Data Quality Analysis of Prostate Cancer Site Specific Factors in Metropolitan Detroit SEER Data, 2004 - 2012

Jeanne Whitlock, MSLS, CTR
Julie George, MS
Ron Shore, MPH
Fawn D. Vigneau, JD, MPH
Metropolitan Detroit Cancer Surveillance System

- Founding participant in the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute.
  - SEER started in 1973
  - Covers Wayne, Oakland and Macomb counties (Detroit and surrounding counties)
- MDCSS is the designated arm of the Michigan Dept of Community Health for collecting cancer information on residents of Wayne, Oakland and Macomb counties
  - Contributes about 40% of cases to the state cancer registry
- Reports 24-25K new cases annually
- Recent reorganization of operational structure included forming a data quality metrics workgroup. The goal of this workgroup is to integrate research expertise in evaluating and improving registry data quality
Project Background

- Noticed a decrease in 2012 prostate cancer cases. This led us to think about examining prostate cancer data in greater detail.

- Possible explanations for decrease:
  - Reporting gaps: decrease seemed to be across the board for our reporting facilities and labs
  - Changes in guidelines for PSA screening
  - Reported decrease for other SEER registries for prostate cancer cases
MDCSS Prostate Cancer Cases by Year of Diagnosis
2004 - 2013

*2013 - 90% complete
Project Background

- During this process, discussed coding changes in prostate cancer, specifically SSFs and how they interrelate. This led us to focus on coding inconsistencies in SSFs as a good target for analysis and for developing strategies to address any needed changes to the data.

- This project draws on work by Dr. Leon Sun and the quality group at SEER on logical checking algorithms for improving data quality, presented at NCRA in 2012
SSF Data Quality Overview

- Source: Prostate cancer data in SEER*DMS
- Timeframe: Diagnosis years 2004 – 2012
- Total number of prostate cases: 31,685
- Avg annual number of cases: 3,520
Reviewed variables

- **PSA:**
  - Lab Value (SSF1) – test result
  - Level (SSF2) – test interpretation
    - Positive, negative or borderline
    - Physician statement or reference range for the lab conducting the test

- **CS Path Extension (SSF3)**
  - Extent of primary tumor at prostatectomy

- **Gleason Score/Pattern (2010-2012)**
  - Pattern
    - SSF7 from biopsy or TURP
    - SSF9 from prostatectomy
  - Score (can be obtained by adding the primary and secondary pattern)
    - SSF8 from biopsy or TURP
    - SSF10 from prostatectomy
Methods

- For each variable, preliminary analysis of unknown/missing codes to determine categories needing further review.
- Based on number of affected cases, we reviewed all or a sample of cases coded to unknown or missing values from each year.
- Quality coordinator reviewed text and other available information in identified records to determine where the coding issues occurred.
- Coding issues were then analyzed to determine reasons for unknown/missing codes and whether reasons for missing data might be explained by data in other related variables.
PSA: Value (SSF1) and Level (SSF2)

- Definitions
  - PSA value (SSF1) is the numerical value from the PSA test
  - PSA level (SSF2) is the test interpretation: whether PSA is elevated, within normal limits or equivocal.

- Missing data evaluation
  - Codes defined as “missing”
    - 997 Test ordered, results not in chart
    - 998 Test not done (test not ordered/not performed)
    - 999 Unknown, no information, not documented in pt record
**PSA Value (SSF1) vs Level (SSF2): Identifying possible cases for review**

<table>
<thead>
<tr>
<th>Scenario</th>
<th>n</th>
<th>Percent</th>
<th>Need to Review?</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSF1 coded, SSF2 coded</td>
<td>23,764</td>
<td>75</td>
<td>No</td>
</tr>
<tr>
<td>SSF1 missing, SSF2 missing</td>
<td>4,753</td>
<td>15</td>
<td>No</td>
</tr>
<tr>
<td>SSF1 coded, SSF2 missing</td>
<td>1,036</td>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>SSF1 missing, SSF2 coded</td>
<td>2,257</td>
<td>7</td>
<td>Yes</td>
</tr>
</tbody>
</table>
PSA Value vs. Level: Preliminary analysis

<table>
<thead>
<tr>
<th>SSF1 code</th>
<th>n</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>997</td>
<td>181</td>
<td>8</td>
</tr>
<tr>
<td>998</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>999</td>
<td>2076</td>
<td>92</td>
</tr>
</tbody>
</table>

- Obvious logical inconsistency: if SSF2 was coded as known (positive, negative or equivocal) then the test was clearly ordered, so SSF1 would be better coded to 997 (test ordered, results not in record) rather than 999 (unknown or no documentation; not documented in patient record).

- We decided to sample 91 cases based on approximately 10 cases per year.
# PSA Value vs Level

## Results

- We sampled 91 cases to review

<table>
<thead>
<tr>
<th>SSF1 code</th>
<th>Number reviewed</th>
<th>PSA value found?</th>
<th>Coding issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>997</td>
<td>7</td>
<td>2</td>
<td>Coded appropriately for the most part</td>
</tr>
<tr>
<td>999</td>
<td>84</td>
<td>23 (27.4%)</td>
<td>Incomplete consolidation of later incoming records</td>
</tr>
</tbody>
</table>

If no PSA value found, SSF1 code should be changed to 997

Explore need for an edit to prevent coding SSF1 999 if SSF2 is coded to a known value (010, 020, 030)
PSA Value vs Level
Resolution and Implications

- **Existing data**
  - Quality project established to review SSF1=999, SSF2 coded cases starting with 2012 cases and working back.
    - 2012: 200 cases reviewed, values found for 20 cases

- **Training**
  - Review general and site-specific instructions in CS Manual part 1 section 2 (Lab tests and Tumor Markers)
  - Emphasize difference between “unknown” codes
  - Emphasize importance of thoroughly consolidating records from other facilities
    - SEER*DMS has a compare difference feature

- **Edits**
  - Consider an edit requiring a value or 997 for SSF1 if SSF2 is coded to 010, 020 or 030.

- **Future plans**
  - Re-visit this question in a year to assess 2013
CS Path Extension (SSF3)

- Unknown/missing values defined as codes 950 or greater
- Total number of cases coded to 950 or greater: 21,284 (67%)
- Avg annual number of cases=2,365
- Found that approach to analysis depended on the specific code; some required analysis of surgery code or reporting source to help determine trends and possible corrective action.
CS Path Extension (SSF3)  
Related variables – surgery code

<table>
<thead>
<tr>
<th>SSF3 code</th>
<th>Definition</th>
<th>Expected Surg03 codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>000, 200-750</td>
<td>Extension codes</td>
<td>30 with neg margins, 50, 70, 80</td>
</tr>
<tr>
<td>950</td>
<td>No evidence of primary tumor</td>
<td>30 with neg margins, 50, 70, 80</td>
</tr>
<tr>
<td>960</td>
<td>Unknown if prostatectomy done</td>
<td>00 (autobuilt cases), 99</td>
</tr>
<tr>
<td>970</td>
<td>No prostatectomy done within 1st course of tx</td>
<td>00, 10-26 (TURP, cryosurgery etc)</td>
</tr>
<tr>
<td>980</td>
<td>Prostatectomy done, not 1st course</td>
<td>00, 10-26 (TURP, cryosurgery etc)</td>
</tr>
<tr>
<td>990</td>
<td>Prostatectomy done, extension not stated, can’t be assessed, not documented</td>
<td>30 with neg margins, 50, 70, 80</td>
</tr>
</tbody>
</table>
Related variables – reporting source

- Reporting source refers to the source of information used to abstract the case.
- Can help explain missing values for lab only or physician only-reported cases.

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hospital IP/Clinic</td>
</tr>
<tr>
<td>2</td>
<td>Rad tx or med onc center</td>
</tr>
<tr>
<td>3</td>
<td>Lab only</td>
</tr>
<tr>
<td>4</td>
<td>Physician</td>
</tr>
<tr>
<td>5</td>
<td>Nursing home/hospice</td>
</tr>
<tr>
<td>6</td>
<td>Autopsy only</td>
</tr>
<tr>
<td>7</td>
<td>Death cert only</td>
</tr>
<tr>
<td>8</td>
<td>Other hospital OP units</td>
</tr>
</tbody>
</table>
## CS Path Extension (SSF3): Preliminary analysis of “missing” codes

<table>
<thead>
<tr>
<th>Code</th>
<th>Code Definition</th>
<th>n</th>
<th>%</th>
<th>Review decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>950</td>
<td>No evidence of primary tumor</td>
<td>23</td>
<td>0.1</td>
<td>Review all</td>
</tr>
<tr>
<td>960</td>
<td>Unknown if prostatectomy done</td>
<td>1521</td>
<td>7.1</td>
<td>Sample. Review surgery code and reporting source</td>
</tr>
<tr>
<td>970</td>
<td>No prostatectomy in 1&lt;sup&gt;st&lt;/sup&gt; course of treatment</td>
<td>19538</td>
<td>91.8</td>
<td>Verify surg code otherwise no review needed</td>
</tr>
<tr>
<td>980</td>
<td>Prostatectomy but not 1&lt;sup&gt;st&lt;/sup&gt; course</td>
<td>106</td>
<td>0.5</td>
<td>Review all</td>
</tr>
<tr>
<td>985</td>
<td>Autopsy, extension unknown</td>
<td>22</td>
<td>0.1</td>
<td>Review all</td>
</tr>
<tr>
<td>990</td>
<td>Prostatectomy done, extension unknown</td>
<td>74</td>
<td>0.3</td>
<td>Sample</td>
</tr>
</tbody>
</table>
CS Path Extension (SSF3)
Findings: Code 950

• Definition: No evidence of primary tumor
  • Findings
    • 23 cases: all reviewed
    • 22 cases coded correctly; no residual tumor on prostatectomy
    • Most likely all of tumor removed at biopsy
  • Resolution
    • No issue to resolve
CS Path Extension (SSF3) Findings: Code 960

- Definition: unknown if prostatectomy done
- 1521 cases
- Step 1: Review surgery code for concordance

<table>
<thead>
<tr>
<th>Surgery code</th>
<th>Number of cases (%)</th>
<th>Further review needed?</th>
</tr>
</thead>
<tbody>
<tr>
<td>99 (unknown)</td>
<td>162 (11%)</td>
<td>No</td>
</tr>
<tr>
<td>Some code other than prostatectomy</td>
<td>37 (2%)</td>
<td>No, change SSF3 code to 970</td>
</tr>
<tr>
<td>00</td>
<td>1322 (87%)</td>
<td>Review for reporting source</td>
</tr>
</tbody>
</table>
CS Path Extension (SSF3)

Findings: Code 960

- Step 2: Determine cases for review using reporting source where surgery code = 00 cases (n=1322 cases)

<table>
<thead>
<tr>
<th>Reporting source</th>
<th>Number</th>
<th>Percent</th>
<th>Further review needed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lab only</td>
<td>912</td>
<td>69%</td>
<td>None-default value = 00 in autobuild</td>
</tr>
<tr>
<td>Hospital/clinic</td>
<td>304</td>
<td>23%</td>
<td>Sample</td>
</tr>
<tr>
<td>Physician report</td>
<td>106</td>
<td>8%</td>
<td>Sample</td>
</tr>
<tr>
<td>Rad onc/med onc only</td>
<td>4</td>
<td>0.3%</td>
<td>Review all</td>
</tr>
</tbody>
</table>
**CS Path Extension (SSF3)**

**Findings: Code 960**

- We sampled 29 of 404 cases for review: 25 cases where reporting source is hospital or physician and all 4 radiation oncology only cases.

<table>
<thead>
<tr>
<th>Identified issue</th>
<th>Suggested remedy</th>
<th>Number affected</th>
<th>% of sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autobuild from radiation only record default SSF3 = 960</td>
<td>Change SSF3 default for these record types records to 970</td>
<td>9</td>
<td>31%</td>
</tr>
<tr>
<td>Coding or consolidation errors</td>
<td>Recode SSF3 to 970</td>
<td>16</td>
<td>55%</td>
</tr>
<tr>
<td>Incorrect reporting source</td>
<td>Change reporting source to lab only</td>
<td>3</td>
<td>10%</td>
</tr>
<tr>
<td>Physician report with no treatment information</td>
<td>No change</td>
<td>1</td>
<td>3%</td>
</tr>
</tbody>
</table>
CS Path Extension (SSF3)
Findings: Code 970 – no prostatectomy done in first course treatment

<table>
<thead>
<tr>
<th>Surgery code</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>17325</td>
<td>88.7</td>
</tr>
<tr>
<td>99</td>
<td>6</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>&lt; 50 check less than 30</td>
<td>2203</td>
<td>11.3</td>
</tr>
<tr>
<td>00 (prostatectomy but not 1\textsuperscript{st} course)</td>
<td>4</td>
<td>&lt;0.1</td>
</tr>
</tbody>
</table>
CS Path Extension (SSF3)
Findings: Code 970 – no prostatectomy done in first course treatment

- Checked surgery code to verify concordance
  - Acceptable surgery codes:
    - 30 – Simple prostatectomy with negative margins
    - 50 – Radical prostatectomy/total prostatectomy
    - 70 – Prostatectomy with resection in continuity with other organs
    - 80 – Prostatectomy, NOS

- Minor coding/consolidation issues in 10 of 19538 cases
CS Path Extension (SSF3)
Findings: Codes 980 and 985

- Code 980: Prostatectomy but not 1st course of treatment
  - 106 cases: all reviewed, no issues
- Code 985: Autopsy, extension unknown
  - 22 cases: all reviewed, no issues
CS Path Extension (SSF3) Findings: Code 990

- Definition: Prostatectomy done, but extension is unknown
- 74 cases: 25 sampled
  - 19 cases (76%) needed CS extension recoded
  - 6 cases (24%) were coded correctly
## CS Path Extension (SSF3):
### Summary

<table>
<thead>
<tr>
<th>Code</th>
<th>Code Definition</th>
<th>n</th>
<th>%</th>
<th>Review decision</th>
<th>Issues?</th>
</tr>
</thead>
<tbody>
<tr>
<td>950</td>
<td>No evidence of primary tumor</td>
<td>23</td>
<td>0.1</td>
<td>Review all</td>
<td>No issues</td>
</tr>
<tr>
<td>960</td>
<td>Unknown if prostatectomy done</td>
<td>1521</td>
<td>7.1</td>
<td>Sample. Review surgery code and reporting source</td>
<td>Coding/consolidation issues.</td>
</tr>
<tr>
<td>970</td>
<td>No prostatectomy in 1\textsuperscript{st} course of treatment</td>
<td>19538</td>
<td>91.8</td>
<td>Verify surg code otherwise no review needed</td>
<td>No issues</td>
</tr>
<tr>
<td>980</td>
<td>Prostatectomy but not 1\textsuperscript{st} course</td>
<td>106</td>
<td>0.5</td>
<td>Review all</td>
<td>No issues</td>
</tr>
<tr>
<td>985</td>
<td>Autopsy, extension unknown</td>
<td>22</td>
<td>0.1</td>
<td>Review all</td>
<td>No issues</td>
</tr>
<tr>
<td>990</td>
<td>Prostatectomy done, extension unknown</td>
<td>74</td>
<td>0.3</td>
<td>Sample</td>
<td>Coding/consolidation issues.</td>
</tr>
</tbody>
</table>
CS Path Extension (SSF3) Resolution and Implications

- **Existing data**
  - Code 960 and surgery = 00 are the default codes for cases autobuilt from pathology casefinding records. Are these the appropriate defaults? This will require discussion with registry leadership and SEER*DMS vendor.
  - Reporting source 3, lab only—does this get changed if more definitive source record comes in? This is a consolidation issue that will need more discussion. Currently this is a routine quality project to manually update.

- **Training**
  - Emphasize site-specific factor coding instructions and meanings of the various “unknown” codes
  - Importance of thorough consolidation

- **Edits**
  - Consider developing edits to check discrepancies between surgery codes and SSF3

- **Future plans**
  - Compare 2013 results for problematic codes
Gleason Pattern and Score (SSF7-10)

- Data items for these measures changed from CSv1 to CSv2. We limited our analysis to diagnosis year 2010 forward to ensure that we were looking at CSv2 coded cases.
  - Prior years have cases coded in either version depending on when they were abstracted.
- CSv2: Two sources of Gleason pattern and score
  - Gleason pattern reported as primary and secondary; these can be added together to obtain the Gleason score
    - Example: Gleason 3+4=7 where 3 and 4 are the patterns, and 7 is the score.
    - Coded in SSF7 or SSF9 as 034
    - Coded in SSF8 or SSF10 as 007
  - Focused on Gleason score rather than pattern
  - Biopsy/TURP: Gleason pattern (SSF7) and Gleason score (SSF8)
  - Prostatectomy/Autopsy: Gleason pattern (SSF9) and Gleason score (SSF10)
Gleason Score: SSF8 and SSF10

- 10618 prostate cancer cases 2010-2012
- Unknown/missing codes
  - 998: Test not done
    - SSF8: Needle biopsy or TURP not done
    - SSF10: Prostatectomy not done
  - 999: Unknown or no information; not documented in pt record

<table>
<thead>
<tr>
<th>Data item</th>
<th>Frequency missing</th>
<th>% missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSF8</td>
<td>588</td>
<td>5.5</td>
</tr>
<tr>
<td>SSF10</td>
<td>7061</td>
<td>66.5</td>
</tr>
</tbody>
</table>
SSF8 Gleason score from biopsy or TURP

Initial review: Sampling method

- 588 cases coded to test not done (998) or unknown (999)
- Cases fitting four diagnosis scenarios were selected for review

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Cases for review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biopsy</td>
<td>32</td>
</tr>
<tr>
<td>No biopsy, TURP done</td>
<td>15</td>
</tr>
<tr>
<td>No biopsy, no TURP, prostatectomy done</td>
<td>9</td>
</tr>
<tr>
<td>No biopsy, no TURP, no prostatectomy</td>
<td>5</td>
</tr>
</tbody>
</table>
### SSF8 Gleason score from biopsy or TURP

#### Initial Review: Results

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Number reviewed</th>
<th>Number (%) requiring change in SSF8</th>
<th>Number (%) requiring change in other variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biopsy</td>
<td>32</td>
<td>16 (50%)</td>
<td>2 (6%)</td>
</tr>
<tr>
<td><strong>Coding issue</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incomplete consolidation</td>
<td></td>
<td></td>
<td>Incorrect code for dx procedure or dx confirmation</td>
</tr>
<tr>
<td>No biopsy, TURP done</td>
<td>15</td>
<td>7 (47%)</td>
<td>1 (7%)</td>
</tr>
<tr>
<td><strong>Coding issue</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incomplete consolidation or code from path report missed</td>
<td></td>
<td></td>
<td>Incorrect surgery code</td>
</tr>
<tr>
<td>No biopsy, prostatectomy done</td>
<td>9</td>
<td>0</td>
<td>3 (33%)</td>
</tr>
<tr>
<td><strong>Coding issue</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Many of cases in this category are incidental dx at cystoprostatectomy for bladder cancer and don’t need review for missing SSF8</td>
<td></td>
<td></td>
<td>Incorrect surgery code-these were actually suprapubic prostatectomies for BPH</td>
</tr>
<tr>
<td>No biopsy, no TURP, no prostatectomy</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Coding issue</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 of these were cryosurgery with no path specimen and no documented biopsy</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SSF8 Gleason Score from biopsy or TURP

Review considerations

- Several variables need consideration when analyzing this data item to identify cases needing review including:
  - **Histology**
    - Gleason score expected for adenocarcinoma only
  - **Diagnostic Confirmation**
    - Can eliminate cases not confirmed by primary tissue
  - **Diagnostic procedure and surgery of primary site:**
    - If there are acceptable codes in Diagnostic Procedure (i.e., biopsy of primary site) **OR** Surgery (i.e., TURP codes) then missing Gleason Score is problematic.
    - If Diagnostic Procedure is coded to 00 and Surgery code of 70 then Gleason score should be 998 (not done) – these are usually incidentally-found cases.
    - If Diagnostic Procedure is coded to 00 and Surgery code is 50 then these should be investigated to determine if the code for diagnostic/staging procedure was missed or if the surgery code needs to be corrected (i.e., to 70) Radical prostatectomies are typically not done without a prostate biopsy.

- Complex analysis lends itself to a decision-tree type tool to select cases for inclusion
SSF8 Gleason score from biopsy or TURP
Suggested Review Strategy

Total cases missing SSF8
588 (5.5% of total)

Review histology

Adenocarcinoma
479 cases
(81.5%)

Non-Adenocarcinoma
109 cases
16 reviewed to confirm histology

Primary Tissue
392 (81.5%)

Not primary Tissue
87 (18.2%): no review

Select based on dx procedure and surgery codes

Biopsy
265 (68%)

No biopsy, TURP
15 (4%)

No biopsy, prostatectomy
101 (26%)

No biopsy, no TURP, no prostatectomy
11 (3%)

Review dx confirmation (Detroit-specific data item has codes for primary or mets tissue)
SSF10 Gleason Score from Prostatectomy
Selecting cases for review

- 7,061 cases coded to 998 (test not done) or 999 (unknown or no information)

<table>
<thead>
<tr>
<th>Surgery code</th>
<th>Number</th>
<th>Percent</th>
<th>Further review needed?</th>
</tr>
</thead>
<tbody>
<tr>
<td>No surgery or unknown surgery</td>
<td>6249</td>
<td>88%</td>
<td>No (all were coded to 998, no prostatectomy)</td>
</tr>
<tr>
<td>Surgery other than prostatectomy</td>
<td>706</td>
<td>10%</td>
<td>No (all were coded to 998, no prostatectomy)</td>
</tr>
<tr>
<td>Prostatectomy</td>
<td>106</td>
<td>2%</td>
<td>Yes</td>
</tr>
</tbody>
</table>
SSF10 Gleason Score from Prostatectomy

Findings

SSF10 Coding Issues

n=25

- Data available incomplete consolidation (16%)
- No information, no path report available (56%)
- Neoadjuvant hormone therapy (16%)
- No residual on prostatectomy (12%)
Gleason Score: SSF8 and SSF10
Resolution and Implications

- **Existing data**
  - SSF10 minor issue (2% of cases in database). Most missing cases could be resolved through text review or path report review where available.
  - SSF8 requires multi-tiered review
    - Source in 2 different sets of data items
    - Worthwhile to limit review to adenocarcinomas.
    - Incorrect coding in diagnostic/staging procedure, surgery and diagnostic confirmation will confound results.

- **Training**
  - Emphasize site-specific factor coding instructions and meanings of the various “missing” codes
  - Emphasize importance of coding diagnostic and surgical procedures correctly.
  - Importance of thorough consolidation

- **Edits**
  - Consider developing edits to check discrepancies between Gleason score coded 998 when dx/staging procedure – TURP or prostatectomy done.

- **Future plans**
  - Compare 2013 results using developed review strategy
In conclusion

• Don’t be afraid to take a deeper dive into the data! What you find might surprise you—but will lead to higher quality data.