**Calculating Zone-based Cancer Rates**

This process describes how to calculate zoned-based cancer incidence rates. It assumes that you have access to individual-level cancer data geocoded to census tracts in SEER\*Stat and that these data are linked with the tract-level population denominators needed for calculating age-adjusted rates.

* For SEER registries, a database with incidence data linked to census tract attributes is available and can be used for this process. The currently available database is: Incidence - SEER Specialized Limited-Field Data, 22 Registries (excl AK and IL), Nov 2021 Sub (2006-2019) <Vintage 2019 Pops by Tract 2010 Geographies>. This database will be referred to as the “Incidence - SEER Specialized Limited-Field Database” throughout this document.
* For non-SEER registries, you can create a local SEER\*Stat database using census tract 2010 geography linked to census tract attributes. For more information, see the topic [“Resources for Calculating Rates for Small Areas and by ABSMs”](https://www.naaccr.org/gis-resources/#1652972825105-a3aa7424-9fb9) under GIS Resources on the NAACCR website.

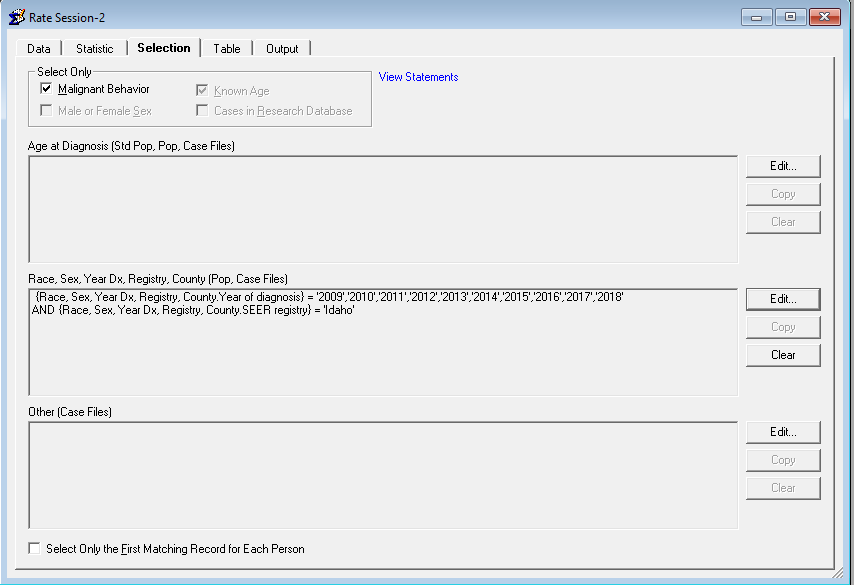
We also assume you have access to SAS. This process is designed to produce SEER\*Stat matrix files that can be used to generate the cancer rate tables needed for the web-based tool for cancer reporting by zones. If desired, Westat can assist with this processing once any necessary data usage agreements are in place.

For this process, we assume files reside in a Windows folder called C:\Work\ZoneRateCalcs. This folder can be located anywhere, including on a network drive. The SEER\*Stat output files contain tables with small cells that will ultimately be suppressed based on your desired suppression rules. Thus, access to the folder where the SEER\*Stat output files are kept should be limited to authorized users.

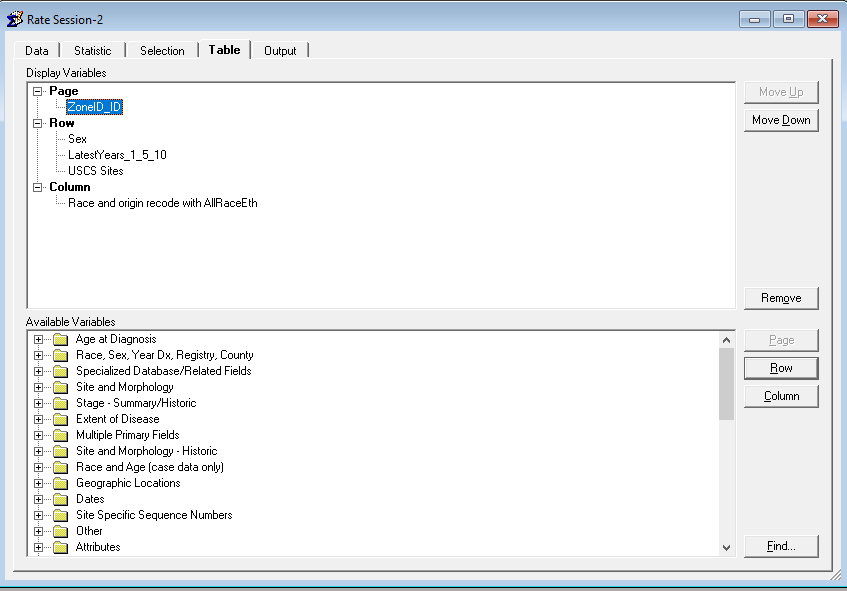
In this process, we refer to variable parameters using square brackets. For example, [StAbbr] refers to the state abbreviation for your registry.

**If you do not have access to the Incidence - SEER Specialized Limited-Field Database specified above and/or your zone definitions are not yet available in the SEER-provided Census Tract Attributes file, you may still be able to calculate zone-based cancer incidence rates. Please see the Appendix for additional steps and requirements.**

1. Define and run a SEER\*Stat rate session to calculate zoned-based cancer rates
   * Create a SEER\*Stat rate session file for zone-level rates named “[RunNum]zone\_RateCalcs.si”. Parameters:
     + Data tab: select the SEER\*Stat database with tract IDs for cancer cases, or the Incidence - SEER Specialized Limited-Field Database if using
     + Create a user-defined variable based on the variable Cancer Reporting Zone in the database.
       - The Cancer Reporting Zone variable is under the section ‘Tract Attributes Geography and Year’
       - Select the Cancer Reporting Zone values applicable to your state/catchment area. Name the new user-defined variable “ZoneID\_[StAbbr]”.
       - Note: the Cancer Reporting Zone values in the SEER\*Stat database for your state/catchment area should match the ZoneIDFull values in the “ZonedTracts\_[RunNum]\_final.xlsx” file Westat provided to your registry as part of the final cancer reporting zone materials. If the values do not match the ZoneIDFull values or if the “ZonedTracts\_[RunNum]\_final.xlsx” file your registry has does not include a ZoneIDFull variable, please contact Westat for an updated file.
     + Import other user variables (.fmx files included with the distribution folder) and modify as needed:
     + LatestYears\_1\_5\_10: defines the groups of years you wish to generate rates for. The sample has 1-year, 5-year, and 10-year groups ending in 2015.
       - Note: if modifying the years defined in the variable, keep the name of the variable as “LatestYears\_1\_5\_10”.
     + USCS\_Sites: the cancer sites to be included. The sample includes the cancer sites reported by USCS.
     + RaceOriginRecode\_wAllRaceEth (variable name: Race and origin recode with AllRaceEth): race and ethnic origin groups. The sample includes non-Hispanic White, non-Hispanic Black, non-Hispanic API, non-Hispanic AIAN, non‑Hispanic unknown race, Hispanic, and a combined grouping of all races and ethnic origins.
     + CTCertainty\_ReportingZone\_Recode (variable name: Census Tr Certainty 2010 for Reporting Zone Analysis recode): categorization of cases into two categories – “High certainty (1,2,6) or low certainty but tract not required for zone” and “All values”.
     + Statistics tab: age-adjusted rates, show standard errors and CIs
     + Selection tab: needed years, registry/state of interest
       - It is generally recommended to limit cases to those with high census tract certainty or with low certainty but the tract is not required for assigning Reporting Zone. To do this, add to the selection under the ‘Other (Case Files)’ section using the ‘Census Tr Certainty 2010 for Reporting Zone Analysis recode variable’ selecting for “High certainty (1,2,6) or low certainty but tract not required for zone”. This selection may not be needed if the registry has very high census tract certainty values and it is important that the total cases for all reporting zones match the total cases for the state as a whole.



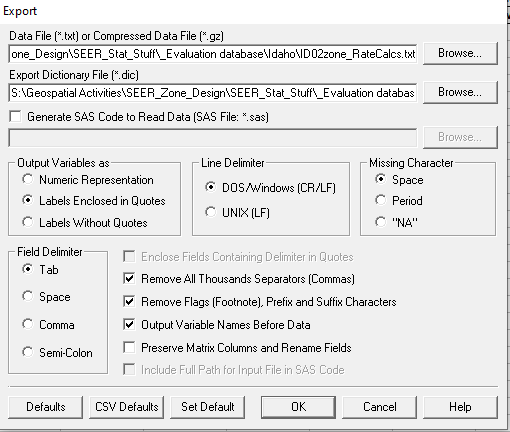
* + - Table tab:
      * Page: ZoneID\_[StAbbr]
      * Row: Sex, LatestYears\_1\_5\_10, USCS sites
      * Column: Race and Origin with AllRaceEth



* + - Output tab: 3 decimal places, don’t hide statistics for small cells. (Note: this enables a summary of cell suppression in to be generated when processing the results. If desired, suppression rules can be implemented here.)



* + Run the rate session and save the resulting matrix file.
  + Export the matrix using the Export Results as Text File dialog with the “Labels Enclosed in Quotes” option selected. The program that builds the tables for the web tool assumes the filename will be “[RunNum]zone\_RateCalcs.txt”. Here is a screenshot of the Export dialog:



1. If desired, define and run a SEER\*Stat rate session to calculate the same cancer rates for the state as a whole. These state-level rates are needed for the web tool.
   * Copy the zone-level rate session and remove the zone field in the Tables tab.

A screenshot of a computer

Description automatically generated

* + Run the rate session and save the resulting matrix file
  + Export the matrix using the Export Results as Text File dialog with the “Labels Enclosed in Quotes” option selected. The program that builds the tables for the web tool assumes the filename will be “[StAbbr]state\_RateCalcs.txt”.

1. If desired, define and run a SEER\*Stat rate session to calculate the same cancer rates by county. These county-level rates are not needed for the web tool, but can be used to add county-level rates as an option.
   * Create a user-defined variable based on the variable State-county in the database.
     + - The State-county variable is under the section ‘Race, Sex, Year Dx, Registry, County’
       - Edit the variable:
         * Keep only the groupings associated with your state/registry.
         * In addition, delete the overall state grouping (e.g., Delaware, California) from the ‘Groupings’ section, as well as the unknown county grouping (if applicable).
         * Name the new user-defined variable “County\_[StAbbr]”.
   * Copy the zone-level rate session and remove the zone field in the Tables tab.
   * Add the County\_[StAbbr] variable under the Page section in the Table tab.

A screenshot of a computer

Description automatically generated

* + Run the rate session and save the resulting matrix file
  + Export the matrix using the Export Results as Text File dialog with the “Labels Enclosed in Quotes” option selected. The program that builds the tables for the web tool assumes the filename will be “[StAbbr]county\_RateCalcs.txt”.

**Appendix**

These procedures are to be used by registries whose zone definitions have not yet been included in either the Incidence - SEER Specialized Limited-Field Database (SEER registries) or in the SEER Census Tract Attributes file (non-SEER registries) but whose definitions have been finalized and provided by Westat. The procedures outline how to create a user-defined zone variable to be used in a local SEER\*Stat database with census tract 2010 geography data. If following these procedures, do these steps first before continuing with the steps above.

1. Create a copy of the ZoneRateCalcs distribution folder in the desired location (we assume this is at “C:\Work\ZoneRateCalcs” for this process). Copy the final zoned tracts and zone list Excel files provided with the zone design results (“ZonedTracts\_[RunNum]\_final.xlsx” and “ZoneList\_[RunNum]\_final.xlsx”) to this folder.
2. In SEER\*Stat, generate a user-variable export file for all of the tracts in the state:
   * In a rate session, select the incidence database with tract geocodes
   * Open the Dictionary by clicking the dictionary icon: 
   * Select the tract field for the SEER\*Stat database (possibly named something like “State-County-Census Tract 1990/2000/2010 Recode”) and click “Create” to create a user variable based on this field
   * Change the name to Tracts\_[StAbbr].
   * Clear all of the entries in the Grouping window by clicking in the window, typing Ctrl-A to select all, and then clicking Delete.
   * In the Values window, select all of the tracts for the state (click on the first tract for the state, then Shift-click on the last. You may need to know the state FIPS code for your state to identify the relevant tracts.
   * Click Add…
   * Select the “Added as separate groupings (one for each value)” and be sure the “Use underlying data values instead of names” option is not checked. Click OK.
   * Click OK to save this as a user variable and return to the Dictionary window
   * Click Export in the Dictionary window
   * Use “Browse…” to navigate to the “C:\Work\ZoneRateCalcs” folder and name the file “Tracts\_[StAbbr].fmx”
   * Select the “Tracts\_[StAbbr]” variable and click OK to create the file
   * Click “Close” to close the Dictionary window. Exit SEER\*Stat.
3. Run the 50\_GenSEERStatVar\_public SAS program to create a SEER\*Stat user variable that defines each zone as a set of tracts
   * Set the %let parameters at the beginning of the code to specify the state abbreviation, run number, and data path if different from “C:\Work\ZoneRateCalcs”
   * Investigate any warning messages generated by the program as these usually mean something is not working as expected
   * The program creates a SEER\*Stat “fmx” file to define a user variable called “ZoneID\_[StAbbr]”