Introduction to SEER*Stat: Frequencies

Diane Nishri
Senior Research Associate, Surveillance
February, 2011
SEER

• “The Surveillance, Epidemiology, and End Results (SEER) Program is the most authoritative source of information on cancer incidence and survival in the United States”

• Case ascertainment for SEER began on January 1, 1973, in the states of Connecticut, Iowa, New Mexico, Utah, and Hawaii and the metropolitan areas of Detroit and San Francisco-Oakland

• “Considered the gold standard for cancer registries around the world”
SEER*Stat

• Software created and maintained by the US National Cancer Institute

• Produces statistics useful for studying the impact of cancer on a population. Sessions available are:
  ▪ Frequencies
  ▪ Rates
  ▪ Survival
  ▪ Limited-duration prevalence

• Designed to work with SEER data, but adaptable to any cancer registry anywhere
Three basic steps:

- Define the parameters of your analysis in a SEER*Stat session
- Execute the session, and review the results in the results matrix window
- Use the results by printing, exporting or copying them
SEER*Stat Sessions

- Specify the statistics to calculate
  - Select the appropriate session
  - Choose statistics & set parameters (Statistic)

- Define the records to be used
  - Select a database (Data)
  - Make selections if a subset of records is desired (Selection)

- Define the layout for output
  - Stratify results by one or more variables (Table)
  - Define groupings of variables in the Dictionary
  - Provide titles & other output options (Output)
Frequencies: Start at the beginning

- One-way frequencies for all categorical variables
  - Invalid or missing values?
  - Unusual distribution?

- Cross tabulations for 2 or more variables to check consistency
  - Site recode by sex
  - Site recode by morphology by age group
Frequency example

- Create a table showing the frequency of Renfrew County and District HU cases by age group and sex for 2003-2007
- Age groups:
  - 0-14, 15-44, 45-54, 55-64, 65-74, 75+
Creating user-defined variables

• Open a session in SEER*Stat
• Select the incidence database
• Go to File -> Dictionary
• Choose the SEER*Stat variable needed to create the user-defined variable
• On the right-hand menu, click Create
• Define the groupings needed
• Name the new variable appropriately
• Variables are in the User-Defined folder
Frequency Session & Data Tab
File -> Dictionary


- Age at Diagnosis
  - Age recode
- Sex, Year Dx, Geography
- Site and Morphology
- Race and Age (case data only)
- Dates
- Other
Original variable definition
Create new groupings and labels
How to share User-Defined variables

- File -> Dictionary… -> Export… or Import…
- A dictionary is specific to a database
Statistic Tab

The Statistic Tab in the SEER*Stat 6.6.2 software interface is shown. It includes options for selecting statistical methods such as Frequencies and Trends, and parameters like P-Values, Trend Variable, PC End Points, and APC Calculation. The interface also allows for the display or exclusion of confidence intervals.
Selection Tab
Select Only the First Matching Record…

• If you check this option, your selection may result in fewer cases than expected

• SEER*Stat will select the first cancer case, regardless of the site, and then select the cases which meet the other selection criteria

• If your database contains all incident cases, and a person has a non-colorectal case before their colorectal case, they would be excluded from a colorectal analysis if this option is checked
Table Tab
Selection Tab vs. Table Tab

• Use the Selection tab to reduce the number of records analyzed based on specific variables.

• Use the Table tab to choose variables to include in the output table.

• A case is excluded from your analysis if it does not pass the selection criteria of the Selection tab and match at least one table definition of the Table tab. In general, it is best to select your case subset on the Selection tab rather than relying on the table definitions to exclude cases.
Output tab

Title (Up to 5 lines)
Cancer Incidence by Age Group and Sex
Renfrew County & District Health Unit, 2003-2007

Number of Decimal Places for Trends: 0.1
Hide Statistics When Fewer Than 6 Cases
Include Individual Year Counts
# Results Matrix

### Frequency Matrix

Renfrew County & District Health Unit cancer incidence by age group and sex, 2003-2007

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Male and female</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>Column %</td>
<td>Cum %</td>
</tr>
<tr>
<td>00-14 years</td>
<td>20</td>
<td>0.66%</td>
<td>0.66%</td>
</tr>
<tr>
<td>15-44 years</td>
<td>164</td>
<td>5.44%</td>
<td>6.11%</td>
</tr>
<tr>
<td>45-54 years</td>
<td>313</td>
<td>10.39%</td>
<td>16.50%</td>
</tr>
<tr>
<td>55-64 years</td>
<td>638</td>
<td>21.18%</td>
<td>37.68%</td>
</tr>
<tr>
<td>65-74 years</td>
<td>669</td>
<td>26.85%</td>
<td>66.53%</td>
</tr>
<tr>
<td>75+ years</td>
<td>1,008</td>
<td>33.47%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Total</td>
<td>3,012</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

^Statistic not displayed due to fewer than 6 cases
“^ Statistic not displayed due to fewer than 6 cases”

• If we had restricted the time period to 2005-2007, there would have been fewer than 6 male cases aged 0-14

• The total would not be reported, and there would be no percentages calculated for males

• Options:
  - More years
  - Different age groups (0-44)
  - Combine sexes
Matrix Menu

SEER*Stat 6.6.2

**Cancer incidence by age group and sex: 2005-2007**

<table>
<thead>
<tr>
<th>Male</th>
<th>Column %</th>
<th>Cum %</th>
<th>Count</th>
<th>Column %</th>
<th>Cum %</th>
</tr>
</thead>
<tbody>
<tr>
<td>00-14</td>
<td>0.50%</td>
<td>0.60%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-44</td>
<td>4.93%</td>
<td>5.43%</td>
<td>36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-54</td>
<td>10.26%</td>
<td>15.69%</td>
<td>73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55-64</td>
<td>21.12%</td>
<td>36.81%</td>
<td>201</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65-74</td>
<td>27.54%</td>
<td>64.34%</td>
<td>232</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75+</td>
<td>35.56%</td>
<td>100.00%</td>
<td>323</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100.00%</td>
<td>100.00%</td>
<td>1823</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statistic not displayed due to fewer than 6 cases.
## Matrix Menu

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrieve Session</td>
<td>Extract the session that created the matrix</td>
</tr>
<tr>
<td>Options</td>
<td>View or edit matrix options</td>
</tr>
<tr>
<td>Export</td>
<td>Export the results to a text file for use in other applications</td>
</tr>
<tr>
<td>Order</td>
<td>Change the order of row/column variables</td>
</tr>
<tr>
<td>Lock</td>
<td>Lock a row/column while scrolling</td>
</tr>
</tbody>
</table>
Matrix Options

- Choices that you make here will affect the display of the matrix on your screen as well as in printed, copied, or exported data.

- Calculation of statistics is not affected.

- If you change the title in Matrix Options and then retrieve the session, the session will contain the original, not the edited title.
Saving your work

- To save the parameters of a session or a results matrix: **File -> Save or Save As.**
- A session file may be saved at any time, even if it has not been executed.
- It is not necessary to save both the session and its results matrix. A saved matrix contains a snapshot of the session at the time of execution, or a session may be run again to recreate the matrix.
- You can also save your session's parameters as a text file in Windows’ INI format: **File -> Export As Text.**
Renfrew County & District health unit cancer incidence by age group and sex, 2003-2007

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Count</th>
<th>Column %</th>
<th>Cum %</th>
<th>Count</th>
<th>Column %</th>
<th>Cum %</th>
</tr>
</thead>
<tbody>
<tr>
<td>00-14 years</td>
<td>20</td>
<td>0.66%</td>
<td>0.66%</td>
<td>10</td>
<td>0.65%</td>
<td>0.65%</td>
</tr>
<tr>
<td>15-44 years</td>
<td>164</td>
<td>5.44%</td>
<td>6.11%</td>
<td>58</td>
<td>3.79%</td>
<td>4.45%</td>
</tr>
<tr>
<td>45-54 years</td>
<td>313</td>
<td>10.39%</td>
<td>16.50%</td>
<td>124</td>
<td>8.11%</td>
<td>12.56%</td>
</tr>
<tr>
<td>55-64 years</td>
<td>638</td>
<td>21.18%</td>
<td>37.68%</td>
<td>332</td>
<td>21.71%</td>
<td>34.27%</td>
</tr>
<tr>
<td>65-74 years</td>
<td>860</td>
<td>28.65%</td>
<td>66.53%</td>
<td>506</td>
<td>33.00%</td>
<td>67.36%</td>
</tr>
<tr>
<td>75+ years</td>
<td>1,008</td>
<td>33.47%</td>
<td>100.00%</td>
<td>499</td>
<td>32.64%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Total</td>
<td>3,012</td>
<td>100.00%</td>
<td>100.00%</td>
<td>1,529</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

Statistic not displayed due to fewer than 6 cases.
ALWAYS copy the Session Information!
Create a graph to display your results

New cancer cases by age group  
Renfrew County & District Health Unit 2003-2007

- 00-14 years: 1%  
- 15-44 years: 5%  
- 45-54 years: 10%  
- 55-64 years: 21%  
- 65-74 years: 29%  
- 75+ years: 34%
Exercise #1: Frequency Session

• For your Health Unit, create a frequency table of cancer incidence by age group and sex. Choose an appropriate time period. If possible, use Excel or other software to graph your results

• BONUS: Repeat with mortality data