The Ontario Cancer Registry and its Data Quality

Diane Nishri
Senior Research Associate, Surveillance
February, 2011
Objectives

- Become familiar with cancer registration in Ontario, including issues related to data quality and geography
- Gain proficiency in the use of SEER*Stat
- Generate specific indicators for cancer incidence as defined by APHEO to meet the Ontario Public Health Standards
Outline

• Definitions
  ▪ Terminology
  ▪ Measuring the cancer burden

• The Ontario Cancer Registry
  ▪ Sources of information
  ▪ Record linkage and ‘Case Resolution’

• Data Quality
  ▪ Four dimensions of cancer data quality
  ▪ How does the OCR stack up?
What is cancer?

- A term used for diseases in which abnormal cells divide without control and are able to invade other tissues
- There are more than 100 types of cancer
- Cancer cells can spread to other parts of the body through the blood and lymph systems ("metastasis")
- Tumour/neoplasm: “An abnormal mass of tissue that results when cells divide more than they should or do not die when they should”
Classifying cancer

- Topography: a code indicating the site of origin of a neoplasm

- Morphology: a code describing the type of cell what has become neoplastic and its biologic activity
  - 4 digits cell type (histology)
  - 1 digit behaviour
5th digit behaviour code for neoplasms

- /0 Benign
- /1 Uncertain whether benign or malignant (borderline)
- /2 Carcinoma in situ
- /3 Malignant, primary site
- /6 Malignant, metastatic or secondary site
- /9 Malignant, uncertain whether primary or metastatic site
Cancer coding systems over time

- **Primarily Topography**
  - ICD-6 1948 WHO
  - ICD-7 1955 WHO
  - ICD-8 1965 WHO
  - ICD-9 1975 WHO
  - ICD-10 1992 WHO

- **Primarily Morphology**
  - MOTNAC 1951 ACS
  - Chapter II Neoplasms

- **Both Topography and Morphology**
  - STAT CODE 1956 WHO
  - SNOP 1965 Sections 8.9 Neoplasms CAP
  - MOTNAC 1968 ACS
  - ICD-O 1976 WHO
  - ICD-O-2 (1990) WHO
  - ICD-O-3 (2000) WHO
  - SNOMED Morphology Sections 8.9 Neoplasms CAP

**Abbreviations:**
- WHO, World Health Organization
- ACS, American Cancer Society
- CAP, College of American Pathologists
- ICD, International Classification of Diseases
- MOTNAC, Manual of Tumor Nomenclature and Coding
- STAT, Statistical Code for Human Tumours
- SNOP, Systematized Nomenclature of Pathology
- SNOMED, Systematized Nomenclature of Medicine
Measures of cancer burden

- Incidence
- Mortality
- Survival
- Prevalence
Incidence

- **Definition:** the number of *new* cases of a disease (e.g. cancer) diagnosed in a given population within a *specified period of time*

- Typically expressed as a rate:

\[
\frac{\text{# new cases in a specified time period}}{\text{# persons at risk of disease during the same time period}} \times 100,000
\]
Example – Ontario colorectal cancer incidence

Number of new cases and incidence rates of colorectal cancer, Ontario, 1977–2008

Source: Cancer Care Ontario (Informatics, 2007)
Mortality

- **Definition:** the number of *deaths* due to a disease (e.g. cancer) in a given population within a *specified period of time*

- Typically expressed as a rate:

  \[
  \frac{\text{# deaths in a specified time period}}{\text{# persons at risk of dying during the specified time period}} \times 100,000
  \]

- **Note:** the denominator includes the entire population at risk of dying from the disease, including those that are currently disease-free.
Example – Ontario colorectal cancer mortality

Number of deaths and mortality rates due to colorectal cancer, Ontario, 1977–2008

Source: Cancer Care Ontario (Informatics, 2007)
The Ontario Cancer Registry

- Ontario currently has a population of 13.2 million
  - 38.7% of the Canadian population
- The OCR is population-based
  - Incidence from 1964; mortality from 1950
  - Unique cancer registration methods
    - Passive registration
    - Reliant on administrative records created by others
    - Computerized record linkage and automated medical logic
  - 63,660 malignant incident cases diagnosed in 2007
Passive registration

- The reporting of cancer is *not* legally mandated in Ontario
- There is no staff who visit hospital or non-hospital facilities to find/abstract cases
- The Ontario Cancer Act protects those who provide information on cancer cases to CCO, although it does not mention a cancer registry explicitly
Data Source 1: Hospitals

- Provided by:
  - April 1986+: Canadian Institute for Health Information (CIHI)
- Coding:
  - April 2002+: ICD-10-CA / ICD-O-2
  - morphologies only provided for ~5% of records
Data Source 1: Hospitals

- Coverage:
  - Discharge Abstract Database (DAD): all years
  - Same Day Surgeries (SDS): April 1993-March 2001
  - National Ambulatory Care Reporting System (NACRS): April 2001 to present
Data Source 2: Regional Cancer Centres

- Number of RCCs (including Princess Margaret Hospital)
  - 1981: 8 RCCs
  - c1985 – Sudbury RCC opened
  - 2003-2009: 6 more RCCs opened

- Topography coding:
## Data Source 2: Regional Cancer Centres

<table>
<thead>
<tr>
<th>Coding System</th>
<th>RCCs</th>
<th>PMH</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICD-O-2</td>
<td>1993-2001</td>
<td></td>
</tr>
<tr>
<td>ICD-O-3</td>
<td>2002 +</td>
<td>1999 +</td>
</tr>
</tbody>
</table>
Data Source 3: Pathology

- Phases:
  - 1988-2002: Continued growth
  - 2003: electronic pathology system introduced

- Coding:
  - 2002+: ICD-O-3
Data Source 4: Death Certificates

- Extreme delays in the receipt of coded death certificates from the Registrar General of Ontario hamper timeliness of registration
  - 2006 deaths received May 2009
  - 2007 deaths received March 2010
  - 2008 deaths received January 2011

- Coding:
  - 2000+: ICD-10-CA
Probabilistic Record Linkage

• “the bringing together of information from two records that are believed to relate to the same individual” (Newcombe, 1988)

• You can calculate the likelihood of a correct linkage by comparing the individual identifiers (names, sex, dates, health numbers) and the outcome of these comparisons (exact agreement, partial agreement, disagreement)
“Case Resolution”

• A COBOL program with thousands of lines
• Automated medical logic written in 1981-2
• Tables and logic based in ICD-9 / ICD-O-1
• Some quirks based on beliefs of MD/epidemiologist involved in design (“Breast wins”)
• Conservative multiple primary rules used due to data quality concerns
Ontario’s Multiple Primary Rules

- Timing and laterality do not count (= IACR)

- A new tumour must differ from previous primaries on **both** topography (ICD-9 3 digits) and morphology (“Breg table”, pg xxxvii, ICD-O-2) to be identified as a multiple primary

- IACR rules: different topography or morphology
Geography

- Residents of NW Ontario frequently go to Winnipeg for treatment
  - File received annually from CancerCare Manitoba with Ontario residents registered in Manitoba
- Residents of Gatineau, QC frequently go to Ottawa for treatment
- New York, Michigan, Minnesota??
Which OCR data sources use what codes?

<table>
<thead>
<tr>
<th>Source</th>
<th>SGC</th>
<th>MOH</th>
<th>Postal Code</th>
<th>RG SGC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Path Labs</td>
<td>No</td>
<td>No</td>
<td>Maybe</td>
<td>No</td>
</tr>
<tr>
<td>RCCs</td>
<td>No</td>
<td>Yes&amp;No</td>
<td>Yes?</td>
<td>No</td>
</tr>
<tr>
<td>Deaths</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Residence at diagnosis

- PCCF+ is used to assign the postal code on every source record to an SGC code (CD+CSD)
  - Postal codes which cross boundaries are randomly assigned to an SGC using population weights & a SAS program
- The SGC code on the record closest to the date of diagnosis is taken as the residence at diagnosis
  - Usually the same day, but can be earlier or later
- CDs & CSDs map nicely to PHUs in southern Ontario; it’s more complicated in northern Ontario, but I’ve tried to do the best assignment possible…
  - Brant & Haldimand/Norfolk – Six Nations & New Credit IR
Completeness of postal codes by source

![Graph showing completeness of postal codes by source over time.](image-url)
Missing residence

- Varies by site and time

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Melanoma</td>
<td>10.72%</td>
<td>0.77%</td>
</tr>
<tr>
<td>Oral Cavity</td>
<td>3.10%</td>
<td>0.53%</td>
</tr>
<tr>
<td>Thyroid</td>
<td>1.82%</td>
<td>1.58%</td>
</tr>
<tr>
<td>All Cases</td>
<td>1.37%</td>
<td>0.37%</td>
</tr>
</tbody>
</table>

- Historically, has been more of a problem for sites with a high percentage of pathology only cases
- Tends to be a characteristic of the hospital/clinic/lab
Postal code or MOH code?

- 355,787 hospital records, 2001
- PCCF+ run on the postal code
  - 1.9%  no match
  - 20.1% assigned using population weights
  - 76.7%  no problem
- MOH to SGC conversion table
  - 0000 patient w no fixed address
Postal code or MOH code, cont’d

- SGCs compared for 272,981 records with no problems
  - 97.9% postal code & MOH agree
  - 1.4% CD agrees, but CSD disagrees
  - 0.7% nothing agrees

- Common disagreements?
  - MOH says Toronto, postal code says no: 28%
  - Postal code says Toronto, MOH says no: 6.5%
  - Wrong MOH code for 2 places with the same name
Differential misclassification bias in residence assignment

- In 1991, Marrett & Weir looked at the accuracy of the residence data for Frontenac County cases for 1988
  - 14% of cases had an incorrect residence
  - 7% involved Kingston, with people living outside Kingston were being incorrectly attributed residence in Kingston
- As a result, Kingston rates were too high and the rest of Frontenac County rates too low
Age Adjusted Incidence Rate
Ontario, 1996-2005

Distribution by Census Division

+: Rate ratio significantly high (p<.05)
-: Rate ratio significantly low (p<.05)
O: Incidence case is zero
Data Quality in cancer registries

- The uses of cancer registries are changing
  - Used to be mainly surveillance and epidemiology
  - Now expected to be used to monitor screening programs, evaluate treatment and prevention programs, project future incidence, etc.

- The ability of a cancer registry to effectively perform the above functions depends on the registry’s data quality
Data Quality in cancer registries

- Both internal and external factors affect data quality

<table>
<thead>
<tr>
<th>Internal factors</th>
<th>External factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Registry operations</td>
<td>• Purpose of data collection</td>
</tr>
<tr>
<td>• Registry resources</td>
<td>• Resources of external data sources</td>
</tr>
<tr>
<td>• Access to additional data sources</td>
<td>• Reporting diligence</td>
</tr>
</tbody>
</table>
Dimensions of data quality in cancer registries

• Four dimensions of data quality (Parkin & Bray (2009)):
  1. Comparability
  2. Completeness
  3. Accuracy (validity)
  4. Timeliness
Comparability

- **Definition:** *the extent to which registry practices (e.g. criteria for registration, coding practices, etc.) adhere to standard guidelines and can be compared over time and across registries.*

- Important when analyzing and interpreting variations in cancer burden across space and time

- **Methods to ensure comparability:**
  - Documentation of registry practices
  - Standards for registration and coding (e.g. NAACCR, CCR guidelines)
North American cancer registries

United States

SEER

NPCR

NAACCR

Canada

PTCR

CCR
Comparability

• Classification and coding systems
  - Example: the coding of behaviour for borderline malignant ovarian cancers

• Definition of incident cases
  - Behaviour codes captured
    - In situ (/2) bladder; benign brain cancers

• Definition of a primary cancer
  - SEER, IACR, CCR rules for counting multiple primaries
Completeness

- **Definition:** The extent to which all incident cancers in the population are included in the registry (i.e. case ascertainment).

- Indicates how closely incidence rates and survival proportions reflect their true value

- **Methods to ensure completeness:**
  - Multiple data sources to register cases
  - Legislation for reporting incident cases
  - Record linkage with other databases (e.g. other registries, national/provincial death clearance)
Completeness

- **Semi-quantitative** methods for assessing completeness:
  - Historic data methods
    - E.g. compare counts/rates to historic values or to other jurisdictions
  - Incidence:mortality (mortality:incidence) ratios
  - Number of sources per case
  - % single sources (e.g. death certificate only, hospital only, etc.)

- **Quantitative** methods for assessing completeness:
  - Independent case ascertainment
    - Case-finding audits (i.e. rescreening of a sample of sources)
Completeness

- Reporting sources for incident cases in the OCR, 2007

**Hospital discharges**: 69%

**Path reports**: 83%

**Day surgery**: 34%

**Deaths**: 14%

<table>
<thead>
<tr>
<th>% Single Source</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hosp only</td>
<td>5%</td>
</tr>
<tr>
<td>Path only</td>
<td>7%</td>
</tr>
<tr>
<td>DC only</td>
<td>2%</td>
</tr>
<tr>
<td>Clinics only</td>
<td>3%</td>
</tr>
<tr>
<td>(Day surgery only +2%)</td>
<td></td>
</tr>
</tbody>
</table>
Accuracy (validity)

- **Definition:** *the proportion of registered cases with a given characteristic that truly have that attribute.*

- **Methods to ensure accuracy:**
  - Multiple sources to register cases
  - Training of registry personnel
  - Electronic ‘edits’
  - Record linkage to other databases (e.g. provincial health insurance plans)
  - Trace-back and follow-back procedures
Quantitative methods for assessing accuracy

• Reabstraction studies and recoding audits
  ▪ Assess agreement with source records or among coders

• Histologic verification:
  ▪ % microscopically verified (e.g. histology, cytology, autopsy)

• Death certificate only (DCO) cases
• Missing information
• Internal consistencies
**Timeliness**

- **Definition:** *the speed with which a registry can collect, process, and report complete and accurate cancer data*

- Depends on two time intervals:
  - Time until receipt: *time from diagnosis to receipt of report*
  - Process time: *time from receipt to data availability*

- Ensures up-to-date data is available to health researchers, providers, and planners

- A trade-off exists between the timeliness of the data and the completeness and accuracy
Data quality standards for cancer registries

- Several standards for acceptable data quality (e.g. NAACCR criteria, SEER, CCR guidelines, etc.)

- Data quality should always be considered in context of use
  - Example: Standards for assessing common indicators for accuracy

<table>
<thead>
<tr>
<th>Indicator Name</th>
<th>CCR Optimal Value</th>
<th>NAACCR Gold Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microscopically Confirmed Cases</td>
<td>≥ 93%</td>
<td>--</td>
</tr>
<tr>
<td>Death Certificate Only†</td>
<td>≤ 3%</td>
<td>≤ 3%</td>
</tr>
</tbody>
</table>
Data quality implications for the OCR

• Comparability
  ▪ Multiple primary rules
  ▪ Population data

• Completeness
  ▪ AB/MB/ON Case Ascertainment Study

• Accuracy/Validity
  ▪ Ontario re-abstraction study

• Timeliness
  ▪ Not so good…