Analyze and Interpret Surveillance Data

Presenter's Name
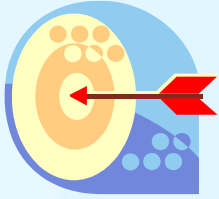
Presenter's Title

Title of Event

Date of Event

Learning Objectives

At the end of this training, you will be able to:

- Describe data to collect based on the objective of a surveillance system

- Identify how to present surveillance data

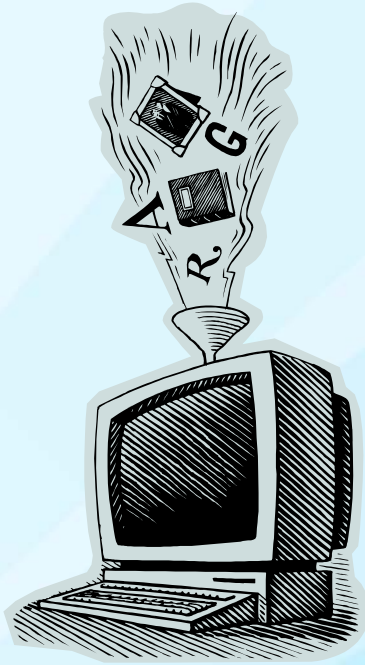- Interpret surveillance data, including trends and patterns

Lesson Overview

Review of data management

- Categories of data
- Confidentiality
- Data quality

Analyze and interpret data

- Presentation
- Descriptive epidemiology
- Risk factors
- Limitations when interpreting data

REVIEW OF DATA MANAGEMENT

Categories of Data

- Identifying
- Demographic
- Clinical
- Laboratory
- Risk factor
- Source

Question: What kinds of information would you find in each of these categories of data?

Confidentiality

- All data must be kept confidential.

- Some surveillance systems do not enter identifying information to protect confidentiality.

Maintaining Confidentiality

Question: Why is it important to keep surveillance data confidential?
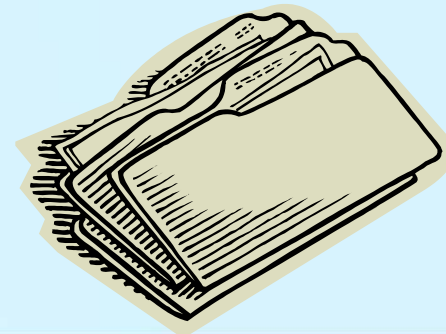
Ways to protect confidentiality:

- Assign a unique ID number to each case or record

- Avoid unintentional disclosure

Data Quality: Completeness

Completeness can refer to many things:

- Completeness of data collected
  - How much of the data are missing?

- Completeness of reporting
  - Did the surveillance system capture all of the events?

Question: Completeness

Table 1. Diabetes

Sex	n	%
Male	491	49.1
Female	423	42.3
Missing	86	8.6
Age	**n**	**%**
Median	38 years	n/a
Missing	173	17.3
Total	1000	100

Table 2. Asthma

Sex	n	%
Male	921	46.0
Female	874	43.7
Missing	205	10.3
Age	**n**	**%**
Median	35 years	n/a
Missing	129	6.4
Total	2000	100

Data Quality: Validity

- Validity refers to the accuracy of the collected data.

- Sources of errors:
  - Respondent provides inaccurate information
  - Data recorded inaccurately when collected
  - Data entered inaccurately into database

Question: Validity

| Name	Sex	Date of Birth	Date of Diagnosis	Date of Death	Age at Death
Person A	Female	15/07/1954	16/03/2006	03/12/2008	47
Person B	Male	28/04/1963	24/09/2004	27/07/2009	46
Person C	Both	02/09/1948	30/02/2005	18/10/2008	60
Person D	Male	08/11/1943	05/06/2009	11/04/2009	65
Person E	Male	21/01/1792	19/10/2006	09/09/2010	38

ANALYZING SURVEILLANCE DATA

Analysis in Support of Objectives

- Analysis of the data needs clear, measurable objectives.

- Surveillance objectives guide:
  - Types of data collected
  - Types of analyses performed

Presentation of Results from Analyses

- Results should be presented in a manner that is easy to understand and interpret.

- Formats:
  - Tables
  - Graphs
  - Charts
  - Maps
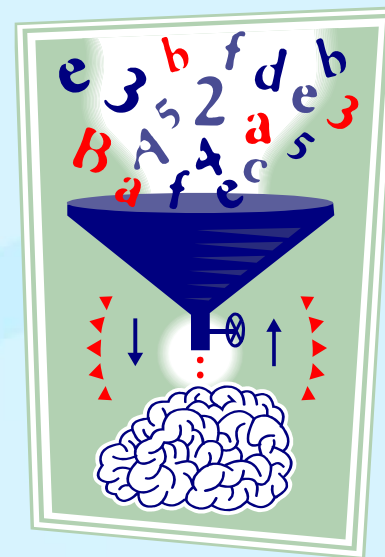
Descriptive Epidemiology

- Data used to describe the distribution of a health condition or event in a community.
  - Person – who?
  - Place – where?
  - Time – when?

Person

Who?

- Age
- Sex
- Marital status
- Occupation

Question: Person

Prevalence of obesity (body mass index ≥ 30 kg/m^2) among men aged 18-59 years old, Brazil, 2006

Age (years)	Prevalence (%)	95% Confidence Interval
18-25	4.3	3.1, 5.5
26-32	10.9	8.3, 13.6
33-40	14.9	12.7, 17.1
41-48	14.1	12.0, 16.3
49-59	16.5	13.9, 19.2

Source: Moura EC, Claro RM. Estimates of obesity trends in Brazil, 2006-2009. Int J Public Health. 2011 May 26.
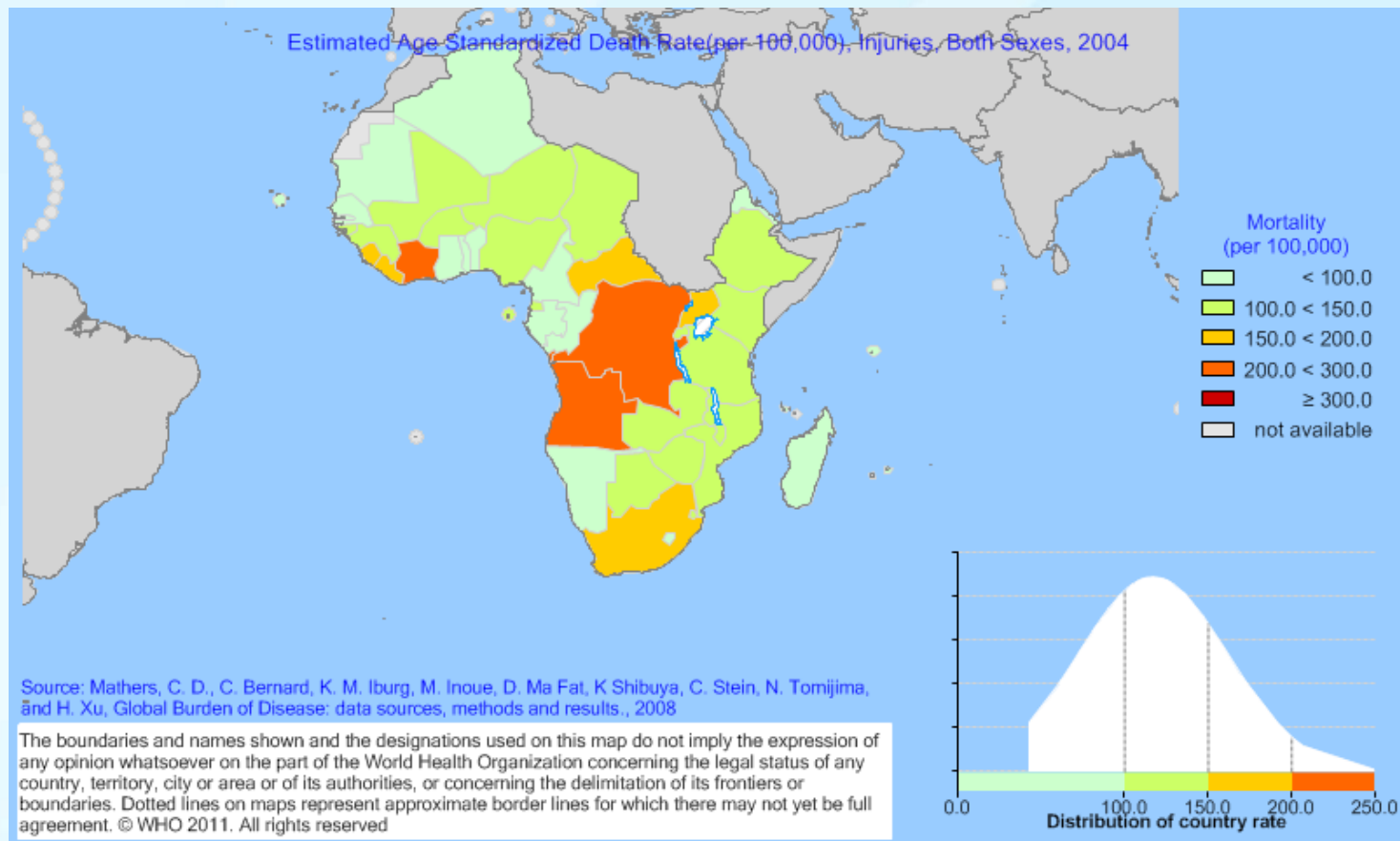
Place

- Where?
  - Residence
  - Workplace
  - Location of exposure
  - Location of diagnosis and/or treatment

- Maps can describe health conditions by place.

Question: Place

Estimated Age Standardized Death Rate (per 100,000), Injuries, Both Sexes, 2004

Source: WHO Global Infobase. https://apps.who.int/infobase/Comparisons.aspx
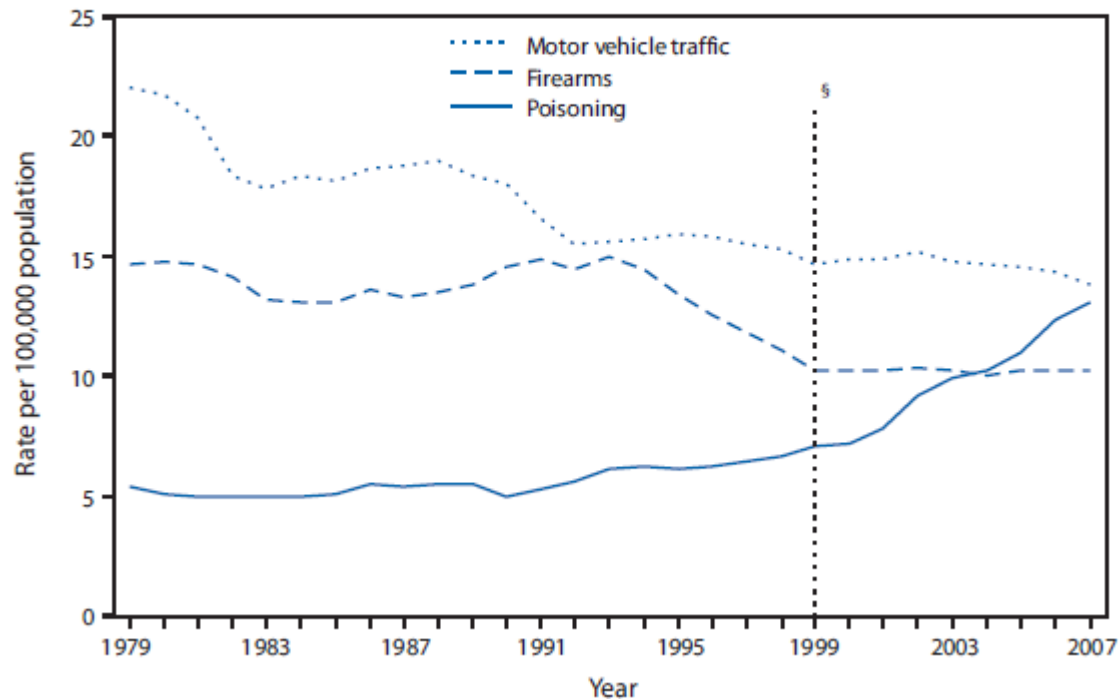
Time

- When?

- Look for trends over time

- Examples: year, day of week, season

Question: Time

Death Rates* for the Three Leading Causes of Injury Death[†] — United States, 1979–2007

Risk Factors

- Risk factors commonly associated with NCDs
  - Alcohol consumption
  - Diet and nutrition
  - Genetics
  - Lack of physical activity
  - Tobacco use

Types of Risk Factors

- Modifiable – can be changed by the individual
  - Lifestyle choices

- Nonmodifiable – cannot be changed by the individual
  - Age
  - Family history (hereditary)
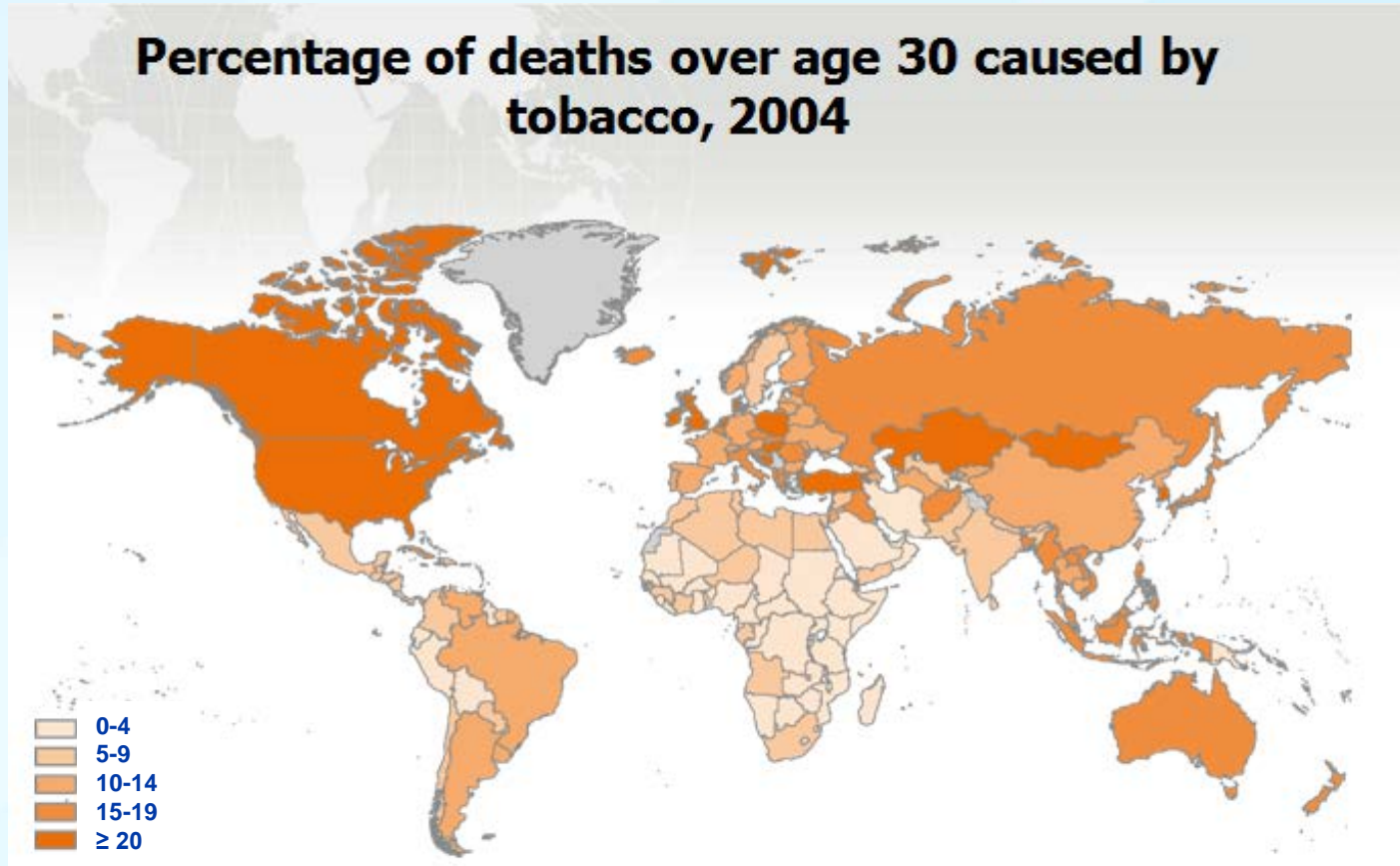 - Race or ethnicity
  - Sex

Question 1: Risk Factors

Adapted from Chronic Disease Risk Factors Among Participants in Medical Examination, by Selected Demographic Characteristics

High Blood Pressure	Age Groups				
	18-34 % (SE)	35-49 % (SE)	50-64 % (SE)	≥ 65 % (SE)	Total
Self-reported	2.5 (0.095)	11.3 (1.87)	35.9 (4.05)	34.1 (6.82)	15.2 (1.52)
Measured	9.4 (2.30)	28.3 (3.53)	55.2 (3.78)	61.4 (5.52)	30.2 (1.83)

Question 2: Risk Factors

Percentage of deaths over age 30 caused by tobacco, 2004

Legend:
- 0-4
- 5-9
- 10-14
- 15-19
- ≥ 20

Source: WHO Global Health Risks. Accessed on January 13, 2011 at
http://www.who.int/healthinfo/global_burden_disease/global_health_risks/en/index.html

Interpreting Modifiable Risk Factor Data

- Individuals choose to engage in modifiable factors.

- Recommendations often encourage people to change behavior and make better health-related decisions.

Interpreting Non-modifiable Risk Factor Data Tips

- Use caution and sensitivity

  - People do not choose to have non-modifiable risk factors.

  - Children, women, and people from certain races or ethnic groups may be vulnerable populations.

  - Vulnerable populations may be stigmatized if data are not interpreted appropriately.
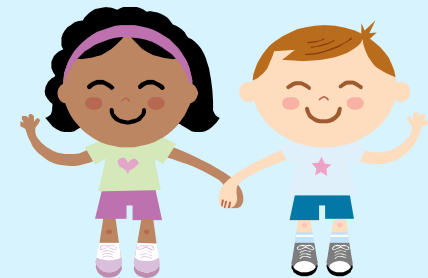
Interpreting Non-modifiable Risk Factor Data

- Be aware of non-modifiable risk factors.

- Cannot recommend changes.

- Can recommend reducing modifiable risk factors among people with non-modifiable factors.

  - Example: Risk of obesity increases with age:

    – Cannot modify age

    – Can increase physical activity or improve diet among older people
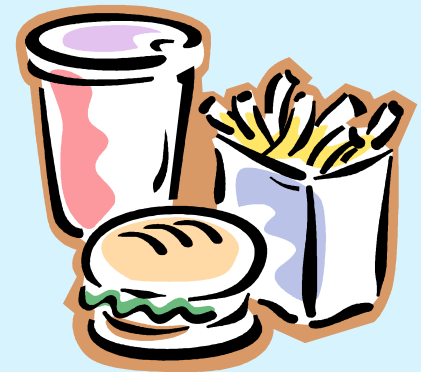
Limitations of Surveillance Data

- Underreporting of cases

- Not representative of entire population

- Changes in case definition over time

Underreporting

- Failure to report a health condition or vital event, as required by law, to proper officials
  - Due to individuals being unaware of their responsibility to report
  - Common in passive surveillance systems
    - Notifiable disease reporting systems
    - Vital events registration
   - Morbidity registries

Example of Underreporting

Can you share an example of underreporting that you may have experienced in your work?

Representativeness

- Definition: How accurately data reflect the occurrence and distribution of a disease in a population

- Affected by
  - Exclusion of particular subpopulations
  - Changes in reporting practices
  - Differences in reporting practices

Representativeness Example
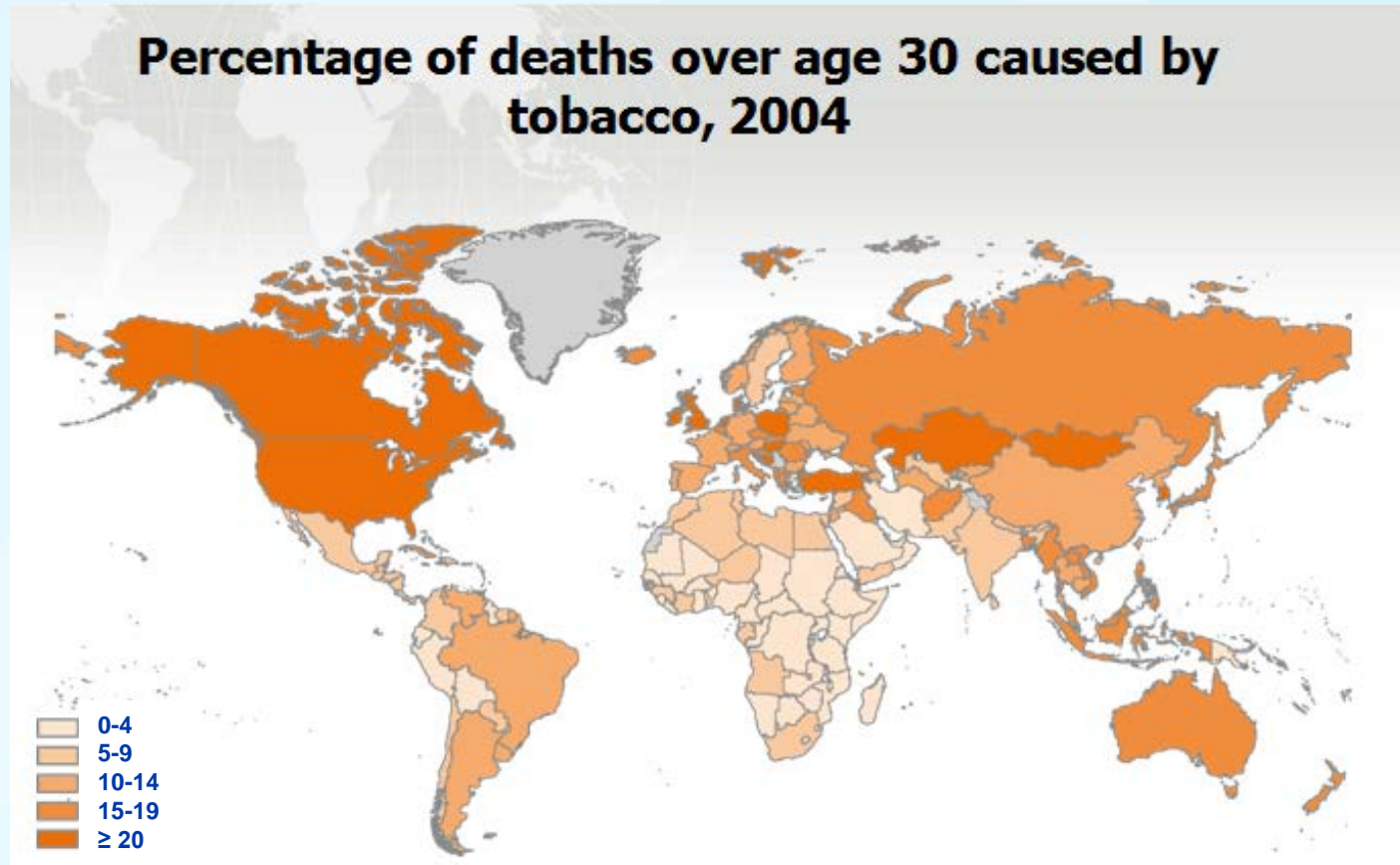
Percentage of deaths over age 30 caused by tobacco, 2004

Legend:
- 0-4
- 5-9
- 10-14
- 15-19
- ≥ 20

Source: WHO Global Health Risks. Accessed on January 13, 2011 at
http://www.who.int/healthinfo/global_burden_disease/global_health_risks/en/index.html

Inconsistent Case Definitions

- Case definitions
  - Standard criteria
  - Used by public health officials to classify a health condition

- Data limitations
  - When definitions are not used consistently
  - When looking at trends, a revised definition can lead to dramatic yet misleading changes
  - Need to consider when interpreting data

Inconsistent Case Definitions Example

Death Rates* for the Three Leading Causes of Injury Death† — United States, 1979–2007
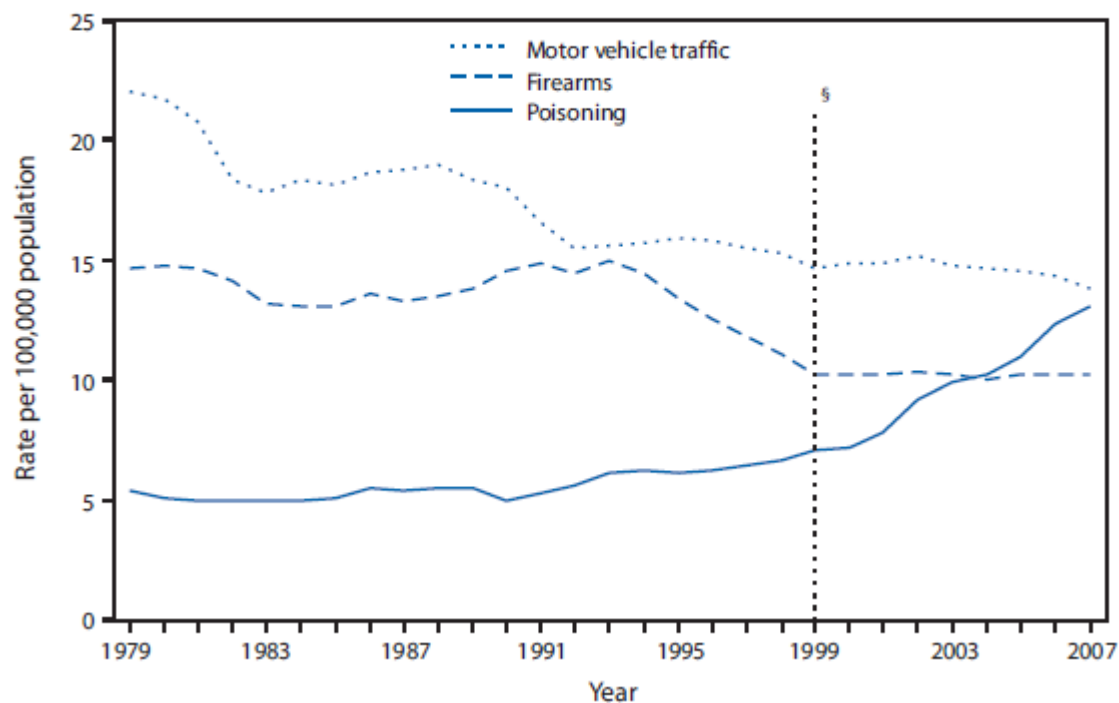
Source: CDC. QuickStats: Death Rates for the Three Leading Causes of Injury Death - United States, 1979-2007. MMWR Weekly Report. Aug 6, 2010;59(30):957.

Ensuring Consistent Case Definitions

- International Classification of Diseases (ICD) codes
  - International set of criteria used to classify health conditions and deaths
  - Used for clinical, epidemiological, and managerial purposes
  - Permits comparison of calculated morbidity and mortality between countries
  - Periodic revisions to incorporate new health conditions as well as advances in science and technology

REVIEW

Review: Questions 1-3

1. Name at least three categories of data collected as part of surveillance.

2. What is unintentional disclosure?

3. When we talk about data quality, _____ refers to how much data are missing and _____refers to the accuracy of the data.

Review: Answers 1-3

1. Name at least three categories of data collected as part of surveillance. *Identifying, demographic, clinical, laboratory, and risk factor*

2. What is unintentional disclosure? *When nonidentifying information is released that is specific enough that the identity of an individual can be inferred.*

3. When we talk about data quality, _____ refers to how much data are missing and _____ refers to the accuracy of the data. *1) Completeness, 2) Validity*

Review: Questions 4-6

4. Data should be described by which three epidemiologic attributes?

5. What is a modifiable risk factor? Give an example.

6. Name two of three limitations to be aware of when interpreting surveillance data.

Review: Answers 4-6

4. Data should be described by which three epidemiologic attributes? *Person, place, and time*

5. What is a modifiable risk factor? Give an example. *A modifiable risk factor is a behavior that an individual chooses to engage in or not to engage in. Smoking and little physical activity are examples.*

6. Name two of three limitations to be aware of when interpreting surveillance data. *Underreporting, representativeness, changes in case definition*

Review: Questions 7-8

7. What is a case definition and how is it used in public health surveillance?

8. What are two risk behaviors that you could change to decrease your risk of chronic disease?

Review: Answers 7-8

7. What is a case definition and how is it used in public health surveillance? *Case definitions standardize the data by providing uniform criteria to be used throughout the surveillance system – ensuring that all those who report use the same case definition.*

8. What are two risk behaviors that you could change to decrease your risk of chronic disease? *Diet, exercise, smoking, alcohol, using safety equipment, etc.*

Skills Assessment

1. You will work in a small group to complete the assessment.

2. Use the tables, graphs, and charts to analyze and interpret obesity and cardiovascular diseases in the United States. Materials and questions are in your *Participant Guide*.

3. Spend approximately 1 hour completing the assessment.

Centers for Disease Control and Prevention (CDC). Analyze and Interpret Surveillance Data. Atlanta, Georgia: Centers for Disease Control and Prevention (CDC); 2013.

For more information please contact Centers for Disease Control and Prevention
1600 Clifton Road NE, Atlanta, GA 30333
Telephone: 1-800-CDC-INFO (232-4636)/TTY: 1-888-232-6348
Visit: www.cdc.gov | Contact CDC at: 1-800-CDC-INFO or www.cdc.gov/info

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Department of Health and Human Services

Centers for Disease Control and Prevention