cdc.gov/epiinfo/maps.htm. Updated August 14, 2008. Accessed August 22, 2008.

Clarke KC, McLafferty SL, Tempalski BJ. On epidemiology and geographic information systems: A review and discussion of future directions. *Emerg Infect Dis*. 1996; 2(2):85-92.

Time

An easy way to examine surveillance data by time is to describe the distribution of cases over time, and compare the number of cases reported in a particular time period of interest (e.g., weeks, months, years) to the number of cases reported during a similar historical period. Since there is usually a delay (sometimes a long delay of months to years) between disease onset and the date when a disease is reported, it is preferable to use the date of onset, if available, rather than the date of report.

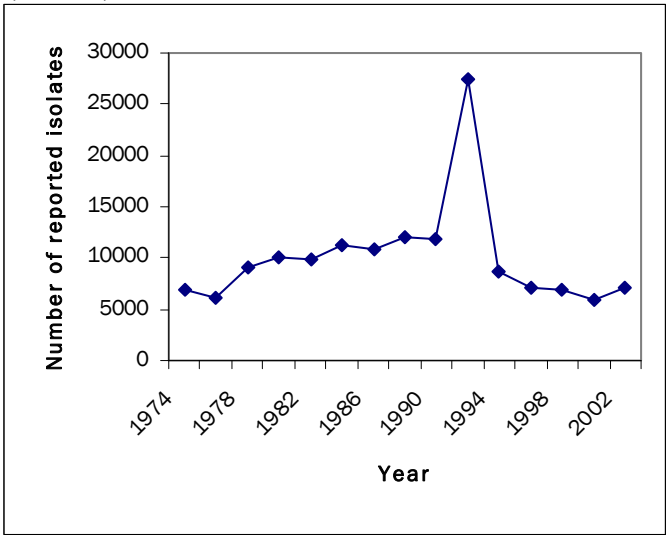
Line graphs are useful for presenting surveillance data by time, and they are especially helpful for examining data that are not likely to have much short term variation (for example, there is limited variation in the number of AIDS cases reported each month). Line graphs provide valuable qualitative information; disease outbreaks are often obvious from visual inspection of the data and may not require a quantitative analysis.

Figure 4 gives an example of this method using fabricated data. It shows the number of reported cases of *Salmonella typhimurium* for 2-year time intervals from 1974 to -2002. The spike in 1994 indicating an outbreak of *S. typhimurium* is obvious without quantitative analysis.

If the data are available, a line graph may be used to plot incidence rates for a time period of interest. An incidence rate is the number of new cases of a disease that occur during a specified time interval in a population at risk for developing the disease. For surveillance purposes, the number of new cases of disease that are reported may be used as a proxy for overall disease occurrence. Often this value is multiplied by 1,000 or 100,000 to improve interpretability.

Reporting incidence *rates* rather than *numbers* is particularly important if the population has changed in size or characteristics (e.g., the addition of towns to a surveillance region has increased population size, or an influx of migrant workers has significantly changed the demographics).

Figure 4. *Salmonella typhimurium*: reported isolates by year (fictitious), 1974-2002



Standardization

One issue that arises when using rates in the context of surveillance is *standardization*. As noted above, a rate is made up of a numerator and a denominator. Surveillance data are often numerator data (the number of cases reported in a specific time period), but the utility of these raw numbers is limited because they do not take into account the size of the population or the distribution of demographic factors such as age or gender.

Rates allow for more meaningful comparisons of disease experience over time within a population, among subpopulations, or between populations because rates take into account the size of the population and the time period involved. (3)

Crude rates are often calculated using surveillance data. A crude rate is the number of events of interest (such as reported cases of disease) for a specific period of time for the entire population. It is only appropriate to compare crude rates if populations are similar with respect to factors related to the disease of interest, such as age, gender, or race. For example, it would be inappropriate to compare the rate of prostate cancer in a population that had a high proportion of elderly men with the rate in another population that contained mostly young men, since the risk of prostate cancer increases with age.

Standardization is a method that is used to remove the effects of differences in confounding variables such as age when comparing two or more populations. Standardization results in adjusted rates, and is particularly useful when comparing rates in different populations (e.g., comparing state data to national data) when comparison of crude rates may be misleading if the populations differ on key variables.

Several techniques are used for standardization; the most common of these uses weighted average rates specific to potential confounding variables, based on specified distribution of the variables. (5) However, a detailed discussion of standardization methods is outside the scope of this *FOCUS* issue.

Data Presentation

As noted at the beginning of this issue, once data have been collected and organized, it is important to share any results with the public health and medical communities. This means that surveillance data must be presented in a way that is easy to understand and interpret.

There are many different ways of displaying surveillance data. Line graphs are useful for displaying data by time, while maps are useful for presenting data in a geographic context. Other ways of presenting surveillance data include graphical displays such as histograms, frequency polygons, box plots, scatter diagrams, bar charts, pie charts, or stem-and-leaf displays; spot or chloropleth maps; and single/multivariable tables. (3) The choice of a particular graph or table depends on the type of data, but the presentation should be simple and easy to follow.

When developing graphs, tables, or maps for presentation, you should provide all the information necessary to interpret the figure without referring to the text. This includes a concise title that describes the subject or disease and the time and place, when relevant. Additional guidelines for tables and graphs are shown in Figure 5. You should also define any abbreviations or symbols and note any data exclusions. (3)

Conclusion

Surveillance is a valuable epidemiologic tool that can serve many purposes. When surveillance data is collected, analyzed, interpreted, and reported appropriately, these data can provide important information about disease patterns to inform public health practice and policy.

Figure 5. Guidelines for developing tables and graphs (3)

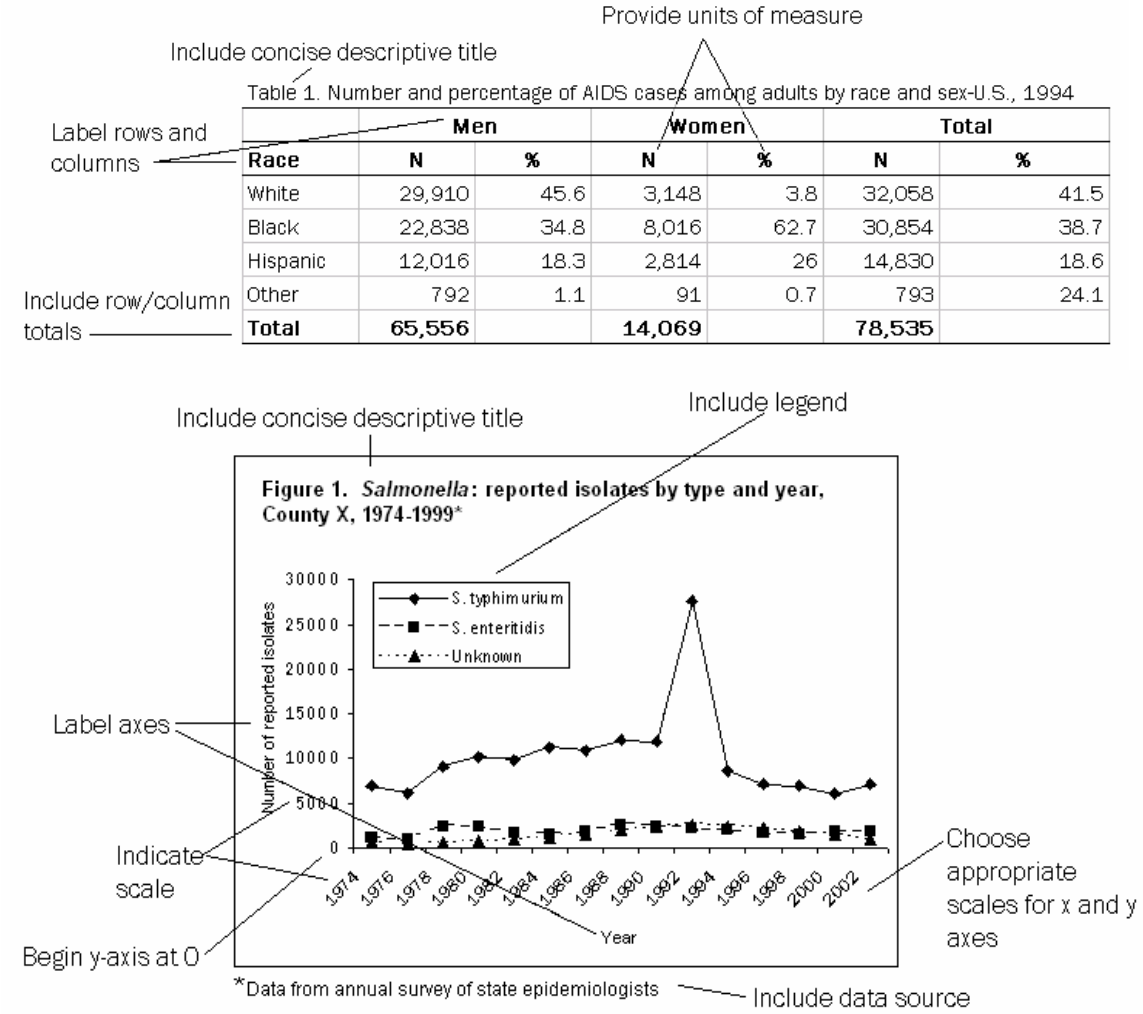


Table 1. Number and percentage of AIDS cases among adults by race and sex-U.S., 1994

	Men		Women		Total	
Race	N	%	N	%	N	%
White	29,910	45.6	3,148	3.8	32,058	41.5
Black	22,838	34.8	8,016	62.7	30,854	38.7
Hispanic	12,016	18.3	2,814	26	14,830	18.6
Other	792	1.1	91	0.7	793	24.1
Total	**65,556**		**14,069**		**78,535**	

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
- Meredith Anderson, MPH, CPH
- Lauren N. Bradley, MHS
- Jennifer A. Horney, MA, MPH, CPH
- Pia D.M. MacDonald, PhD, MPH, CPH
- Amy Nelson, PhD, MPH, CPH
- Tara P. Rybka, MPH
- Rachel A. Wilfert, MD, MPH, CPH

REFERENCES:

1. Thacker SB, Berkelman RL. Public health surveillance in the United States. *Epidemiol Rev*. 1988;10:164-190.
2. Thomas TN, Reef S, Neff L, Sniadack MM, Mootrey GT. A review of the smallpox vaccine adverse events active surveillance system. *Clin Infect Dis*. 2008;46 Suppl 3:S212-S220.
3. Janes GR, Hutwanger L, Cates Jr W, Stroup DF, Williamson GD. Descriptive Epidemiology: Analyzing and Interpreting Surveillance Data. In: Teutsch SM, Churchill RE, eds. *Principles and Practice of Public Health Surveillance*. New York, NY: Oxford University Press, inc, 2000:112-167.
4. Centers for Disease Control and Prevention. Active Bacterial Core Surveillance Report (ABCs), Emerging Infections Program Network, *Streptococcus pneumoniae*, 2006. http://www.cdc.gov/ncidod/dbmd/abcs/survreports/spneu06.pdf2007. Published 2007. Accessed August 21, 2008.
5. Last JM, ed. *A Dictionary of Epidemiology*. 3rd ed. New York, NY: Oxford University Press, Inc, 1995.
6. Eng SB, Werker DH, King AS, et al. Computer-generated dot maps as an epidemiologic tool: Investigating an outbreak of toxoplasmosis. *Emerg Infect Dis*. 1999;5 (6):815-819.

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://nccphp.sph.unc.edu/focus/

Coming In 2009: FOCUS goes deeper!

Starting with Volume 6, *FOCUS* will be a semi-annual publication with longer, more in-depth discussions of field epidemiology topics based on the Applied Epidemiology Competencies from CSTE/CDC.

UPCOMING TOPICS!

- **Public Health Surveillance Systems**

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