The Quality of Cancer Registry Data: Preliminary Results from CDC-NPCR’s Patterns of Care Study

Robert R. German, DrPH, MPH; Tim Byers, MD, MPH; Holly Wolf, PhD, MSPH; Jen Wike, MBA, MPH; Linda Alley, PhD; and PoC Study Group

June 2006  NAACCR annual conference
Acknowledgement of Authors

Robert R. German, DrPH, MPH*
Tim Byers, MD, MPH†
Holly Wolf, PhD†
Jen Wike, MBA, MPH*‡
Linda Alley, PhD*

* CDC’s National Program of Cancer Registries, Atlanta, GA
† University of Colorado Health Services, Denver, CO
‡ Northrop Grumman Corporation, Atlanta, GA

The findings and conclusions in this presentation are those of the authors and do not necessarily represent the views of CDC.
Agenda

• Background on CDC-NPCR’s Breast, Colon, and Prostate Cancer Data Quality and Treatment Patterns of Care (PoC) Study
  – Data quality component

• Results of data quality analysis

• Discussion and concluding remarks
PoC Study, #1

- In 1999, the Institute of Medicine (Hewitt and Simone, 1999) recommended the use of existing data, such as CDC-NPCR’s,
  - To measure variations in the use of appropriate standards of care
  - To assess the quality of cancer care in the US.
- A year later, the IOM (Hewitt and Simone, 2000) noted that some cancer patients may not be receiving the care known to be most effective for their condition.
PoC Study, #2

• Collaborative study (McDavid et al., 2004)
  – CDC’s National Program of Cancer Registries and statewide cancer registries in CA, CO, IL, LA, NY, RI, SC, and DC

• Main objectives
  – Assess quality and completeness of cancer data collected by registries
  – Evaluate the extent to which patients are receiving stage-specific, guidelines-based cancer treatments
PoC Study, #3

• Simple random samples
  – Localized female breast cancer (C50.0-C50.9)
  – Regional 3 or 4 colon cancer (C18.0-C18.9)
  – Localized prostate cancer (C61.9)

• Re-abstraction of data on patient, tumor, treatment
  – Hospital charts
  – Physician offices
  – Ambulatory surgery centers
  – Radiation therapy centers
Data Quality Analysis in PoC, #1

• Data from all states in PoC except DC
• Comparison of re-abstracted data with those originally coded in the registry
  – Original data archived or “frozen”
  – Merged by state and case identification number
  – Single primaries only
Data Quality Analysis in PoC, #2

• All three stage-specific cancers in PoC
  – Stage at diagnosis with CONCORD data in PoC
• Measurements of agreement: kappa statistic, % agreement, sensitivity (Szklo and Nieto, 2000)
  – Re-abstracted data as ‘Gold Standard’
  – Good quality or better: $k \geq 0.60$ and % agreement and sensitivity $\geq 80\%$ (Cicchetti, 2001)
• SEER summary stage 1977, race/ethnicity, breast and prostate grade, breast histology, surgery, radiation, chemotherapy, hormone therapy, and biologic response modifier
SEER Summary Stage Agreement, CONCORD data in PoC Study, DY 1997-1998

Female breast cancer (n=3,653)

- In situ (k=93) ☺
- Local (k=0.88) ☺
- Regional (k=0.88) ☺
- Distant (k=0.77)
- Unk/unstag (k=0.32)

% Agree  Sn (%)

Colon cancer (n=2,566)

- In situ (k=0.86) ☺
- Local (k=0.69) ☺
- Reg 2,5 (k=0.61)
- Reg 3,4 (k=0.85)
- Distant (k=0.85)
- Unk/unstag (k=0.45)

% Agree  Sn (%)

k = kappa statistic
% Agree = percent agreement
Sn = sensitivity

☺ Good quality or better based on kappa >= 0.60 and each of percent agreement and sensitivity >= 80% (Cicchetti, 2001).
SEER Summary Stage Agreement, #2

Prostate cancer (n=3,503)

- Local (k=0.60)
- Regional (k=0.69)
- Distant (k=0.80)
- Unk/unstag (k=0.35)

Summary of n-by-n tables

<table>
<thead>
<tr>
<th>Cancer</th>
<th>% unknown/unstaged re-abstracted to stage</th>
<th>% staged re-abstracted to unknown/unstaged</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female breast</td>
<td>66%</td>
<td>2%</td>
</tr>
<tr>
<td>Colon</td>
<td>55%</td>
<td>2%</td>
</tr>
<tr>
<td>Prostate</td>
<td>61%</td>
<td>6%</td>
</tr>
</tbody>
</table>

k = kappa statistic
% Agree = percent agreement
Sn = sensitivity

Good quality or better based on kappa >= 0.60 and each of percent agreement and sensitivity >= 80% (Cicchetti, 2001).

More unknown/unstaged female breast cancers in re-abstracted data (n=101) than in original data (n=99).

In situ or localized female breast cancer (n=3,551)

- NH WH (k=0.87)
- NH BL (k=0.95)
- NH Oth (k=0.83)
- Hisp (k=0.78)

Regional 3 or 4 colon cancer (n=1,328)

- NH WH (k=0.93)
- NH BL (k=0.97)
- NH Oth (k=0.96)
- Hisp (k=0.82)

Stage at diagnosis is SEER summary stage 1977 in 2003 re-abstraction.

NH WH = non-Hispanic white
NH BL = non-Hispanic black
NH Oth = non-Hispanic other
Hisp = Hispanic, all races

k = kappa statistic
% Agree = percent agreement
Sn = sensitivity

摲 □ Good quality or better based on kappa >= 0.60 and each of percent agreement and sensitivity >= 80% (Cicchetti, 2001).
Race/Ethnicity Agreement, #2

Localized prostate cancer (n=3,281)

<table>
<thead>
<tr>
<th>Race/Ethnicity</th>
<th>% Agree</th>
<th>Sn (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH WH (k=0.90)</td>
<td>☺</td>
<td></td>
</tr>
<tr>
<td>NH BL (k=0.95)</td>
<td>☺</td>
<td></td>
</tr>
<tr>
<td>NH Oth (k=0.80)</td>
<td>☺</td>
<td></td>
</tr>
<tr>
<td>Hisp (k=0.78)</td>
<td>☺</td>
<td></td>
</tr>
</tbody>
</table>

Stage at diagnosis is SEER summary stage 1977 in 2003 re-abstraction.

NH WH = non-Hispanic white
NH BL = non-Hispanic black
NH Oth = non-Hispanic other
Hisp = Hispanic, all races

k = kappa statistic
% Agree = percent agreement
Sn = sensitivity

☺ Good quality or better based on kappa >= 0.60 and each of percent agreement and sensitivity >= 80% (Cicchetti, 2001).

Summary of n-by-n tables

<table>
<thead>
<tr>
<th>Cancer</th>
<th>% unknown race/ethnicity re-abstracted to race/ethnicity</th>
<th>% race/ethnicity re-abstracted to unknown race/ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>In situ or localized female breast</td>
<td>76%</td>
<td>0.4%</td>
</tr>
<tr>
<td>Regional 3,4 colon</td>
<td>80%</td>
<td>0.2%</td>
</tr>
<tr>
<td>Localized prostate</td>
<td>74%</td>
<td>0.5%</td>
</tr>
</tbody>
</table>
Grade Agreement, PoC Study, DY 1997-1998

In situ or localized female breast cancer (n=3,549)

<table>
<thead>
<tr>
<th>Grade</th>
<th>% Agree</th>
<th>Sn (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade I (k=0.79)</td>
<td>😊</td>
<td></td>
</tr>
<tr>
<td>Grade II (k=0.74)</td>
<td>😊😊</td>
<td></td>
</tr>
<tr>
<td>Grade III (k=0.75)</td>
<td>😊😊😊</td>
<td></td>
</tr>
<tr>
<td>Grade IV (k=0.33)</td>
<td>😊😊😊😊</td>
<td></td>
</tr>
<tr>
<td>Grade Unk (k=0.69)</td>
<td>😊😊😊😊😊</td>
<td></td>
</tr>
</tbody>
</table>

Localized prostate cancer (n=3,281)

<table>
<thead>
<tr>
<th>Grade</th>
<th>% Agree</th>
<th>Sn (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade I (k=0.79)</td>
<td>😊</td>
<td></td>
</tr>
<tr>
<td>Grade II (k=0.69)</td>
<td>😊😊</td>
<td></td>
</tr>
<tr>
<td>Grade III (k=0.74)</td>
<td>😊😊😊</td>
<td></td>
</tr>
<tr>
<td>Grade IV (k=0.13)</td>
<td>😊</td>
<td></td>
</tr>
<tr>
<td>Grade Unk (k=0.32)</td>
<td>😊😊😊😊😊</td>
<td></td>
</tr>
</tbody>
</table>

Missing responses are excluded.
Stage at diagnosis is SEER summary stage 1977 in 2003 re-abstraction.
k = kappa statistic.
% Agree = percent agreement
Sn = sensitivity
😊 Good quality or better based on kappa >= 0.60 and each of percent agreement and sensitivity >= 80% (Cicchetti, 2001).
Grade agreement, #2

Summary of n-by-n tables

<table>
<thead>
<tr>
<th>Cancer</th>
<th>% unknown grade re-abstracted to grade</th>
<th>% grade re-abstracted to unknown grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>In situ or localized female breast</td>
<td>21%</td>
<td>9%</td>
</tr>
<tr>
<td>Localized prostate</td>
<td>67%</td>
<td>3%</td>
</tr>
</tbody>
</table>

- More unknown grade for female breast cancers in re-abstracted data (n=919) than in original data (n=877)
Histology Agreement, PoC Study, DY 1997-1998

In situ or localized female breast cancer (n=3,551)

<table>
<thead>
<tr>
<th>Type</th>
<th>% Agree</th>
<th>Sn (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ductal (k=0.77)</td>
<td>☻</td>
<td>☻</td>
</tr>
<tr>
<td>Lobular (k=0.87)</td>
<td>☻</td>
<td>☻</td>
</tr>
<tr>
<td>Mixed lob/duc (k=0.67)</td>
<td>☻</td>
<td>☻</td>
</tr>
<tr>
<td>Other Adeno (k=0.73)</td>
<td>☻</td>
<td>☻</td>
</tr>
<tr>
<td>Other hist (k=0.60)</td>
<td>☻</td>
<td>☻</td>
</tr>
<tr>
<td>NOS (k=0.34)</td>
<td>☻</td>
<td>☻</td>
</tr>
</tbody>
</table>

Stage at diagnosis is SEER summary stage 1977 in 2003 re-abstraction.

k = kappa statistic
% Agree = percent agreement
Sn = sensitivity

😊 Good quality or better based on kappa >= 0.60 and each of percent agreement and sensitivity >= 80% (Cicchetti, 2001).
Treatment Agreement, PoC Study, DY 1997

In situ or localized female breast cancer (n=3,019)

- Surgery (k=0.11)
- Radiation (k=0.84)
- Chemotherapy (k=0.79)
- Hormone (k=0.47)

% Agree Sn (%)

Regional 3 or 4 colon cancer (n=1,059)

- Surgery (k<0)
- Radiation (k=0.81)
- Chemotherapy (k=0.64)
- BRM (k=0.40)

% Agree Sn (%)

Stage at diagnosis is SEER summary stage 1977 in 2003 re-abstraction.

Chemotherapy for colon cancer excludes missing responses.

k = kappa statistic; k<0 is not significant.

% Agree = percent agreement

Sn = sensitivity

😊 Good quality or better based on kappa >= 0.60 and each of percent agreement and sensitivity >= 80% (Cicchetti, 2001).
Treatment Agreement, #2

Summary of n-by-n tables
- Across these 12 treatment variables, there were no more than 15 unknowns for any variable in the original data.
- Except for colon chemotherapy, all of the unknowns in the original were re-coded in the re-abstraction.
- For 9 of the 12 variables (75%), there were more unknowns in the re-abstracted data (n<=108) than in the original data (n<=15).

Stage at diagnosis is SEER summary stage 1977 in 2003 re-abstraction

k = kappa statistic
% Agree = percent agreement
Sn = sensitivity

Good quality or better based on kappa >= 0.60 and each of percent agreement and sensitivity >= 80% (Cicchetti, 2001).
Discussion, #1

• Results provide a check on quality of data recorded in population-based central cancer registries

• In this analysis, e.g.,
  – Localized stage achieved good quality or better for all three cancers, but other stages did not
  – Information on non-Hispanics whites and blacks was good quality for all three cancers, but not other groups
  – Very few treatment variables met the data quality criteria
Discussion, #2

• The analysis
  – Indicates the quality of variables routinely collected by cancer registries can vary by cancer site and across variables
  – Appears to support the observation that the use of supplemental sources of data (e.g., non-hospital) may be necessary to achieve good quality information on cancer treatment
Discussion, #3

• Limitations
  – Wide variation in state-specific measurements
  – Analysis is one measure of quality in cancer registry data
  – Changes in reporting requirements over time
  – “Gold standard” may not as pure or complete as preferred
  – Analyses are preliminary
Concluding Remarks

• Changes in cancer treatment regimens require assessing quality of cancer registry data
• The data quality analysis in the PoC studies guide efforts to improve the quality of data disseminated by a cancer registry
• Future PoC studies will apply lessons learned from the current analyses
References


