Using Text Fields to Determine Out Of State Diagnoses in Central Cancer Registries

Laura E Soloway, PhD MPH
Francis P Boscoe PhD
New York State Cancer Registry
Bureau of Cancer Epidemiology
New York State Department of Health, Albany, NY
Introduction

• Because of the mobile nature of the NY population and the worldwide reputation of NY hospitals, many non-residents are reported to the New York State Cancer Registry (NYSCR).

• Over 10% of all cancer patients seen in NYS hospitals were not NYS residents (1995-2009)
Introduction

Frequency of Cases from Out of State Residents Seen in NYS Hospitals 1995-2009

N=144927

MEMORIAL SLOAN-KETTERING CANCER...
NY PRESBYTERIAN HOSPITAL-NY WEIL...
NY PRESBYTERIAN HOSPITAL-COLUMBIA...
MOUNT SINAI HOSPITAL
ROSWELL PARK CANCER INSTITUTE
STATEN ISLAND UNIVERSITY HOSPITAL:
BETH ISRAEL MEDICAL CENTER-PETRIE DIV.
NYMC ST. VINCENT'S MANHATTAN
LENNOX HILL HOSPITAL
OUR LADY OF LOURDES MEMORIAL...
**Introduction**

- Medical Tourism: the practice of traveling to another city, state, or country for the sole purpose of obtaining medical care
- Worldwide, 60,000-85,000 medical tourists a year*
- 40% of these are searching for the world’s most advanced technology and head to the US*

*Data from McKinsey &Co Survey 2008*
Introduction

- **www.medicaltourismnyc.com**
  - Manhattan plastic surgeon creates an app to boost medical tourism to NYC
  - by David Harvey, Editor on May 16, 2011, American News Report
- **www.onlinemedicaltourism.com**
- **www.meditravels.org**
Introduction

- Many hospitals have international services specifically for foreigners
- Goes for out of state or even out of city patients as well
Aim

• To examine the text fields of cases on the New York State Cancer Registry database for references to diagnosis in other countries or states.
Methods

• Out of Country
  • All text fields (mostly from 1995-2009) were scanned for occurrences of country names using the index function in SAS 9.2 (Cary, NC).

• Eliminated various common combinations of text strings found in text that indicated non-country values such as: “Beth Israel”; “Dr. Jordan”; “Jamaica, NY” and “Cuba, NY”; and “Vietnam veteran”.
Methods

- Out of State
  - All text fields (mostly from 1995-2009) were scanned for occurrences of (non-NY) state names using the index function in SAS 9.2 (Cary, NC).
  - Instances of New Jersey and Georgia were removed during the out of country process.
Methods (Summary)

• After cleaning the data, we used the “findw” function in a macro in SAS to locate instances of country and state names.

• These instances were marked and then were able to be pulled out using arrays.

• We eliminated common phrases that showed non-out of country cases.
Methods

- Common phrases deleted:

| JAMAICA HOSPITAL                     | DR MONACO                      |
| JAMAICA HOSP                       | DR. MONACO                    |
| JAMAICA SO                        | DR. JORDAN                    |
| JAMAICA, NY                      | DR. JORDAN                    |
| BETH ISRAEL                     | DR. EDWARD JORDAN             |
| BRONX LEBANON                   | JORDAN STERN                  |
| COLOMBIA PRESBYTERIA            | CUBA, NY                      |
| NEW JERSEY                    | CUBA, NEW YORK                |
| CONGO RED                    | VIETNAM VET                   |
Methods

- Manual review to determine if the case was actually diagnosed in New York or the patient was not a resident of New York at the time of diagnosis.
Results

- Out of Country
  - 1666 cases flagged for review
  - 1130/1666 or 68% were determined to be out of country diagnoses and/or not actually New York State residents at the time of diagnosis
Results

Frequency of Original Counties Out of Country Diagnoses Were Assigned

\[ N = 1057 \]
Results

Countries Cases Were Reassigned to After Analysis

N=884
Results

- Out of State
  - 1095 cases flagged for review
  - 302/1095 or 28% were determined to be NYS residents/diagnoses
  - 362/1095 or 33% were determined to be FL diagnoses
  - 55/1095 or 5% were determined to be VT diagnoses
Results

Frequency of Actual Diagnosis States for Cases found as Out Of State

<table>
<thead>
<tr>
<th>State</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL</td>
<td>400</td>
</tr>
<tr>
<td>NY</td>
<td>350</td>
</tr>
<tr>
<td>VT</td>
<td>250</td>
</tr>
<tr>
<td>CA</td>
<td>200</td>
</tr>
<tr>
<td>PA</td>
<td>150</td>
</tr>
<tr>
<td>VA</td>
<td>100</td>
</tr>
<tr>
<td>TX</td>
<td>50</td>
</tr>
<tr>
<td>NC</td>
<td>50</td>
</tr>
<tr>
<td>OH</td>
<td>50</td>
</tr>
<tr>
<td>AZ</td>
<td>50</td>
</tr>
<tr>
<td>MD</td>
<td>50</td>
</tr>
<tr>
<td>SC</td>
<td>50</td>
</tr>
<tr>
<td>MA</td>
<td>50</td>
</tr>
</tbody>
</table>
Results

Frequency of Actual Diagnosis States for Cases Found as Out Of State

- VT: 60
- CA: 50
- PA: 40
- VA: 30
- TX: 20
- NC: 10
- OH: 20
- AZ: 10
- MD: 10
- SC: 10
- MA: 10
Results

- Total Incident Cases Changed:
  - Approximately 1500 cases changed from NYS residents/diagnoses to non-NYS residents/diagnoses
Conclusions

• Out of Country
  - 68% of flagged cases were determined to be out of country diagnoses

• Most reports came from major metropolitan areas with large cancer facilities: NYC metro area, Buffalo, Rochester
Conclusions

• **Out of State**
  - 66% of flagged cases were probably NYS residents at the time of diagnosis
  - 33% of flagged cases were probably not NYS residents at the time of diagnosis

• References to “Wyoming County” and “Delaware Valley” could have been removed systematically but were not.
Conclusions

- An important factor in determining whether or not a person was diagnosed and/or lived in NY at the time of diagnosis.

- Has implications for various types of analysis as public reports of incidence rates are limited to state residents and we often limit our studies to those who live in NYS as well.
Strengths

- Allows for more complete data for NYSCR
- Allows for better analysis of NYS resident data for outside researchers
Limitations

- Cases must be manually reviewed by cancer registry staff

- Have to manually review many cases that are “false positives” because not all contingencies can be weeded out

- Cannot know residence history through current mechanisms
Acknowledgements

• Funding Sources:
  • This work was supported in part by the Centers for Disease Control and Prevention’s Cooperative Agreement U58/DP000783, awarded to the New York State Department of Health through the National Program of Cancer Registries.
Thank You!

Questions?