Development of Methods to Evaluate the Pre-Diagnostic Breast Cancer Patient Trajectory Linking Existing Population-Level Databases

Marcy Winget, PhD
Alberta Health Services, University of Alberta
April 2013
Outline

• Objectives
• Data
• Results
• Discussion
Objective

• Developing and validating algorithm for identifying patient presentation through symptom or screening
• Understand the patterns diagnostic imaging utilization for breast cancer patients
• Feasibility of using physician billing data to assess these care patterns
• Assess the time from presentation to diagnosis
Inclusion Criteria

- Female patients diagnosed with first ever invasive breast cancer in Alberta, 2007-2010
- Residents in Alberta at diagnosis
- Histologic confirmation on breast cancer
- At least one diagnostic imaging one-year prior to breast cancer diagnose date
## Data Sources

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient ID</td>
<td>Patient ID</td>
<td>Patient ID</td>
<td>Patient ID</td>
</tr>
<tr>
<td>Demographics</td>
<td>Dates/results of screening mammogram</td>
<td>Dates/results of all screening and diagnostic mammogram, biopsies</td>
<td>Diagnosis procedures</td>
</tr>
<tr>
<td>Tumor details</td>
<td>Dates/results of subsequent diagnostic imaging</td>
<td></td>
<td>Anonymized physician ID</td>
</tr>
<tr>
<td>Identify cohorts</td>
<td></td>
<td></td>
<td>Visit Date</td>
</tr>
</tbody>
</table>

Data Sources: www.albertahealthservices.ca
- Separate 1000 patients for testing purpose
- Remaining 5994 patients in the training data pool for algorithm development
Algorithm flow chart for dataset II

Total 5994

No Screening mammogram N=2817 (47%)

With Screening Mammogram N=3177 (53%)

1st test >9m and 1st test = s_mammo N=229 (3.8%)

1st test >9m and 1st test ≠ s_mammo N=22 (0.36%)

1st test <9m and 1st test = s_mammo N=2926 (49%)

Symptom led N=3007 (50%)

Only one s_mammo and it is >9m N=190 (3.2%)

Multi s_mammo and last one <9m N=39 (0.65%)

s_mammo in middle and last one <9m N=22 (0.36%)

Screening led N=2987 (50%)
Venn Diagram for datasets

- Screening: 424 patients
- ASR: 6836 patients
- Billing: 8045 patients

Total patients: 8379

264 patients could not be found in any of the three datasets.
## Results

<table>
<thead>
<tr>
<th>Billing data &amp; Screening data</th>
<th>ASR data &amp; Screening data</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Screening led</td>
<td>Symptom led</td>
<td>Total</td>
</tr>
<tr>
<td>Screening led</td>
<td>2591</td>
<td>396</td>
<td></td>
<td>2987</td>
</tr>
<tr>
<td>Symptom led</td>
<td>118</td>
<td>2889</td>
<td></td>
<td>3007</td>
</tr>
<tr>
<td>Total</td>
<td>2709</td>
<td>3285</td>
<td></td>
<td>5994</td>
</tr>
</tbody>
</table>

Concordance: 91%

Dis-concordance: 9%

N=25, Scr. mammo $>9$ months, but negative results

N=93, Showed up as diag. mammo, but screening led in ASR
Frequency of diagnostic imaging in dataset I & dataset II

- Screening mammogram:
  - Dataset I: 4%
  - Dataset II: 6%

- Diagnostic mammogram:
  - Dataset I: 68%
  - Dataset II: 84%

- Ultrasound:
  - Dataset I: 75%
  - Dataset II: 92%

- Screening mammogram:
  - Dataset I: 100%
  - Dataset II: 93%

- Diagnostic mammogram:
  - Dataset I: 75%
  - Dataset II: 79%

- Ultrasound:
  - Dataset I: 100%
  - Dataset II: 79%

Legend:
- Blue: Screen led in DATA I (N=2709)
- Red: Symptom led in DATA I (N=3285)
- Green: Screen led in DATA II (N=2987)
- Purple: Symptom led in DATA II (N=3007)
Kaplan - Meier curve for time from first relevant test to diagnosis in dataset I

**Analysis Time**

- **Screening Led**
- **Symptom Led**

Graph showing the Kaplan-Meier curve with analysis time ranging from 0 to 200, illustrating the time from first relevant test to diagnosis for different led strategies.
Kaplan-Meier curve for time from first relevant test to diagnosis in dataset II **
Kaplan - Meier curve for time from first relevant test to diagnosis in dataset I & II: Screening Led patients only **
Kaplan-Meier curve for time from first relevant test to diagnosis in dataset I & II: Symptom Led patients only.
Discussion