Evaluation of NAACCR Survival Data
June 2012

Hannah K Weir, PhD
Division of Cancer Prevention and Control
Centers for Disease Prevention and Control
Atlanta, GA

Chris J Johnson, MS
Cancer Data Registry of Idaho
Boise, ID

And the NAACCR Survival Analysis Workgroup (SAWG)
### NAACCR Survival Analysis Workgroup Members

<table>
<thead>
<tr>
<th>Name</th>
<th>State, Province or Agency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deb Hurley</td>
<td>SC (co-chair)</td>
</tr>
<tr>
<td>Chris Johnson</td>
<td>ID (co-chair)</td>
</tr>
<tr>
<td>Glenn Copeland</td>
<td>MI</td>
</tr>
<tr>
<td>Larry Ellison</td>
<td>Stat Cam</td>
</tr>
<tr>
<td>Monique N. Hernandez, Ph.D.</td>
<td>FL</td>
</tr>
<tr>
<td>Bin Huang</td>
<td>KY</td>
</tr>
<tr>
<td>Angela Mariotto</td>
<td>NCI</td>
</tr>
<tr>
<td>Zoran Miladinovic</td>
<td>Stat Can</td>
</tr>
<tr>
<td>Cyllene Morris</td>
<td>CA</td>
</tr>
<tr>
<td>Xiaoling Niu</td>
<td>NJ</td>
</tr>
<tr>
<td>Arti Parikh-Patel</td>
<td>CA</td>
</tr>
<tr>
<td>Paulo S. Pinheiro, MD PhD</td>
<td>NV</td>
</tr>
<tr>
<td>Trevor Thompson</td>
<td>CDC</td>
</tr>
<tr>
<td>Donna Turner</td>
<td>MB</td>
</tr>
<tr>
<td>Baozhen Qiao</td>
<td>NY</td>
</tr>
<tr>
<td>Zhenguo Qiu</td>
<td>AB</td>
</tr>
<tr>
<td>Kevin Ward</td>
<td>GA</td>
</tr>
<tr>
<td>Hannah Weir</td>
<td>CDC</td>
</tr>
<tr>
<td>Reda Wilson</td>
<td>CDC</td>
</tr>
<tr>
<td>Brad Wohler</td>
<td>FL</td>
</tr>
<tr>
<td>Kevin Zhang</td>
<td>MACRO</td>
</tr>
</tbody>
</table>
Overview

• What is population-based survival and how is it used?
• Data evaluation
• Putting it all together
• Next steps
What is Population-Based Survival

• Measures survival *achieved* in the population regardless of age, race, stage of disease, access to health care, etc.

• Can be used to:
  • Target and monitor cancer control and health policy initiatives
  • Evaluate the effectiveness of healthcare delivery (measure of cancer system performance)
Advantages and Disadvantage of Relative vs. Cause Specific Survival

<table>
<thead>
<tr>
<th></th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Relative</strong></td>
<td>Relies on fact of death not cause of death</td>
<td>Life tables may not be available for all populations</td>
</tr>
<tr>
<td><strong>Cause Specific</strong></td>
<td>Not limited to populations with life tables</td>
<td>Death Certificates may not be reliable (e.g., may be coded to site of mets or recurrence)</td>
</tr>
</tbody>
</table>
Overview

- What is population-based survival and how is it used?
- Data evaluation
- Putting it all together
- Next steps
Data

– First year requested follow-up data
– Excluded Canadian data due to coding of vital status variable
– Registries
  • SEER: CA (LA, SF), Detroit, HI, IA, KY, LA, NJ, NM, UT, Seattle
  • NPCR: remaining states
    – 2 NPCR state cancer registries not included
Data Elements

- Patient Demographics
  - date of birth
  - sex
  - race/ethnicity
  - name
  - SS#

- Tumor Record
  - site
  - histology
  - behavior
  - stage
  - date of diagnosis
  - type of reporting source

Incidence

Follow-Up
- date of last follow-up
- vital status
- cause of death
- follow-up source central

Alive

Death
Evaluation Criteria

• % Sex, Age or Race Unknown
• % DCO/AO
• % Vital status Unknown
• % Edi Errors
• % MV

• % Missing Cause of Death
• % Alive with 0 Survival Time
• % Death within 1 Month of Diagnosis
The validity of population-based survival comparisons is clearly dependent on the validity of the incidence data. Berrino, 2003
Factors that Impact Incidence

• NAACCR Certification
  – Completeness of case ascertainment
  – DCO/ autopsy
  – Missing critical information (age, sex, race)
  – Edits
  – Duplicates
Gold and Silver Level Certification Status of NAACCR U.S. Cancer Registries for 2008 Data

http://www.naaccr.org/Certification/WhoisCertified.aspx
Factors that Impact Incidence

• NAACCR Certification
  – Completeness of case ascertainment
  – DCO/ autopsy
  – Missing critical information (age, sex, race)
  – Edits
  – Duplicates

• Population Coverage
  – 1995 - 19 US registries NAACCR Certified
  – 2008 - 53 US registries Certified
Factors that Impact Incidence

• NAACCR Certification
• Completeness of Case Ascertainment
  – Clinical vs. Microscopically Verified (%MV)
% Type Diagnostic Confirmation
SEER (1992-2008)
% Type Diagnostic Confirmation
NPCR (1995-2008)

- MV
- Clinical
- Other
- Unk
Follow-Up

Incidence

- Alive
- Death
Demographic Variables

• Variable: Name (last, first), Sex, Date of birth, Social Security No (SS#)
• Critical for enhancing race/ethnicity, follow-up information through linkage
• Results from Melissa Jim – IHS linkage project
### % Missing - Linkage Variables

<table>
<thead>
<tr>
<th></th>
<th>SS#</th>
<th>Birth Date</th>
<th>Sex</th>
<th>Last Name</th>
<th>First Name</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SEER - range</strong></td>
<td>0.00-3.93</td>
<td>0.00-0.09</td>
<td>0.00-0.02</td>
<td>0.00-&lt;0.00</td>
<td>0.00-&lt;0.00</td>
</tr>
<tr>
<td>- No. states w/missing</td>
<td>7/10</td>
<td>4/10</td>
<td>3/10</td>
<td>1/10</td>
<td>3/10</td>
</tr>
<tr>
<td><strong>NPCR - range</strong></td>
<td>0.00-2.58</td>
<td>0.00-0.07</td>
<td>0.00-0.03</td>
<td>0.00-&lt;0.02</td>
<td>0.00-&lt;0.02</td>
</tr>
<tr>
<td>- No. states w/missing</td>
<td>30/41</td>
<td>21/41</td>
<td>22/41</td>
<td>9/10</td>
<td>15/41</td>
</tr>
</tbody>
</table>

Source: M Jim, IHS linkage data, variable years of diagnosis
Follow-Up Variables: Inter-Field and Intra-Record Edits

Data Variables and Edits

- Date of last contact
- Vital status
- Cause of death
- ICD revision number
- Follow-up source central
- Types of reporting source

- All NPCR and SEER registries reported <1% edit errors for any individual edits

Edits associated with vital status variables needed for survival analysis
- Age, Histologic Type, COD, ICDO3 (SEER IF43)
- Cause of Death (SEER COD)
- Date of Last Contact (NAACCR DATEEDIT)
- Date of Last Contact Flag (NAACCR)
- Date of Last Contact, Date Flag (NAACCR)
- Date of Last Contact, Date of Diag. (NAACCR IF19)
- Follow-Up Source (COC)
- Follow-Up Source Central (NAACCR)
- Follow-Up Source Central, Vital Status (NPCR)
- Follow-Up Source, Vital Status (COC)
- ICD Revision Number (NPCR)
- ICD Revision Number, Cause of Death (SEER IF37)
- ICD Revision, Vital Stat, Date Last Contact (NPCR)

Type of Report Srce (DC/AO), Date of Dx (SEER IF02)
Type of Report Srce (DC/AO), COD(SEER IF09)
Type of Report Srce (DC/AO), Diag Conf (SEER IF05)
Type of Report Srce (DC/AO), Vital Stat (SEER IF08)
Type of Reporting Source (SEER RPRTSRC)
Vital Status (Subm)
Vital Status, Cause of Death (Subm)

Verify cause of death same on all records for a patient (SEER IR11)
Verify date of follow-up same on all records for a patient (SEER IR08)
Verify vital status same on all records for a patient (SEER IR10)
Follow-Up Requirements

**Alive Status**
- SEER Program requires all SEER registries to follow alive patients
  - 95% patients have last contact date within 18 months of the annual date of submission
- NPCR registries are not required to follow patients

**Death Status**
- All registries conduct death clearance with state DC
- SEER and NPCR provide support for registries to link with the National Death Index and the Social Security Death Index
Events in Follow-Up

Incidence

Alive

Death
The Importance of Death Ascertainment


**OBJECTIVE:** to measure the impact of variation in patient follow-up on survival statistics.

**METHODS:** SEER data used to construct datasets simulated scenarios of complete (SEER), incomplete, and no follow-up (NPCR) of alive patients; and complete and incomplete death ascertainment.

**CONCLUSIONS:**

- Complete death ascertainment important for producing accurate cancer survival statistics, and
- Ascertainment of deaths only should generally be sufficient for survival analysis.
Full Dates vs. Partial Dates

- Date of Birth
- Date of diagnosis
- Date of last contact

Age at diagnosis needed for Life Tables
Survival interval

SEER Program uses month and year
Example: Patient diagnosed April 2000 and dies May 2000. Survival interval could be 1 – 60 days

NAACCR / NPCR uses month, day and year
Survival Interval

Full Dates vs. Partial Dates

Full dates (day, month, year) should be used in population-based cancer survival studies.
Day of Diagnosis (2004-2008)
Day of Death among Decedents (2004-2008)

![Graph showing the distribution of days of death among decedents from 2004 to 2008, comparing SEER and NPCR data.](image-url)

- **SEER**
- **NPCR**

Legend:
- **valid**
- **15**
- **1**
- **30/31**
- **Blank**

NAACCR Logo
Cause of Death among Decedents
SEER 1995-2008

- Unknown/missing/invalid COD
- State DC not available or state DC available but no COD
- Non cancer death
- In situ, benign or unknown behavior neoplasm
Cause of Death among Decedents
NPCR 1995-2008

- Unk
- DC not available or DC available but no COD
- Non Cancer
- In situ, benign or unknown behavior neoplasm
- All Malignant
Events in Follow-Up

- Incidence
  - Alive
  - Death
Follow-Up Requirements

Alive Status
• SEER Program requires all SEER registries to follow alive patients
  - 95% patients have last contact date within 18 months of the annual date of submission
• NPCR registries are not required to follow patients

Death Status
• All Registries conduct death clearance with state DC
• SEER and NPCR provide support for registries to link with the National Death Index and the Social Security Death Index
Follow-Up Requirements

Alive Status
- SEER Program requires all SEER registries to follow alive patients
  - 95% patients have last contact date within 18 months of the annual date of submission
- NPCR registries are not required to follow patients
  - impute follow-up date to be the end of study (e.g., 12/31/08)

Death Status
- All Registries conduct death clearance with state DC
- SEER and NPCR provide support for registries to link with the National Death Index and the Social Security Death Index
Overview

• What is population-based survival and how is it used?

• Data evaluation

• Putting it all together

• Next steps

Female Breast Cancer

[Graph showing survival rates for SEER and NPCR datasets, with a focus on the survival rates at 60% and 90% marks.]
Overview

• What is population-based survival and how is it used?
• Data evaluation
• Putting it all together
• Next steps
SAWG - Next Steps

- Deceased with 0 survival time (and not a DCO/AO case)
  - E.g., Physician only reporting source, follow up source central (State or NDI). These events are included in analysis whereas DCO/AO cases are excluded

- Immortal cases
- Survival using full dates - SEER*Stat enhancement
- State specific life tables – available in 2012
- NAACCR Webinar June 14, 2012
- CONCORD Study
The findings and conclusions in this presentation are those of the presenter and do not necessarily represent the official position of the Centers for Disease Control and Prevention.