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

<table>
<thead>
<tr>
<th>Sex</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>491</td>
<td>49.1</td>
</tr>
<tr>
<td>Female</td>
<td>423</td>
<td>42.3</td>
</tr>
<tr>
<td>Missing</td>
<td>86</td>
<td>8.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>38 years</td>
<td>n/a</td>
</tr>
<tr>
<td>Missing</td>
<td>173</td>
<td>17.3</td>
</tr>
</tbody>
</table>

**Total** | 1000 | 100 |

#### Table 2. Asthma

<table>
<thead>
<tr>
<th>Sex</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>921</td>
<td>46.0</td>
</tr>
<tr>
<td>Female</td>
<td>874</td>
<td>43.7</td>
</tr>
<tr>
<td>Missing</td>
<td>205</td>
<td>10.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>35 years</td>
<td>n/a</td>
</tr>
<tr>
<td>Missing</td>
<td>129</td>
<td>6.4</td>
</tr>
</tbody>
</table>

**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

<table>
<thead>
<tr>
<th>Name</th>
<th>Sex</th>
<th>Date of Birth</th>
<th>Date of Diagnosis</th>
<th>Date of Death</th>
<th>Age at Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person A</td>
<td>Female</td>
<td>15/07/1954</td>
<td>16/03/2006</td>
<td>03/12/2008</td>
<td>47</td>
</tr>
<tr>
<td>Person C</td>
<td>Both</td>
<td>02/09/1948</td>
<td>30/02/2005</td>
<td>18/10/2008</td>
<td>60</td>
</tr>
<tr>
<td>Person D</td>
<td>Male</td>
<td>08/11/1943</td>
<td>05/06/2009</td>
<td>11/04/2009</td>
<td>65</td>
</tr>
<tr>
<td>Person E</td>
<td>Male</td>
<td>21/01/1792</td>
<td>19/10/2006</td>
<td>09/09/2010</td>
<td>38</td>
</tr>
</tbody>
</table>
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/m²) among men aged 18-59 years old, Brazil, 2006

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

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

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

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

Question 2: Risk Factors

Percentage of deaths over age 30 caused by tobacco, 2004

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

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


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.
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.
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?
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.