BCCDC Informatics Activities

BC Centre for Disease Control
An agency of the Provincial Health Services Authority

Environmental Health Surveillance Workshop

February 26, 2013
Public Health Informatics

• Application of key disciplines to Public Health
  – information science
  – computer science
  – cognitive or social sciences
  – supplemented as needed by other domains including mathematics, engineering, biostatistics, and GIS

• A multidisciplinary approach to solving complex informational, analytical, and technical problems, guided by scientific principles and applied systematically.
Data Warehouse
BCCDC Context

• 2007 MOU identifies data linkage as key BCCDC function
• 2008 BCCDC Surveillance Workshop identifies centralization of data as priority theme
  • Common data standards
  • Centralized data cleaning
  • Linkage of all data sets within BCCDC
Public Health Reporting Data Warehouse

- Automated and standardized:
  - Data integration
  - Patient-matching; case-matching
  - Geo-coding (health boundaries and derivation of address at time of event)
  - Episode date derivation
- Patient- and specimen-centric data views
- Data provided to end users anonymized or audited
Principles

- Analyses run on real-time data to support surveillance
- Data source cooperation
  - Joint access to marts
- Improved Privacy
  - De-identified data in marts; line lists audited
  - Copies of data are minimized
  - Analyses exported, but data is not
Advantages

• Time savings from automated data reconciliation
• Standardization and data quality
• On-going access to health outcomes data to evaluate PH programs
Data Sources

Public Health Reporting Data Warehouse (PHRDW)

- Panorama
  - CD
- iPHIS
- Enhanced Surveillance Databases (e.g. IGAS)
- Vital Statistics
- MOH Administrative datasets (e.g. MSP, PharmaNet, DAD)
- Other PHSA registries (e.g. Perinatal Health, Cancer)
- Legacy lab data
  - HIV
  - HCV
- Mysis
- PLIS
- HAISYS
- STIS

Green = Current
Yellow = 2012/13
Orange = Near term
Red = Long term
Data Marts

- Subset of data for a specific surveillance purpose
  - Automated algorithms applied (e.g. lab interpretation, case definitions)
  - Mart-specific dimensions created (e.g., age groups)
What is a Cube?

Data cube

• a three- (or higher) dimensional array of values  

OLAP cube

• a set of data organized in a way that facilitates non-predetermined queries for aggregated information  
• a tool for Online Analytical Processing  
  <http://en.wikipedia.org/wiki/OLAP_cube>
Transformation of source data for cube creation

Variables are grouped into dimensions

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age</th>
<th>Health Authority</th>
<th>HSDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>15</td>
<td>Vancouver Coastal</td>
<td>Vancouver</td>
</tr>
<tr>
<td>M</td>
<td>25</td>
<td>Fraser</td>
<td>Fraser North</td>
</tr>
<tr>
<td>M</td>
<td>19</td>
<td>Vancouver Coastal</td>
<td>Vancouver</td>
</tr>
<tr>
<td>F</td>
<td>22</td>
<td>Fraser</td>
<td>Fraser South</td>
</tr>
</tbody>
</table>

Values become their members (categories)
Data presented in a Cube format – Example 1

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age</th>
<th>Health Authority</th>
<th>HSDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>15</td>
<td>Vancouver Coastal</td>
<td>Vancouver</td>
</tr>
<tr>
<td>M</td>
<td>25</td>
<td>Fraser</td>
<td>Fraser North</td>
</tr>
<tr>
<td>M</td>
<td>19</td>
<td>Vancouver Coastal</td>
<td>Vancouver</td>
</tr>
<tr>
<td>F</td>
<td>22</td>
<td>Fraser</td>
<td>Fraser South</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>15-19</th>
<th>20-24</th>
<th>25-29</th>
<th>Age Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>M</td>
<td>1</td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Gender</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>
Data presented in a Cube format – Example 2

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age</th>
<th>Health Authority</th>
<th>HSDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>15</td>
<td>Vancouver Coastal</td>
<td>Vancouver</td>
</tr>
<tr>
<td>M</td>
<td>25</td>
<td>Fraser</td>
<td>Fraser North</td>
</tr>
<tr>
<td>M</td>
<td>19</td>
<td>Vancouver Coastal</td>
<td>Vancouver</td>
</tr>
<tr>
<td>F</td>
<td>22</td>
<td>Fraser</td>
<td>Fraser South</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Vancouver</th>
<th>Fraser South</th>
<th>Fraser North</th>
<th>HSDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>1</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>
The Structure of a Cube

• Measures
  e.g. # of patients, # of cases, # of tests, Rate of cases, Average age, Median turn-around-time

• Dimensions
  e.g. Patient, Reported Date, Diagnosis, Health Authority
  - Attributes
    e.g. Gender, Ethnicity
  - Hierarchies
    e.g. Year-Month-Day, HA-HSDA-LHA
Access Methods for Data Mart Cubes

1. Pivot Tables
   – Client identifiers removed

2. Line List Reports
   – Identifiers viewable; users audited

3. Direct SAS script
   – Analytic code
## Pivot Table

### Patients included in the Flu Concise Line List report

<table>
<thead>
<tr>
<th>Result Date: CDC Flu Season</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Result Date: Year</td>
<td>(Multiple Items)</td>
</tr>
<tr>
<td>Proficiency Test: Yes/No</td>
<td>No</td>
</tr>
<tr>
<td>Flu Result</td>
<td>Yes</td>
</tr>
<tr>
<td>Result As of Date</td>
<td>Current</td>
</tr>
<tr>
<td>Patient Health Authority: Health Se (Multiple Items)</td>
<td>Yes</td>
</tr>
</tbody>
</table>

### Tests

<table>
<thead>
<tr>
<th>Virus Type</th>
<th>Virus Subtype</th>
<th>2011 Total</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Not In Hierarchy</em></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>Coronavirus</em></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>Flu A</em></td>
<td></td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td><em>Flu B</em></td>
<td></td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td><em>Human MetaPneumoVirus</em></td>
<td></td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td><em>Respiratory Syncytial virus</em></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>Rhinovirus or Enterovirus</em></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>Grand Total</em></td>
<td></td>
<td>256</td>
<td>256</td>
</tr>
</tbody>
</table>

| 2011 Flu A Total          |               | 256        | 256         |
Online Data Visualization
Data Visualization

- The main goal is to communicate quantitative information clearly and effectively through graphical means.
- As data becomes more widely available and online tools evolve to manipulate, compare, and overlay data, visualizations offer opportunities for:
  - stakeholder engagement
  - situational awareness
  - rapid decision-making
  - advocacy
  - policy support
  - open data
Informative vs. Persuasive

Gun-related murder rates in the developed world

U.S. GUN MURDERS IN 2010

9,595
PEOPLE KILLED

409,280
STOLEN YEARS

This black man was shot in December in Mississippi at the age of 36 by his son during an argument.

Had he not been killed with a handgun, he might have lived to be 84 and died of respiratory disease.
Influenza Surveillance Dashboard

Influenza activity is **moderate** and **stable** this week.

<table>
<thead>
<tr>
<th>Category</th>
<th>Weekly Change</th>
<th>Yearly Change</th>
<th>Historical Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOPC consultation</td>
<td>19.0%</td>
<td>n.a.</td>
<td></td>
</tr>
<tr>
<td>GP consultation</td>
<td>3.5%</td>
<td>3.9%</td>
<td></td>
</tr>
<tr>
<td>Laboratory isolation</td>
<td>5.3%</td>
<td>63.9%</td>
<td></td>
</tr>
<tr>
<td>School absenteeism</td>
<td>1.6%</td>
<td>31.6%</td>
<td></td>
</tr>
<tr>
<td>Hospital admission</td>
<td>70.0%</td>
<td>88.1%</td>
<td></td>
</tr>
</tbody>
</table>

*click figure for details*

Surveillance data displayed: 2010-05-08 to 2011-04-30

Last Updated: 4 May, 2011
Data Visualization Concepts

- Data dimensions, for example
  - disease
  - time
  - age
  - sex
  - geography
- Visual encodings, for example
  - position
  - colour
  - length
  - size
Project Overview

• Goal: To facilitate members of the public, including health care professionals based outside of the BCCDC, to prepare tables and figures that communicate information regarding diseases and public health interventions in British Columbia in a manner that is clear and meaningful.

• Audiences: public, media, researchers, students

• Phase 1 scope: CD Annual Report
Online Data Architecture

Data → Aggregate and De-identify → Public Cube → Online Visualization Service

Re-identification risk policy applied

Firewall

External Programmer → Innovative Use of Data

BCCDC Website

© Provincial Health Services Authority
Privacy

- A public-use data file of aggregate information will be created and will adhere to the re-identification risk policy (comparable to data available in annual reports).
- No personal information will be stored on the site. Only the public-use file will be uploaded.
- The site has no access to internal resources or line-level data.
- A process will be developed to ensure all data is de-identified and that no identifying or quasi-identifying information is ever posted.
Proof of Concept

Acute Hepatitis B

The annual number of acute hepatitis B cases reported in BC continues to decline as seen in Figure 2.4. The national acute hepatitis B rate is not available before 2005; the rate for 2005 onwards is based on only 4 provinces that report acute hepatitis B.

Trend

Living with the Stigma of Hepatitis C
Parental intention to have daughters receive the human papillomavirus vaccine