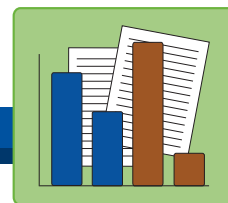# Module 1: An Overview of Public Health Data

Transcript

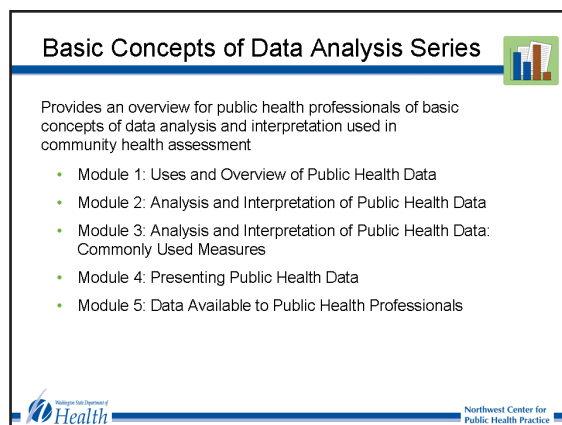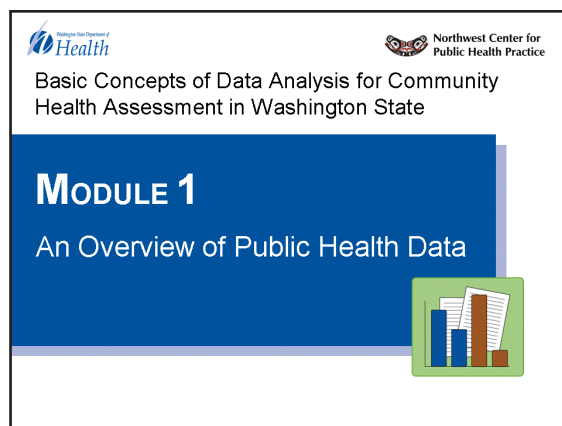## An Overview of Public Health Data

Welcome to "An Overview of Public Health Data," the first module in the series on Basic Concepts of Data Analysis for Community Health Assessment in Washington State. I'm Jane Ballard. I helped create this course, along with a team of other community health assessment experts at both local and state health departments in Washington and Oregon. I have a doctorate in epidemiology and manage the Health Statistics and Assessment Program at Snohomish Health District. I have worked at Snohomish Health District since 2000.

Basic Concepts of Data Analysis Series

This series provides an overview for public health professionals of the basic concepts of data analysis and interpretation used in Community Health Assessment. The training is intended to help professionals who work in public health practice at state and local agencies hone their assessment skills.

- Module 1 provides an overview of public health data sources and uses.
- Module 2 introduces the analysis and interpretation of public health data.
- Module 3 continues the discussion of the analysis and interpretation of the public health data.
- Module 4 provides information on how to present public health data, and
- Module 5 describes data available to public health professionals.

This series was developed by Washington State Department of Health in partnership with the Northwest Center for Public Health Practice. Many of the examples use Washington State or county level data, but the concepts they illustrate are relevant to public health professionals in any location.

Module 1: An Overview of Public Health Data

Transcript
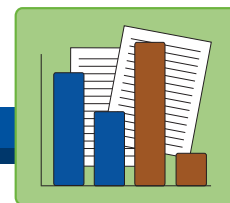
Learning Objectives

This overview module introduces the types of data that are used in public health assessment activities. It takes about 30 minutes to complete.

By the end of this module, you should be able to:

- Identify uses of data in public health core functions
- List at least three common data sources used to characterize the health or disease status of the community
- List five key attributes of data, and
- List three elements to consider when assessing data quality.

What Is Assessment?

Assessment is one of the 12 standards of public health identified in the *Standards for Public Health in Washington State* that the state and every local health department should be able to address.
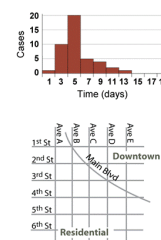
You might ask, What is assessment? Assessment involves collecting, analyzing, and using data to educate communities, develop priorities, acquire resources, develop policies, and evaluate programs.

Assessment is all about data. Health data includes diseases, conditions, health outcomes, and risk behaviors, as well as data on person, place, and time. These data help local health departments and the community decide how, where, and when health threats are occurring and where we should target our efforts.

Uses of Data

Let's cover the uses of data in a little more detail.

Health data are used to educate and inform others of the health needs of the community and to identify target populations. For example, in Snohomish Health District, we developed a set of public health indicators that included a wide variety of measures on diseases, conditions, health outcomes, and behaviors. These indicators inform and educate our own health jurisdiction and the community about what

Module 1: An Overview of Public Health Data

Transcript
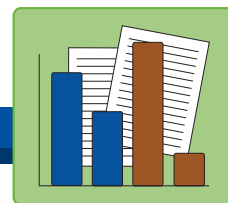
our needs are. Obesity, for example, is a relatively new health threat that affects one in four of our adult population.

The data also help health departments and organizations develop priorities so they can make wise decisions in program planning, setting goals, and prioritizing resources. Further analysis of the obesity data by person, place, and time, for example, can help us decide in which community we should target our interventions and what behaviors to target.
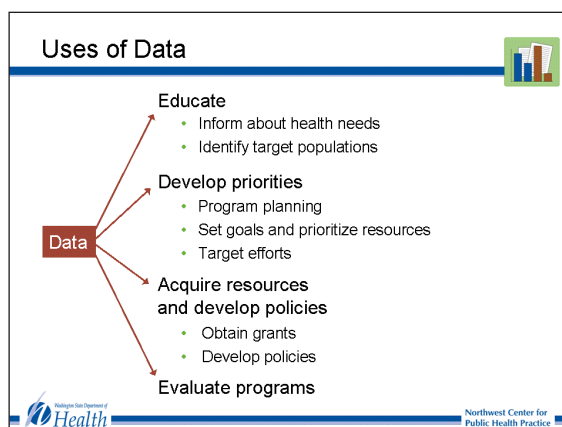
Data are an essential component of acquiring resources and developing policies. For example, we might use data on obesity and related conditions to help us obtain grants to carry out interventions or to develop policies, such as restaurant menu labeling or building more sidewalks.

Another essential part of assessment is to evaluate programs to see if they are effective. Health departments need to regularly evaluate their programs. Also, some grants—tobacco programs are a good example—require that the projects they fund be evaluated in order to show that the money is being spent wisely and actually results in change or improvement.

Uses of Data

Data →

Educate
- Inform about health needs
- Identify target populations

Develop priorities
- Program planning
- Set goals and prioritize resources
- Target efforts

Acquire resources and develop policies
- Obtain grants
- Develop policies

Evaluate programs

Overview of Public Health Data

Before you begin any assessment, you should have a basic understanding of the data sources available to you. In this module, we'll look at data attributes, types of data, data sources, data quality and relevance, and confidentiality. (Please note: Module 5 will look in more detail at some specific data sets available in Washington State.)

Data Attributes

When selecting a data source to use in your analysis, you will want to consider these attributes:

Whether the data are population-based or client/patient data, whether the data are from a population sample, whether the data represent a single person or an event, whether the data are from individual records or are aggregated, and finally whether the data are quantitative or qualitative.

Overview of Public Health Data

Understand available data sources

- Data attributes
- Types of data
- Data sources
- Data quality and relevance
- Confidentiality
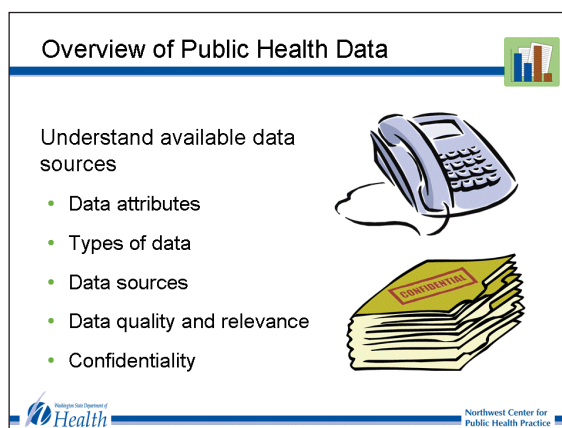
Module 1: An Overview of Public Health Data
Transcript
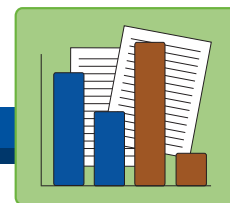
Let's look at these attributes in a little more detail. The first attribute is who is included in the data-set—the entire population or clients from a clinic or agency?

If everyone in the community—or state—could potentially be represented in the data (such as a vital statistics registry, which includes births and deaths), the data are population-based. However, if the data come from the records of a health or social service provider, it will include only those clients or patients served by that agency.

Some data sources use population sampling techniques to collect the information. Unlike a census or disease registry where everyone in the population—or everyone experiencing a specific event—is included, data from a population sample include only a specific proportion of the population. If this proportion or sample is drawn using a random method, and there is a high response rate, it should be representative of the whole.

Data Attributes: Record Definition

A second attribute is how a record is defined.

Most data sets are a compilation of individual records, each of which represent a single person. For example, each record in a data set derived from death certificates represents a person who died. However, some data sets may also be event-based, such as the motor vehicle accident database maintained by the state's Department of Transportation. In this case, each record represents an accident event that may have resulted in injuries to multiple people.

Hospitalization databases provide several different types of records; each record represents one event or episode, but one person may have multiple episodes or hospitalizations. These episodes may be for the same condition (such as asthma) or for different conditions.

Data Attributes

When selecting a data source, consider these attributes:

- Population-based vs. client/patient data
  - Everyone in the community = population-based data
  - Records of a health provider = client/patient data
- Population sample data
  - Include only a specific proportion of the population that is representative of the whole
- Person vs. event
- Record level vs. aggregated
- Quantitative vs. qualitative

Data Attributes: Record Definition

- Individual record represents a single person
- Event-based records can represent multiple people
- Hospitalization records represent one event
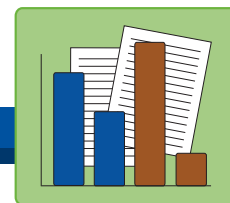 - But one person may have multiple episodes

Individuals   Events

Module 1: An Overview of Public Health Data

Transcript

Data Attributes: Raw vs. Aggregated

Another data attribute to be aware of is whether you have raw data or data that are already aggregated or summarized in some way.

In preparing for your analysis, you may obtain data in its "raw," or record-level, form. If you get raw data, you will need to summarize these individual records. You may also have to do some data management such as reorganizing or cleaning up the data file.

Frequently, though, it is easier to obtain data that have already been aggregated or summarized for you. Although this saves you a great deal of time and effort, the data may not have been summarized in a way that answers your specific question. For example, if you're working on a teen pregnancy report, you may find a published data table that gives you the rate of pregnancy among 15-to 19-year-olds but what you really need is the rate among 15-to 17-year-olds. With the record-level, or raw, data, you can calculate exactly what you need yourself.

Data Attributes: Raw vs. Aggregated

- Raw, or record-level, data
  - May need to reorganize or clean up data file
  - Can calculate exactly what you need
- Aggregated data is easier to obtain
  - Saves time and effort
  - May not be summarized in useful way

Data Attributes: Quantitative vs. Qualitative

A final attribute to note is are the data quantitative or qualitative?

When we talk about data, most people think of information that can be summarized numerically, for example, the number of people who died from heart disease or the median household income. Quantitative data will be the focus of the modules in the series Basic Concepts of Data Analysis.

However, data can also be qualitative or narrative in nature, such as the data collected in a focus group on attitudes toward cancer in an immigrant community. This qualitative data requires special analytic techniques that will not be covered in this series. But, qualitative data can provide a rich source of information that is not captured by quantitative data. It is sometimes useful to use a mix of qualitative data and quantitative data to provide a more complete picture. For example, a county health department finds through analysis of health risk behavior survey data that its maternal smoking rate

Data Attributes: Quantitative vs. Qualitative

- Quantitative can be summarized numerically
  - Example: People who died from heart disease
  - Focus of this series
- Qualitative is narrative
  - Data collected in focus group
  - Rich source of information
- Sometimes useful to combine qualitative and quantitative data
  - Provides more complete picture

Module 1: An Overview of Public Health Data
Transcript
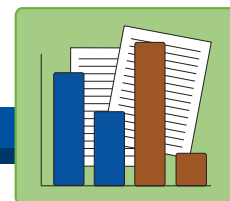
is higher in specific groups. In order to develop effective intervention strategies, it conducts focus groups with the higher-risk populations about their attitudes toward tobacco use.

Let's pause now so you can check your understanding of data attributes.

Practice Exercise: Data Attributes

Types of Data

Several types of data are commonly used in public health assessment. These include health outcome data, risk factors, demographic data, and resource data.

Health outcome data cover a variety of health events experienced by people, such as birth, death, infectious disease, chronic disease, and injuries.

Risk factors are behaviors or circumstances that put people at risk for experiencing a health event. Examples are smoking, lack of access to health care screening, and poor eating habits.

Demographic data not only give us some descriptive information about our population—such as where they live and how old they are—but they also frequently provide information about socio-economic characteristics that have an effect on health.

And, finally, resource data provide information about resources available to the population, such as health and social services, community infrastructure, and economic opportunities.

Types of Data
- Health outcome data
  - Health events such as birth, death, chronic disease, injuries
- Risk factor data
  - Behaviors or circumstances
  - Descriptive information about population, socio-economic characteristics
- Demographic data
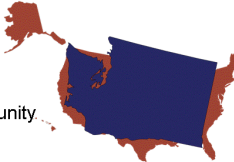 - Resources available to population
- Resource data

Data Sources: Routinely Collected

Data are available from a variety of sources. Every source has its advantages and disadvantages, which I'll discuss in more detail later. Some data sources are more reliable and appropriate than others, depending on your purpose or what questions you're trying to answer.

Data may be collected by national, state, regional, or local entities. Although data collected on a larger geographic scale—such as for the entire

Data Sources: Routinely Collected
- Every source has advantages and disadvantages
- Some sources are more reliable and appropriate
- National, state, regional, local
- Vital statistics registries
- Other registries (cancer, trauma, birth defects, pesticide exposure)
- Notifiable conditions data
- Health-related surveys (BRFSS, HYS, PRAMS)
- U.S. Census, American Community Survey
- Population estimates

Module 1: An Overview of Public Health Data
Transcript
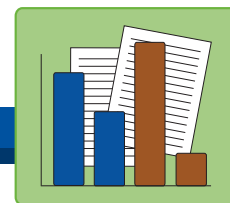
U.S.—has the advantage of broad comparability, it may not provide the level of detail needed by local public health jurisdictions.

Some of the most reliable and useful sources of data are the state vital statistics registries. These registries collect data on significant health events, such as birth and death, from the entire state population. Many state Centers for Health Statistics also collect statewide data on hospitalizations.

Other registries vary from state to state, but they provide valuable information about the occurrence of a significant disease or health event in the state's population. Types of state registries include cancer, trauma or injury, birth defects, and pesticide exposure.

For the purposes of surveillance, states typically require mandatory reporting of notifiable conditions by health care providers, laboratories, and veterinarians of new cases of certain communicable, sexually transmitted, and zoonotic diseases.

For health-related behaviors and risk factors, the only source of data may be population surveys, such as the Behavioral Risk Factor Surveillance System (or BRFSS as it's called), the Healthy Youth Survey (or HYS), and the Pregnancy Risk Assessment Monitoring System (which is frequently referred to as PRAMS).

The U.S. Census Bureau has collected demographic and socio-economic data on the U.S. population every ten years since 1790. In 2001, the Census Bureau began collecting similar data through an annual telephone survey. This survey is known as the American Community Survey.

States also produce official population estimates between the census years for a variety of planning and allocation purposes. These estimates and forecasts are invaluable for public health assessment because they provide a picture of the entire population in which health events occur and allow us to calculate rates for each year.
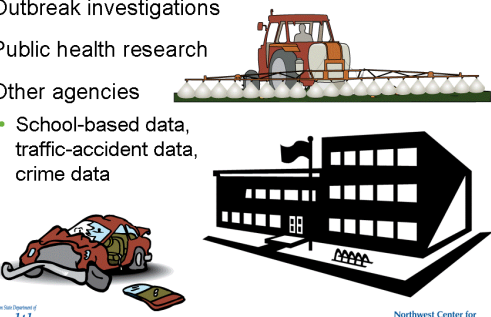
Other Data Sources

You may also be able to access other sources of data. For example, when outbreaks of rare or serious infectious disease occur, the local or state public health agency will often conduct a special investigation to collect additional information beyond what's captured in routine surveillance.

Data from public health agencies is generally more accessible than data collected by research and academic facilities, but research facilities that
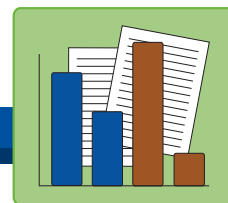

Other Data Sources

- Outbreak investigations
- Public health research
- Other agencies
  - School-based data, traffic-accident data, crime data

Module 1: An Overview of Public Health Data

Transcript

conduct primary data collection may provide more in-depth information about special health problems, such as pesticide exposure among farm workers.

Other public agencies that are not specifically health-related also collect data on topics and events that have public health implications. Examples include school-based data on student enrollment and achievement, traffic accidents and injury data from the department of transportation, and crime data from law enforcement agencies.

Let's pause now so you can check your understanding of the material we have just covered.

Practice Exercise: Data Sources

Data Quality: Accurate, Complete, Representative

Depending on the source and collection methods, data will vary significantly in terms of quality. As a data user you will have to evaluate the quality of the data. A first issue is, how accurate are the data? Consider these questions about data you want to use. Do the data come from a source that is reliable and has a proven track record, or do the data results conflict with other sources?

Another concern is, are the data complete? Do the data include all cases or events? For example, reporting for certain communicable diseases will be more complete than for others because some diseases require more rigorous follow-up or are more likely to result in contact with a health care provider.

You may also ask, are certain data elements missing for a large portion of the records? In many data sets, for example, race and ethnicity are not collected for all records.

And finally, are the data representative? Depending on how the data are gathered, a data set may not include all of the population of interest. Telephone surveys, including the Behavioral Risk Factor Surveillance System (or BRFSS), often face this challenge. The BRFSS survey is conducted only by landline telephone and doesn't include people who use only cell phones. Since people who use only cell phones tend to be younger, exclusion of this group from BRFSS may bias the data toward the health behaviors reported by an older population.

Data Quality: Accurate, Complete, Representative

Depending on the source and collection methods, data will vary significantly in terms of quality.

- Accurate
  - Do data come from a reliable source?
  - Do data conflict with other sources?
- Complete
  - Do data include all cases or events?
  - Are data elements missing?
- Representative
  - Do the data include all of the population of interest?
  - Are the data biased toward a population?

Module 1: An Overview of Public Health Data
Transcript
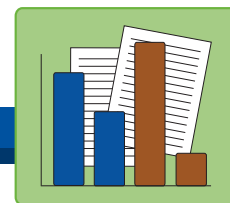
Data Quality: Frequency of Collection

Another question is: How frequently are the data collected? Because we are interested in how health events change over time, it is important to have data that are regularly collected and to have the most current data available. Data sets vary on when and how often the data become available. Even the most recent data sets may still be at least a year old because of the delay in compiling and reporting the data. The death data published in March 2008, for example, actually reports data for the calendar year 2006, not the deaths that happened in 2007.

Data Quality: Frequency of Collection

- How frequently are the data collected?
  - Collected regularly
  - Most current available

Data Quality: Consistent and Relevant

Are the data collection methods consistent over time? As circumstances change, the way certain elements are defined may also change. For example, race and ethnicity, collected by the U.S. Census Bureau, are concepts that have changed dramatically over the decades. As a result, data from the most recent census in 2000 is not directly comparable to that collected just ten years previous. Another area to be watchful for regarding consistency is how questions are asked in population surveys. Changes in the wording of the question make it a challenge to compare results over time.

Data Quality: Consistent and Relevant

- Are the data collection methods consistent over time?
  - Definitions may change
  - Wording of survey questions may change
- Is the data relevant to your question?

And finally, is the data relevant to the question you are trying to answer? The choice of the data you use must be determined by the question or purpose your assessment project seeks to address. If you want to know how many women are receiving mammograms, for example, data on breast cancer death rates won't answer that question.

Let's pause now so you can check your understanding of the material we have just covered.

Practice Exercise: Data Quality

Module 1: An Overview of Public Health Data

Transcript
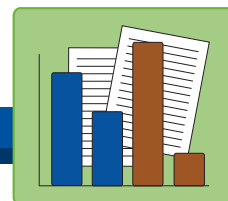
Possible Data Challenges

No matter what public health setting you work in, you will encounter some common challenges in dealing with public health data.

One common challenge you may face is that although many sources of reliable data exist, you may sometimes be confronted with a situation in which no data exists to answer your question.

Another challenge that smaller jurisdictions, in particular, may face is that when dealing with events that are relatively rare, the small numbers can be difficult to interpret. Module two discusses this challenge in more detail.

Data Confidentiality

An important consideration when working with any data set is protecting the confidentiality of the individuals represented in that data. Most agencies have protocols in place to protect the confidentiality of their data. These protocols specify how users are to handle raw data sets, the purposes for which their data can be used, and when analyses based on a small number of records should not be reported.

Accessing Data

When possible, the best way to access data is through the agency or program that collects and maintains it. This provides direct access to the most accurate and reliable version of the data. In addition, when you have questions about data quality or methodology you can go directly to the source for answers.

It is getting easier to do analysis as more and more public agencies develop online query systems for accessing their data. Depending on your assessment project, you may need to obtain raw case-

Possible Data Challenges

- No data exist to answer your question
- Small numbers can be difficult to interpret
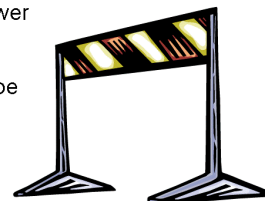
Data Confidentiality

- Must be aware of protecting the confidentiality of the individuals represented in that data
- Agencies have protocols to protect the confidentiality of their data
  - Specify how users are to handle raw data sets
  - Purposes for data use
 - Limitations on reporting small numbers

Accessing Data

- Best to access from program that collects and maintains the data
  - Can ask questions about data quality and methodology
- Download or query public data on Web sites
  - Obtain confidential raw data by signing data use agreement
- Secondary sources provide aggregated data from multiple sources

Module 1: An Overview of Public Health Data

Transcript
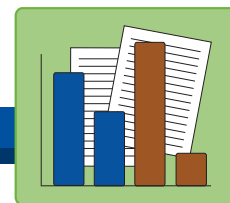
level data to perform your own analyses. The agency may ask you to sign a data use (or sharing) agreement that specifies protocols for maintaining the confidentiality of individuals represented in the data.

Data are also sometimes available from secondary sources that provide aggregated data from multiple sources in an online data warehouse.

Summary

To summarize, in this module we've talked about a variety of aspects of data.

Assessment is one of the core public health functions. Data are used in assessment, for example, to identify the size of health issues, develop policies and programs, obtain grants, and evaluate programs.

Data attributes to consider before analyzing data include: are the data population-based or client- or patient-based, is each record an individual or an event, and are the data raw or aggregated.

Numerous types of data are available: health outcome data, demographics, risk factors, and resource data.

Depending on what type of data you are interested in and what question you are trying to answer, many sources of data are available. More common types are regularly collected at the state or national level, such as vital statistics registries. Some state registries include cancer or hospitalizations. Other rich sources of data include behavioral risk factor survey data such as the Behavioral Risk Factor Surveillance System (or BRFSS) and the Healthy Youth Survey (or HYS).

In choosing the data you use, you need to be aware of its quality and its relevance to your objective.

Understanding protocols for confidentiality is critical before using any data.

And finally, accessing data is best done by working directly with the agency or program that maintains it.

Summary

To summarize, in this module we've talked about a variety of aspects of data.

- Data are used to identify size of health issues, develop policies and programs, obtain grants, and evaluate programs.
- Data attributes to consider before analyzing data include: Are the data population-based or client/patient? Is each record an individual or an event? Are the data raw or aggregated?
- Types of data: health outcome data, demographics, risk factors, and resource data.
- Sources of data: vital statistics registries, behavioral risk factor survey data.
- Be aware of issues of data quality and relevance.
- Understand protocols for confidentiality.
- Access data through the agency or program that maintains it.

Washington State Department of Health

Northwest Center for Public Health Practice

Module 1: An Overview of Public Health Data
Transcript
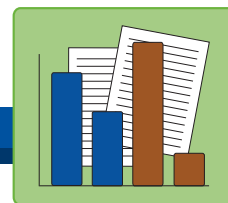
Resources

Related online modules from the Northwest Center for Public Health Practice

What Is Epidemiology in Public Health?
www.nwcphp.org/what-is-epi
Data Interpretation for Public Health Professionals
www.nwcphp.org/data

Other Online Courses

Data Analysis, or "What Do You Mean by That?"
www.pathwayscourses.samhsa.gov/eval102/
eval102_2_pg2.htm
Wading Through the Data Swamp www.pathwayscourses.samhsa.gov/eval201/
eval201_intro_pg1.htm

Resources

Related online modules from the Northwest Center for Public Health Practice
- What Is Epidemiology in Public Health?
- Data Interpretation for Public Health Professionals

Other Online Courses
- Data Analysis, or "What Do You Mean by That?
- Wading Through the Data Swamp