New Frontiers in Defining Data Quality

Measuring Completeness on the Northern Frontier: The Canadian Case Ascertainment Project

Donna Turner, PhD
Epidemiologist and CIHR Sr. Research Fellow
Epidemiology & Cancer Registry
CancerCare Manitoba
Outline

• An Introduction to the Canadian Case Ascertainment Project
  – Purpose
  – Partners
  – Methods

• Results
  - Of Each Method
  - Synthesis

• And now …?
Outline

• An Introduction to the Canadian Case Ascertainment Project
  – Purpose
  – Partners
  – Methods

• Results
  - Of Each Method
  - Synthesis

• And now ...?
Purpose

• To enable estimation of the completeness of case ascertainment using the most accurate and cost-effective method.
Objectives

• To compare 4 protocols to estimate completeness of case ascertainment:
  – Record linkage
  – Case re-finding
  – Capture-recapture
  – I/M ratio

• Applicability? Comparability? Costs?
• Which is the “best”? 
Partners: Participants

– Juanita Hatcher, Rong Huang, Doug Dover, Carol Russel, Maxine Raphael

– Donna Turner, Erich Kliewer, Daojun Mo, Catherine Njue, Jeri Kostyra, Marianne Sutherland, Shelley Derksen, Wendy Fonseca-Holt

– Eric Holowaty, Loraine Marrett, Diane Nishri, Gemma Lee, Darlene Dale
Provinces

Alberta
Population ~ 3.1 million
Incident Cancer Cases ~ 10,800 pa (excl NMSC)

Manitoba
Population ~ 1.1 million
Incident Cancer Cases ~ 5,400 pa (excl NMSC)

Ontario
Population ~ 11.8 million
Incident Cancer Cases ~ 50,200 pa (excl NMSC)
Partners: Funding

- Statistics Canada: ROADMAP
- Participating registries (in kind)
- September 2000 through August 2002
Methodologic Approaches

- Record linkage
- Case re-finding
- Capture-recapture
- I/M ratio
### Design

<table>
<thead>
<tr>
<th>Method</th>
<th>AB</th>
<th>MB</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>Record Linkage</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Case Re-finding</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Capture-Recapture</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>I/M Ratio</td>
<td>X</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>
Outline

• An Introduction to the Canadian Case Ascertainment Project
  – Purpose
  – Partners
  – Methods

• Results
  - Of Each Method
  - Synthesis

• And now ...?
Record Linkage

- Manitoba, Alberta

Second data source (~comprehensive) → Provincial Cancer Registry
Record Linkage

- Manitoba, Alberta

Provincial Hospital Discharge and Day Surgery Files (1994-1996) where Dx=Invasive Malig.

Common Identifiers

• Personal Health Number (PHN)
• Sex
• Date of Birth
• Last Name
• Initials
Record Linkage: Results

- Alberta:
  - 27,811 already reported (1994-1996)
  - 6,898 “candidates” followed back
  - 368 new registrations
    - \((27,811)/(27,811+368) = 98.7\%\) complete
- Manitoba:
  - 15,566 already reported (1994-1996)
  - 1,497 “candidates” followed back
  - 78 new registrations
    - \((15,566)/(15,566+78) = 99.5\%\) complete
Record Linkage: Results

- Alberta: process beneficial for
  - Kaposi’s sarcoma
  - Malignancies of the eye
  - Leukemia
  - Malignancies of male genitalia
- Manitoba: process beneficial for
  - Malignancies of the brain/CNS
  - Leukemia
  - Note: possible issues with older and rural cases
Case Re-finding

- Ontario, Alberta

Labs: 50+ Cases/Yr (1996)

Sample

Provincial Cancer Registry (1996) where Dx=Invasive Malig.
Case Re-finding

Sampling ...

- Laboratories with more than 50 reports per year
- Stratified by size
- Random sample in each strata
- Random start date in 1996
- Period of sample determined to provide required number of cases
- Total sample size 1200 cases
Case Re-finding (ON)

- OCR incident cases 1996
- Stratified random sample: 1200 cases
- 200 - 400 cases per stratum; 2 - 6 mos. sampled

- RCCs 30%
- Large Hosp. 34%
- Small Hosp. 34%
- Private 2%
- PATH 3 / 20
- 4 / 60
- 2 / 4
- 3 / 20
Case Re-finding

Reports examined
• Surgical Pathology
• Cytology
• Bone Marrow
• Autopsy
• (Peripheral Bloods)
Case Re-finding: Results

- Alberta:
  - 1,552 eligible reports identified (1996)
  - 1,425 tumours
  - 17 not previously registered in the ACR
  - **Completeness: 98.8%**

- Ontario:
  - 2,230 eligible reports identified (1996)
  - 1,325 tumours
  - 46 not previously registered in the OCR
  - **Completeness: 97.5%**
Case Re-finding: Results (ON)

Weighted Average Completeness: 97.5% (96.7-98.3)
Capture-Recapture

- Ontario, Manitoba (1994-1996)
## Capture-Recapture

### % Reporting by Information Source

<table>
<thead>
<tr>
<th></th>
<th>FEMALES</th>
<th></th>
<th>MALES</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MB</td>
<td>ON</td>
<td>MB</td>
<td>ON</td>
</tr>
<tr>
<td>Clinic</td>
<td>60.3</td>
<td>52.0</td>
<td>53.4</td>
<td>43.8</td>
</tr>
<tr>
<td>Hosp.</td>
<td>96.2</td>
<td>90.1</td>
<td>93.3</td>
<td>87.8</td>
</tr>
<tr>
<td>Path.</td>
<td>90.7</td>
<td>76.2</td>
<td>90.6</td>
<td>74.7</td>
</tr>
<tr>
<td>Death</td>
<td>36.5</td>
<td>31.7</td>
<td>39.8</td>
<td>35.5</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>7,362</td>
<td>60,276</td>
<td>8,040</td>
<td>64,817</td>
</tr>
</tbody>
</table>
Capture-Recapture: Qualifiers!

- Statistical models: tried simple and multiple sources, log-linear, interactions, ...
- Stability issues!
  - Significant three-way interactions in ON
- Not all registries record all reporting sources
- Source accrual patterns impact estimates
  - Presence/absence of sources
  - Degree of overlap
  - Makes interpretation difficult!
Capture-Recapture: Results

• Manitoba: high degree of overlap of sources
  – Males: 99.1% - 100%
  – Females: 99.5% - 100%

• Ontario:
  – Males: 79.6% - 87.7%
  – Females: 77.9% - 80.8%
I/M Ratio: Variations on a Theme

- Different standard populations?
- Adjusting for case fatality?
- Different standard registries?
  - SEER, CCR, Saskatchewan, Alberta
- First Primary instead of all primaries?
- Different death coding? (Alberta only)
- Stability over age groups (45-64, 65-74, 75+)?
# I/M Ratio

**Different Standard Populations and I/M Standard Male Colorectal**

<table>
<thead>
<tr>
<th>Gold Std I/M</th>
<th>Total US Incid &amp; Mort</th>
<th>White US Incid &amp; Mort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>84.9%</td>
<td>85.4%</td>
</tr>
<tr>
<td>ON</td>
<td>85.3%</td>
<td>85.8%</td>
</tr>
<tr>
<td>MB</td>
<td>90.6%</td>
<td>90.3%</td>
</tr>
<tr>
<td>AB</td>
<td>92.7%</td>
<td>92.6%</td>
</tr>
</tbody>
</table>
## I/M Ratio

### Adjusting for Case-Fatality Colorectal Cancer

<table>
<thead>
<tr>
<th>Sex</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusted for Case Fatality</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Canada</strong></td>
<td>84.0%</td>
<td>87.9%</td>
</tr>
<tr>
<td><strong>ON</strong></td>
<td>84.4%</td>
<td>88.5%</td>
</tr>
<tr>
<td><strong>MB</strong></td>
<td>88.8%</td>
<td>92.4%</td>
</tr>
<tr>
<td><strong>AB</strong></td>
<td>91.1%</td>
<td>92.9%</td>
</tr>
</tbody>
</table>

*(Std Registry=SEER White, Std Popn=US 1970)*
I/M Ratio (cont’d)

• First Primary instead of all primaries?
  – *dramatic increase by using only the first primary*

• Different death coding? (Alberta only)
  – *for breast, colorectal, prostate, lung, using VS death cause leads to higher I/M ratio; for liver, we observe the opposite*

• Stability over age groups (45-64, 65-74, 75+)?
  – *variability over age groups depends on time and the registry to be evaluated*
Outline

• An Introduction to the Canadian Case Ascertainment Project
  – Purpose
  – Partners
  – Methods

• Results
  - Of Each Method
  - Synthesis

• And now …?
## Synthesis

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AB</strong></td>
<td>98.7%</td>
<td>98.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MB</strong></td>
<td>99.5%</td>
<td></td>
<td>99.1% - 100%</td>
<td></td>
</tr>
<tr>
<td><strong>ON</strong></td>
<td>97.5%</td>
<td>77.9% - 80.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1995: 97.5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1996: 96.8%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1995: 84.7%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1996: 92.9%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1995: --------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1996: 91%</td>
</tr>
</tbody>
</table>
Conclusions

• Record linkage
  – Estimates higher than I/M ratio
  – Resource Intensive for Follow Back
• Case re-finding
  – Privacy Legislation problematic
• Capture-recapture
  – Relatively efficient
  – ? Statistical stability
• I/M ratio
  – Simple to compute
  – Issues related to interpretation
Recommendations

• **Case re-finding** appears to be the most resource efficient and accurate method for an overall estimate of completeness.
  – Also enables low precision hospital specific and site specific estimates

• **Record linkage** may be appropriate for resource rich registries

• **If Incidence/Mortality estimates are used** care must be taken over choice of Gold standard Registry, and incidence and mortality rates should be stable
Outline

• An Introduction to the Canadian Case Ascertainment Project
  – Purpose
  – Partners
  – Methods

• Results
  - Of Each Method
  - Synthesis

• And now …?
What we say to cats...

Well, Fluffy, you've clawed the furniture for the last time! I'll not tolerate that behavior any longer!

What they hear
What we say to dogs
Okay, Ginger! I've had it! You stay out of the garbage! Understand, Ginger? Stay out of the garbage, or else!

What they hear
blah blah GINGER blah blah blah blah GINGER blah blah blah GINGER blah...
NAACCR I/M Review Team

- Nancy Weiss, Texas Cancer Registry
- Melanie Williams, Texas Cancer Registry
- Brenda Edwards, NCI
- Maria Schymura, New York State Cancer Registry
- Ron Dewar, Nova Scotia Cancer Registry
- Eric Holowaty, Ontario Cancer Registry
- Larry Ellison, Statistics Canada
- Maxine Raphael, Alberta Cancer Registry
- Donna Turner, Manitoba Cancer Registry (Chair)
NAACCR I/M Review Team

• First: be **clear** on what the elements of the I/M ratio are and how they came to be
• **Then** look into issues:
  – Standard error?
  – Race issues?
  – Use of the IM ratio in the face of changing methods of detection and changing survival?
  – Application of SEER experience to non-SEER areas?
  – First vs subsequent primaries?
  – Influence of unknown primaries?
NAACCR I/M Review Team

- Nancy Weiss, Texas Cancer Registry
- Melanie Williams, Texas Cancer Registry
- Brenda Edwards, NCI
- Maria Schymura, New York State Cancer Registry
- Ron Dewar, Nova Scotia Cancer Registry
- Eric Holowaty, Ontario Cancer Registry
- Larry Ellison, Statistics Canada
- Maxine Raphael, Alberta Cancer Registry
- Donna Turner, Manitoba Cancer Registry (Chair)