47-Year Limited-Duration Prevalence
U.S. Female Breast Cancer for All Races by Age at Prevalence
ProjPrev Example

Objective

Calculate U.S. 47-year limited-duration prevalence for malignant female breast cancer on January 1, 2022 by age for all races combined using ProjPrev.

Key Points

The population of the SEER 8 registries does not have the same age-specific racial breakdown as the U.S. We must project SEER limited-duration prevalence to the U.S. population by race, and sum the races to get all races.

Input File Names

Load SEER limited-duration prevalence statistics calculated in SEER*Stat. A sample data file has been provided in the example folder (c:\users\<username>\documents\IMS\projprev\example). If you would like to learn how to use SEER*Stat to calculate the limited-duration prevalence, refer to the instructions on the following pages. To load the sample data files provided with this exercise:

1. Click the browse button to select the Statistic Dictionary. Select **seer8.ldprev.female.breast.byrace.and.age.dic** from the example folder (**c:\program files\projprev\example**). The data file name will be auto-filled as the dictionary is loaded. (The steps used to create the SEER limited-duration prevalence using SEER*Stat are provided on subsequent pages.)
2. Click the browse button to select the Population Dictionary. Select **uspops.female.byrace.and.age.dic**. The Population Data file name will be auto-loaded as this dictionary is loaded. The populations are for 20021 and 20022 combined. (The steps used to create the populations using SEER*Stat are provided on subsequent pages.)
3. Verify that **Input Populations are for Two Years of Data and Need to Be Averaged** is checked. As ProjPrev loaded the dictionary, it determined that the limited-duration prevalence estimates are for January 1st (rather than mid-year). In this exercise, this box should be checked since the populations are for 2021 and 2022 combined. If our populations were only for one year, we would uncheck this option.
4. Set the desired precision to be used for Prevalence Counts in the output.

Link Variables

Verify that the limited-duration prevalence variables are correctly linked to the population variables (e.g. race with race, age with age). To change the variable linkage, select the output variable of interest in the multi-column list box and then use the population drop-down list box to select the correct population variable.

Note, when there are prevalence variables that are not population variables (e.g. site), you should set the associated population variables to blank. In addition, there may be one population variable that has no matching limited-duration prevalence variable (e.g. state – to project SEER prevalence to each state).

You can also rename output variables, if appropriate. Since we want results by age for all races combined, change the name of the race output variable to: **Race (all races combined)**.

47-Year Limited-Duration Prevalence
U.S. Female Breast Cancer for All Races by Age at Prevalence
ProjPrev Example (continued)

Edit Output Groupings

For each linked variable, use the **Edit Groupings** button to verify the linkage within variables and to specify the output groupings of interest. In our case we would like to have prevalence estimates for all races combined by age.

Select the **Race (all races combined)** variable from the list and click the **Edit Groupings** button to load the **Output Groupings** dialog.

In the **Values** list box, select the values to be summed: W**hite+unk, Black, and Other (American Indian/AK Native, Asian Pacific Islander)**. Enter **All races** in the **Output Label** edit box. Click the **Add** button to create a new grouping. Since, we only want all races combined, select the 3 original groupings in the Groupings list box and click the **Remove** button. Click the **OK** button to return to the main dialog with the changes you have made. You should notice a **Y** in the **Mod** column of the variables list, signifying that the race variable has been modified.

Execute Job and View Results

Click the lightning bolt to execute job. Once the job is complete (almost instantly), ProjPrev will display the results in a new dialog.

There are two methods to export the results into a different application. You can **Save** your results to a data file or **Copy** them to the Windows clipboard. To save the results, you must first specify dictionary and data file names. Use the browse button to select a folder and specify a meaningful name for the Dictionary File (e.g. **us.ldprev.female.breast.byage.dic**). The output data file name will default to the same name with a txt extension.

47-Year Limited-Duration Prevalence
SEER 8 Female Breast Cancer, By Race and Age at Prevalence
ProjPrev Example

Objective

Use SEER*Stat to create a table showing the number of women in the SEER 8 registries diagnosed with a malignant breast cancer in the 47-year period prior to January 1, 2022.

Show the results by race (white + unknown, black, other) and age at prevalence (20 standard age groups).

Key Points

The SEER 8 prevalence percentages will be applied to U.S. populations to estimate U.S. prevalence. The SEER incidence data has cases with unknown race. There are no U.S. populations for unknown race. To account for the incidence cases with unknown race, we have to include them with one of the known races. Since white is the largest racial category, we will include the unknown races with whites when producing an all races estimate. The "SEER 8 white + unknown" percentage will be applied to the U.S. white population, the black and other percentages will be applied to the black and other U.S. populations respectively.

In this exercise, we will use the default Multiple Primary Selection option: "All Tumors Matching the Selection Criteria / One Tumor Per Statistic".

When calculating prevalence percentages, we need a population estimate for the prevalence date. Populations provided with SEER databases are mid-year population estimates. The January 1, 2022, populations are estimated by averaging 2021 and 2022 populations.

In this exercise, we do not need to use all of the standard survival cohort variables. Since this is a sex- and site-specific analysis, we do not need to include sex or site as survival cohort variables. Our selections will only include female breast cancer cases for the entire analysis including survival. Therefore, the survival estimates used in this exercise will be appropriate in regards to sex and site.

Even though we are combining white and unknown race in the prevalence estimates, we will still apply white survival estimates to lost white cases and unknown race survival to lost unknown cases.

Data Tab

Database: Incidence - SEER Research Data, 8 Registries, Nov 2024 Sub (1975-2022)

Selection Tab

Standard Selections:	Use all default values
Sex:	Female
Site recode ICD-O-3/WHO 2008:	Breast
Multiple Primary:	All Tumor Matching Selection Criteria / One Tumor Per Statistic

47-Year Limited-Duration Prevalence
SEER 8 Female Breast Cancer, By Race and Age at Prevalence
ProjPrev Example (continued)

Statistic Tab

Prevalence Date: January 1, 2022
Prevalence Duration: Limited to 47 Years
Statistic: Crude Percent (counts will be included in the results)

Table Tab

Use the dictionary editor to create a new race variable from the Race recode (White, Black, Other) variable. The new race variable should have 3 groupings (White + Unknown, Black, Other). Add the new race variable and Age at Prevalence Date (Calculated) as row variables.

Survival Cohorts Tab

Use the following variables from the System-Provided category as survival cohort variables.

Age recode (<60, 60-69, 70+) - Note this is created from Age of diagnosis variable, as opposed to the Age at prevalence.
Race recode (White, Black, Other, Unknown) - no total
Year of diagnosis (1975-1981,1982-2021 by 5)

Output Tab

Add a title and set the "Decimal Places for Estimated Prevalence Counts" to the maximum precision (0.000000001).

Results

Execute the session and save the resultant matrix as "seer8.ldprev.female.breast.byrace.and.age.spm". Export the results. On the export dialog, use the default settings except, check the options to:

- Remove All Thousands Separators (Commas)
- Remove Flags (Footnote), Prefix and Suffix Characters

U.S. Female Populations
By Race and Age
ProjPrev Example

Objective

In SEER*Stat, use a frequency session to create a table with population counts for females in the U.S. by race and age to be used to estimate U.S. prevalence for January 1, 2022.

Key Points

Populations provided with SEER databases are mid-year population estimates. To estimate January 1, 2022, populations, we need to average mid-year populations for 2021 and 2022. SEER*Stat does not have the ability to average the populations. ProjPrev does have the ability to divide the input populations by 2. Therefore, we will create populations for 2021 and 2022 combined and let ProjPrev average them.

Data Tab

Database: Populations - Total U.S. (1969-2023) <Katrina/Rita Adjustment>

Statistic Tab

Default values (frequencies, no percentages)

Selection Tab

Sex: Female
Year: 2021-2022

Table Tab

Add Race recode (White, Black, Other) and then Age recode with <1 year olds and 90+ as row variables. Note, the Race recode (White, Black, Other) variable does not contain a total in the population databases, so a user-defined variable is not required.

Results

Execute the session and save the resultant matrix as “uspops.female.byrace.and.age.sfm”. Export the results. On the export dialog, use the settings as in step 1. That is, use the default settings except, check the options to:

- Remove All Thousands Separators (Commas)
- Remove Flags (Footnote), Prefix and Suffix Characters