EpiData: The Relevance for Local/Regional Risk Factor Surveillance

CARRFS e-learning session

September 24, 2010
Outline

- Association of Public Health Epidemiologists in Ontario & the EpiData Association
- APHEO EpiData Evaluation & Redevelopment Project
- EpiData Entry and Analysis
- Examples: Uses in Surveillance
- The Future of EpiData
- Q & A
The Association of Public Health Epidemiologists in Ontario - organization of ~ 75 full members who practice epidemiology in Ontario's public health units, as well as more than 100 affiliate members.

Formed in 1991.

APHEO's Mission: To advance and promote the discipline and professional practice of epidemiology in Ontario public health units.

www.apheo.ca
EpiData Association

- Creator of EpiData software - 1999
- Lead by Jens Lauritsen, MD.
- www.epidata.dk
APHEO EpiData Project

- Funded by Public Health Agency of Canada since 2007
- **Purpose:** APHEO supports and influences the further development of EpiData to improve its utility in the field and as a training tool.
- **Project Goals:**
  - Improved EpiData functions that are well suited to the practice of public health in Canada
  - Production of test versions and documentation of EpiData software and Canadian case studies to illustrate the management and analysis of data in the field
  - The software, documentation and case studies will be used by public health practitioners and trainers in Canada to complement existing surveillance software, analysis software and training tools.
- **EpiData Project website:**
EpiData Software Overview

EpiData Entry
EpiData Analysis
History

- **1976–1995:** EpiInfo (DOS program) created by Centers for Disease Control and Prevention (CDC)
- **2000:** CDC releases EpiInfo2000
  - Based on Microsoft Jet (Access) data engine
  - Massive changes
- **2001:** Loyal EpiInfo user group decides it needs real “EpiInfo for Windows”
  - Creates open source public domain program
  - Calls program “EpiData”
EpiData

- Public domain software package for use by public health professionals and researchers
- Two components:
  - EpiData 3.1 for data entry
  - EpiData Analysis (v.2.1)
- Suitable for survey datasets
- Translated in several languages
- Downloaded from the EpiData website: www.epidata.dk
EpiData Entry

- Data entry based on a questionnaire
- Data documentation
- Protect the database from input errors by writing a ‘check’ program
- Import and export data
- Error detection features. E.g. double entry verification
- Support for relational databases
EpiData Entry in Steps

- Define data (QES file)
- Make data file (REC file)
- Checks (CHK file)
- Enter data
- Document
- Export data
Codebooks

- Contain info that helps users decipher data file content and structure
- Includes:
  - Filename(s)
  - File location(s)
  - Variable names
  - Codes and their associated meanings
EpiData Codebook Generators
**File Structure Codebook**

**DATA FILE:** C:\Data\datasets\demo.rec  
**File label:** Class demonstration  
**File size:** 395 bytes  
**Last revision:** 28. Apr 2005 1:48 PM  
**Number of fields:** 5  
**Number of records:** 2  
**Checks applied:** Yes (Last revision 28. Apr 2005 1:45 PM)

Fields in data file:

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Variable label</th>
<th>Field type</th>
<th>Width</th>
<th>Checks</th>
<th>Value labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>fname</td>
<td>Text</td>
<td>Text</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>lname</td>
<td>Upper-case text</td>
<td>Upper-case text</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>dob</td>
<td>Date (mdy)</td>
<td>Date (mdy)</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>sex</td>
<td>Number</td>
<td>Number</td>
<td>1</td>
<td></td>
<td>sex</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1: Male</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2: Female</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9: Unknown</td>
</tr>
<tr>
<td>5</td>
<td>deathage</td>
<td>Number</td>
<td>Number</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
EpiData Analysis

- Descriptive statistics (i.e., sum, mean, min, max, frequencies, etc.)
- Define/modify data
- Produce graphs & tables

  - Summary statistics, confidence intervals
  - Linear regression, correlations
  - Stratified analysis
  - Outbreak focused analysis (i.e., attack rate tables, epi curves)
  - Survival curves Life table and Kaplan-Meir plots
  - Comprehensive data management: recode variables, define missing values, label values and label variables.
Examples

Uses of EpiData in Public Health Surveillance
WHO - STEPwise approach to surveillance (STEPS)

- a simple, standardized method for collecting, analysing and disseminating data in WHO member countries.

- 2 STEPS surveillance systems:
  - Adult chronic disease risk factor
  - Stroke

- EpiData templates available that provide the generic code used during data entry that matches the STEPS Instrument.

- http://www.who.int/chp/steps/en/
Data Entry Assistance – Coding pop- screens

World Health Organization
WHO STEPwise approach to NCD risk factor surveillance
STEP5 Data Entry Instrument v2.0

: Participant Identification Number

LOCATION AND DATE INFORMATION

: District Code
: Cluster/Centre/Village Name
: Cluster/Centre/Village Code
: Interviewer Code
: Date Completed Instrument

STEP 1: Questionnaire

MOGRAPHIC INFORMATION

: Sex
: What is your date of birth (leave blank if 77/77/7777)
: How old are you?
: In total, how many years have you spent at school?
: What is your ethnic background?
: What is the highest level of education you have completed?
: Which of the following best describes your main work status?
: How many people older than 18 live in your household?

[Select value window]

- No formal schooling
- Less than primary school
- Primary school completed
- Secondary school completed
- High school completed
- College/University completed
- Post graduate degree
- don't know
- refused
- missing
## Data Quality Assurance – Check commands

### PHYSICAL MEASURES

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
<th>Check Commands</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>Interviewer ID</td>
<td></td>
</tr>
<tr>
<td>M2A</td>
<td>Device ID for height</td>
<td></td>
</tr>
<tr>
<td>M2B</td>
<td>Device ID for weight</td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>Height</td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>Weight</td>
<td></td>
</tr>
<tr>
<td>M5</td>
<td>Participant pregnant</td>
<td></td>
</tr>
<tr>
<td>M6</td>
<td>Device ID</td>
<td></td>
</tr>
<tr>
<td>M7</td>
<td>Waist circumference</td>
<td></td>
</tr>
<tr>
<td>M8</td>
<td>Interviewer ID</td>
<td></td>
</tr>
<tr>
<td>M9</td>
<td>Device ID</td>
<td></td>
</tr>
<tr>
<td>M10</td>
<td>Cuff Size</td>
<td></td>
</tr>
<tr>
<td>M11A</td>
<td>Systolic Reading 1</td>
<td></td>
</tr>
<tr>
<td>M11B</td>
<td>Diastolic Reading 1</td>
<td></td>
</tr>
<tr>
<td>M12A</td>
<td>Systolic Reading 2</td>
<td></td>
</tr>
<tr>
<td>M12B</td>
<td>Diastolic Reading 2</td>
<td></td>
</tr>
<tr>
<td>M13A</td>
<td>Systolic Reading 3</td>
<td></td>
</tr>
<tr>
<td>M13B</td>
<td>Diastolic Reading 3</td>
<td></td>
</tr>
<tr>
<td>M14</td>
<td>On treatment with drugs?</td>
<td></td>
</tr>
<tr>
<td>M15</td>
<td>Hip circumference</td>
<td></td>
</tr>
<tr>
<td>M16</td>
<td>Heart rate</td>
<td></td>
</tr>
</tbody>
</table>

**Check command example:**

```plaintext
M3 AFTER ENTRY
  IF (M3=.) THEN
    HELP 'A value must be entered. Please enter a value or use: 888.8 for refused 999.9 for missing.'
    GOTO M3
  EXIT
END?
```
Multi-Drug Resistant Tuberculosis

- **Study Objective:** observational study - evaluates effectiveness of standardized regimens for patients with proven multi drug-resistant tuberculosis previously untreated with second-line drugs in low income countries.

- Consenting patients sequentially assigned to one of six standardized treatment regimens.

- Reported the treatment outcome of all patients with laboratory-confirmed, multi drug-resistant tuberculosis enrolled from May 1997 to December 2007.
Treatment records captured in EpiData Entry

Analysis in EpiData Analysis:
- 95% confidence intervals (CI) for proportions and P values where appropriate
- Kaplan-Meier survival statistics with 95% CI

TB Case Registry – Assessing Data Quality

- **Study Objective:** To determine completeness and consistency of information for quarterly reports on case finding and treatment outcome.

- **EpiData Entry:** Data were double-entered and validated for any discordance by comparing the discordant values against the original physical case register and making corrections where appropriate.

- **EpiData Analysis:** frequencies, summaries

Source: Completeness and consistency in recording information in the tuberculosis case register, Cambodia, China and Viet Nam. INT J TUBERC LUNG DIS 14(10):1303–1309
Durham Region Health Unit – ‘DineSafe Durham’ Evaluation

- DineSafe Durham - new program to increase compliance with the Ontario Food Premises Regulations and increase transparency of, and public accessibility to, food safety inspection results and information.

- Colour-coded food safety inspection signs posted at the entrance to food establishments

- Evaluation Question: Have attitudes towards food safety practices changed one year post-implementation?
Method: telephone survey of owners/operators to obtain feedback on the potential impact of DineSafe Durham on business and food safety handling practices

EpiData Entry: real-time data entry by telephone interviewers
Q1. How has DineSafe Durham, the disclosure of the inspection results affected your business?
Would you say very positively (1), somewhat positively (2), has made no impact (3), somewhat negatively (4), or very negatively (5)?

1. Very positively
2. Somewhat positively
3. No impact
4. Somewhat negatively
5. Very negatively
6. Don't know
7. Refused

Q2. How much change, if any, has occurred in your food handling practices as a result of DineSafe Durham?
Would you say there has been significant change (1), some change (2), little change (3), or no change (4)?

1. Significant change
2. Some change
3. Little change
4. No change
5. Don't know
6. Refused

Q3. How much change, if any, has DineSafe Durham made to the way you meet the Ontario Food Premises Regulations?
Would you say there has been significant change (1), some change (2), little change (3), or no change (4)?

(ONLY READ IF AN EXPLANATION IS REQUIRED): The Ontario Food Premises Regulations are the primary pieces of legislation...
Easy to Load Data Files

```
. read
Loading data I:\DineSafe Durham\Evaluation 2009_2010\DineSafe Phase 1\Questionnaire - Oper
File name :I:\DineSafe Durham\Evaluation 2009_2010\DineSafe Phase 1\Questionnaire - Oper
DineSafe Durham Operatory Telephone Survey
Fields: 11 Total records: 53 Included: 53
```
Analyze using easy-to-create Programs
The Future of EpiData

- New: EpiData XML file format "EPX" will combine and replace the current qes, chk and rec files into one physical file
- More access controls
- The re-written software will consist of three parts
  a. EpiData Data Manager
  b. EpiData DataEntry Client
  c. EpiData Analysis
- Public tests of re-written software ongoing
EpiData Benefits

- Easy to use. Although its features may be fewer than more sophisticated packages, the simplicity is a benefit in many situations, particularly for beginners.

- Distributed free of charge.

- Doesn’t require a powerful computer to run it.

- Can export data in formats that can be read by virtually every statistical, database, or spreadsheet package.
Benefits cont’d

- Strong network of software developers – listens to users – enhances software to meet local needs.
- Very portable – small size, carried on USB sticks, install on as many computers as you like (remember – it’s a freeware!). Great for field work.
- Support from other EpiData users via the EpiData listserv.
Population health assessment and surveillance cycle

Data Access, Collection and Management

EpiData Entry

Population Health Assessment and Surveillance

EpiData Analysis

Data Analysis and Interpretation

EpiData Analysis

Report and Dissemination

Action

Source: Population Health Assessment and Surveillance Protocol, Ontario Ministry of Health and Long-Term Care, 2008
EpiData Resources

- EpiData Introduction Guide: A Canadian Example
  (http://www.apheo.ca/index.php?pid=47#Project Documents)

- E-mail based discussion list:
  http://lists.umanitoba.ca/pipermail/epidata-list/
APHEO EpiData Expert Panel

- Brenda Guarda - Project Manager brenmin@rogers.com
- Camille Achonu - Ontario Agency for Health Protection and Promotion cachonu@gmail.com
- Anne Arthur - Toronto Public Health aarthur@toronto.ca
- John Barbaro - Simcoe Muskoka District Health Unit john.barbaro@smdhu.org
- Jason Garay - York Region Community and Health Services Department jason.garay@york.ca
- Kathy Moran - Durham Region Health Department kathy.moran@durham.ca
References and Training Materials

- Data Management for Surveys and Trials. A Practical Primer using EpiData.

- Data Management Using EpiData.
  http://www.idready.org/webcast/fall06_materials/outbreak_investigation/2006-09-12/DE_EpiData_v2.ppt#256,1,Data Management using EpiData

  http://www.who.int/medicines/areas/coordination/training_slides_data_entry.pdf

- STEPS Data Entry Guide.
  http://www.who.int/chp/steps/Part3_Section5.pdf
References and Training Materials

  - Accompanying training documents: http://www.tbrieder.org/epidata/course.htm
- Using of EpiData (questionnaire design and entry). http://www.epinorth.org/dav/FA086CF6D9.ppt#256,1,Use of EpiData (questionnaire design and entry)
References and Training Materials

- EpiData website. [www.epidata.dk](http://www.epidata.dk)
- EpiData Software - Links and Courses. [http://www.epidata.dk/links.htm](http://www.epidata.dk/links.htm)
Thank You!