SEER*Stat

Don Green
Steve Scoppa
Information Management Services, Inc.

October 29, 2021

NCI Analytic Tools SEERies

NIH NATIONAL CANCER INSTITUTE

SEER

Purpose of SEER*Stat

- To provide cancer researchers with a convenient and intuitive system for…
  - Producing statistics useful in studying the impact of cancer on a population
  - Creating analytical datasets to be analyzed with other tools (e.g., DevCan, Joinpoint, ComPrev, JPSurv, R , SAS,…)
- Viewing/exporting the data for analysis in other software

Data Analyzed by SEER*Stat

- Incidence databases (SEER and other)
  - Used for incidence, survival, prevalence statistics, etc.
  - Your data converted to the file format required by SEER*Stat using SEER*Prep
- Mortality databases
  - Available from CDC National Health Center Statistics
  - Covers US and includes all causes of death

Population Databases in SEER*Stat

- Census US populations are available
  - In frequency sessions
- Population files are linked (behind the scenes) by age, sex, race, year of diagnosis, and geographic areas (including SEER registry and county) to produce rates
- Prior to 1990, population data are only available for “white”, “black”, “all other races combined”.
  - Limits analysis of data prior to 1990 to only include those 3 races categories.

SEER Databases At-A-Glance

SEER Over Time				
SEER Group	Year Initiated	Coverage of the US	Race	Ethnicity
SEER-9	1975	9.40%	White, Black, Other	NA
SEER-13	1992	13.40%	White, Black, API, AI/AN	Hispanic
SEER-18	2000	27.80%	White, Black, API, AI/AN	Hispanic
SEER-21	2000	36.70%	White, Black, API, AI/AN	Hispanic

API=Asian Pacific Islander, AI/AN=American Indians and Alaska Natives

SEER-21 has limited variables compared to SEER-21

Race and Ethnicity indicates how SEER data are linked to populations

SEER Data Products

- Starting with the 1975-2017 SEER Data, there are two primary data products available:
  - SEER Research
  - SEER Research Plus

- The changes were motivated because of concerns about the increasing risk of re-identifiability of individuals. The Research Plus databases require a more rigorous process for access that includes user authentication.

SEER Data Products (continued)

- The Research databases exclude variables due to:
  - Confidentiality
    - E.g., County, month of diagnosis, single ages beyond age 84 (through 99), marital status
  - Complexity or limitations in the ways that the variables could be analyzed or interpreted
    - E.g., Treatment fields
- The Research Plus databases provide access to the additional variables, though some are not available at the individual level in SEER*Stat case listing sessions.

SEER*Stat Databases

- All SEER incidence databases are available as Research and Research Plus
- SEER 9 Registry Database, 1975-2018
  - Approximately 9.4% of US population
  - 5,530,535 tumor records
  - Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, and Utah
  - White, Black, Other (Other is American Indian/Alaska Native and Asian/Pacific Islander combined)

- SEER 13 Registry Database, 1992-2018
  - Approximately 13.4% of US population
  - 5,450,399 tumor records
  - SEER 9 plus Los Angeles, San Jose-Monterey, Rural Georgia, and the Alaska Native Tumor Registry
  - White, Black, American Indian/Alaska Native, Asian/Pacific Islander by Hispanic Origin (Hispanic, Non-Hispanic)

SEER*Stat Databases (continued)

- SEER 18 Registry Database, 2000-2018
  - Approximately 27.8% of US population
  - 8,666,662 tumor records
  - SEER 13 plus Greater California, Kentucky, Louisiana, New Jersey and Greater Georgia
  - White, Black, American Indian/Alaska Native, Asian/Pacific Islander by Hispanic Origin (Hispanic, Non-Hispanic)

- SEER 21 Registry Database, 2000-2018
  - Approximately 36.7% of US population
  - 11,865,152 tumor records
  - SEER 18 plus Idaho, Massachusetts, and New York
  - White, Black, American Indian/Alaska Native, Asian/Pacific Islander by Hispanic Origin (Hispanic, Non-Hispanic)
  - Very limited data fields

SEER*Stat Databases (continued)

- Other available incidence databases
  - CiNA+ (NAACCR) incidence database: https://apps.naaccr.org/dart
  - USCS (NPCR + SEER) incidence database: https://www.cdc.gov/cancer/uscs/public-use/index.htm
    - Requires SEER Research Plus access

- Total US Mortality databases (not limited to cancer deaths)
  - All deaths from 1969-2019
  - Available as 1969+ with race White, Black, Other
  - Available as 1990+ with race White, Black, American Indian/Alaska Native, Asian/Pacific Islander by Hispanic Origin (Hispanic, Non-Hispanic)
  - Available with county and grouped year of death, and at state-level with individual year of death

Motivation for New Version

1. SEER*Stat is 25 years old!
2. The current version is 32-bit
   - 32-bit applications will soon be phased out for 64-bit applications.
3. The build tools for SEER*Stat are no longer supported / available
4. The need for a modern User Interface
5. Support for modern relational databases
   - SEER*Stat databases are in a custom format.
6. Future additions / improvements
   - Display configuration, sort / search features, comparison of results.

New SEER*Stat Development Phases

- **Phase I**
  - New User Interface for Client Application
  - Calculations Identical
  - Addition of Graphs
  - Minor Functional Improvements
- **Phase II**
  - Calculation Engine (64 bit)
  - Databases → Relational Databases
  - Updated Client Application (64 bit)
  - New SEER*Prep

Phase I Improvements

- Single Document Interface (SDI) is used (current Windows standard).
  - Multiple Document Interface (MDI) was used in old SEER*Stat.
- Graphs: Addition of basic graphing features with high resolution
- Improved User Interface (UI) for:
  - Case Selection Window:
  - Database Selection Window
- Modern UI: Users are more comfortable with a modern UI (similar to Word or Excel)

Working in SEER*Stat

SEER*Stat Sessions:

- Case Listing
- Frequency
- Rate
- Survival (October 28th)
- Limited-Duration Prevalence
- Multiple Primary Standardized Incidence Ratios
- Left-Truncated Life Tables

SEER*Stat Results:

- Results “Matrix” with Spreadsheet Features
- View
- Print
- Export to Text File
- Copy-and-Paste to other Software (e.g. graphics)

Frequency Session

- Frequencies
- Row and column percentages
- Define cohort based on a person's cancer history

Rate Session

- Crude (non-adjusted) rates
- Age-adjusted rates
- Trends over time expressed as percent change and annual percent change (based on either crude rates or age-adjusted rates)
- Rate ratios

Case Listing Session

- Creates tables showing the actual values stored in the database
- Provides a mechanism for displaying individual cancer records and patient histories (all tumor records for each person selected)
- Features:
  - select records based on values of data fields
  - specify the variables shown in the tables
  - sort by multiple variables
  - export for use in other software

Survival Session

- Observed survival rates – probability of surviving all causes of death for a specified time period
- Relative survival rates – observed survival adjusted for the expected survival (O/E)
- Cause-specific survival – net survival measure
- Conditional survival (special intervals), for example, 5-year survival given survived 3 years
- Period survival – uses only the most recent interval survival estimate of cases diagnosed in different calendar years (cross-sectional estimate of survival)
- Age-standardized survival

Prevalence Session

- Limited-Duration Prevalence Statistics
  - Counts with standard errors and confidence intervals
  - Crude percentages with standard errors and confidence intervals
  - Age-adjusted percentages

MP-SIR Session

- Multiple primaries in first cancer cohorts compared to the general population
- Standardized incidence ratios (SIR) of observed to expected primary cancers (O/E)
- Standardized mortality ratios (SMR) of observed deaths to expected deaths (O/E)
- Observed, expected, O/E, CI, and excess risk and other event related statistics
- Persons, person time at risk, and other cohort related statistics
- Special frequency and case listing tables specific to MP-SIR cohorts

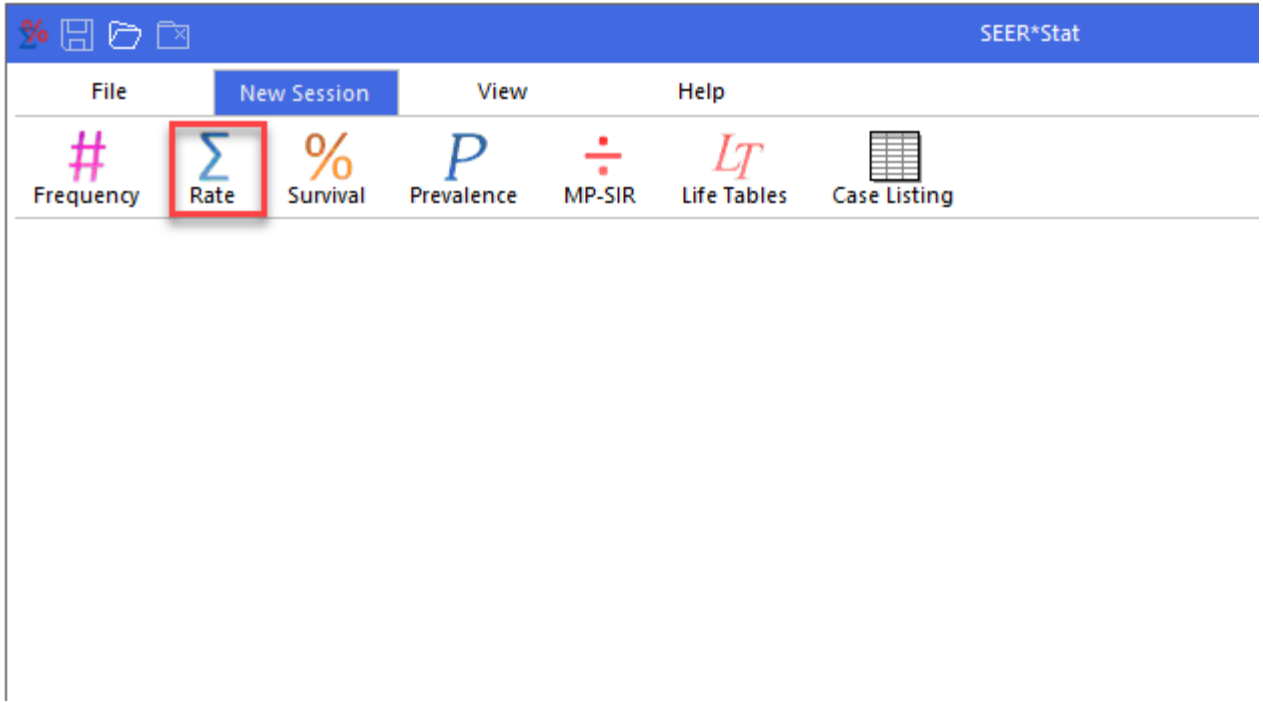
Left-Truncated Life Tables Session

- Allows estimation of net other-cause survival (survival from causes other than diagnosed cancer).
- Employing age at diagnosis as a time axis and taking into account left-truncated feature of the data.
- Both Kaplan-Meier and Actuarial methods are available for calculations of survival probabilities from left-truncated and right-censored data.
- Expected survival probabilities (Ederer II) are also available for comparison.

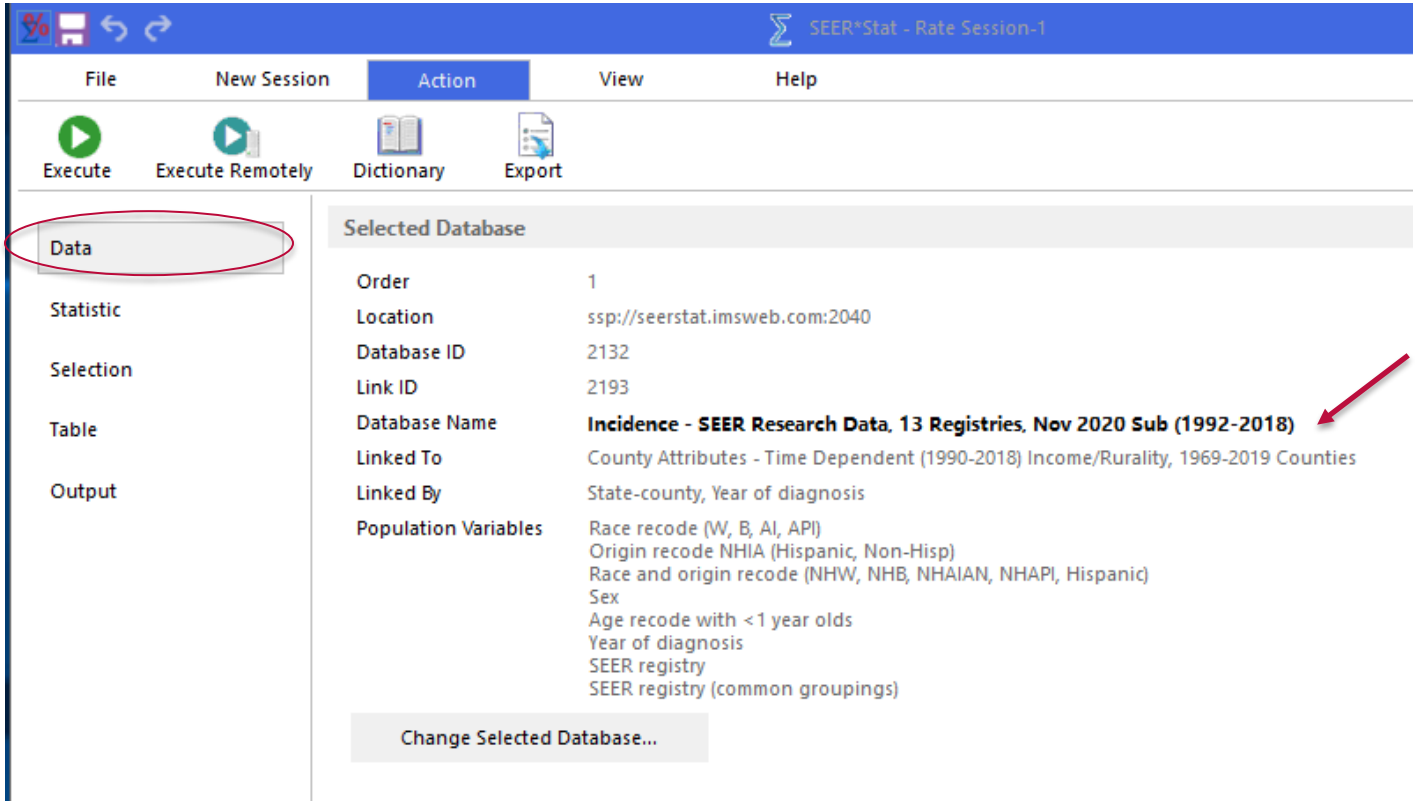
Basic Steps in SEER*Stat

1. Select a session (Frequency, Rate, Survival, etc….)
2. Choose Database (SEER-18, SEER-9, …)
   - Define the Cohort (Inclusion/exclusion)
   - Specify statistics to calculate (age-adjusted rates vs crude rates)
   - Define stratifying variables & output layout
3. Execute the session. SEER*Stat will present the results in a separate “results matrix” window.
4. Use the results: view or print the matrix, export statistics to text or other format or copy them to the Windows clipboard.

Select Session

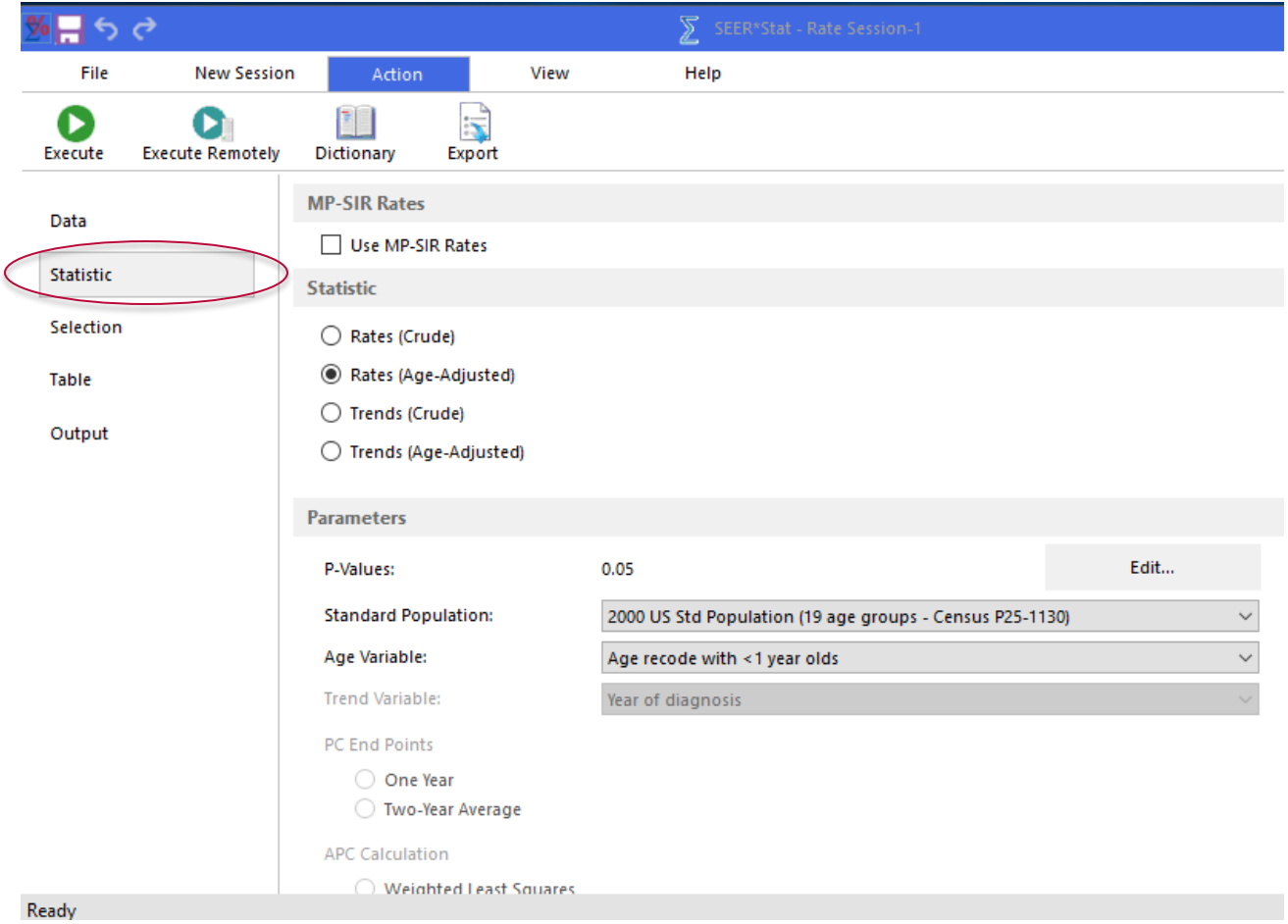


Select Database on Data Tab



Improved description of the database

Statistic Tab



Selection Tab

- Defines the cohort for analysis
- Different sessions have different common/required selection checkboxes
- In rate and prevalence sessions there are 3 edit boxes:
  1. The standard population and population linked age variable
  2. Other variables in the linked population data
  3. All other variables

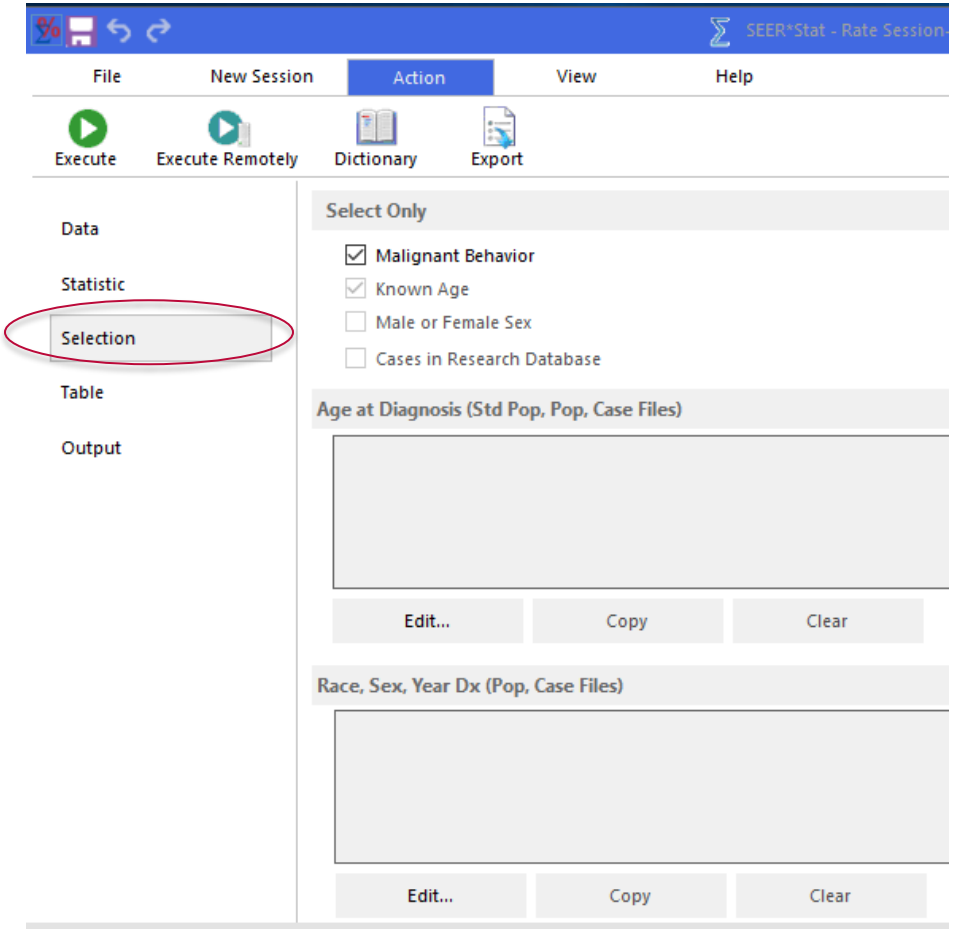
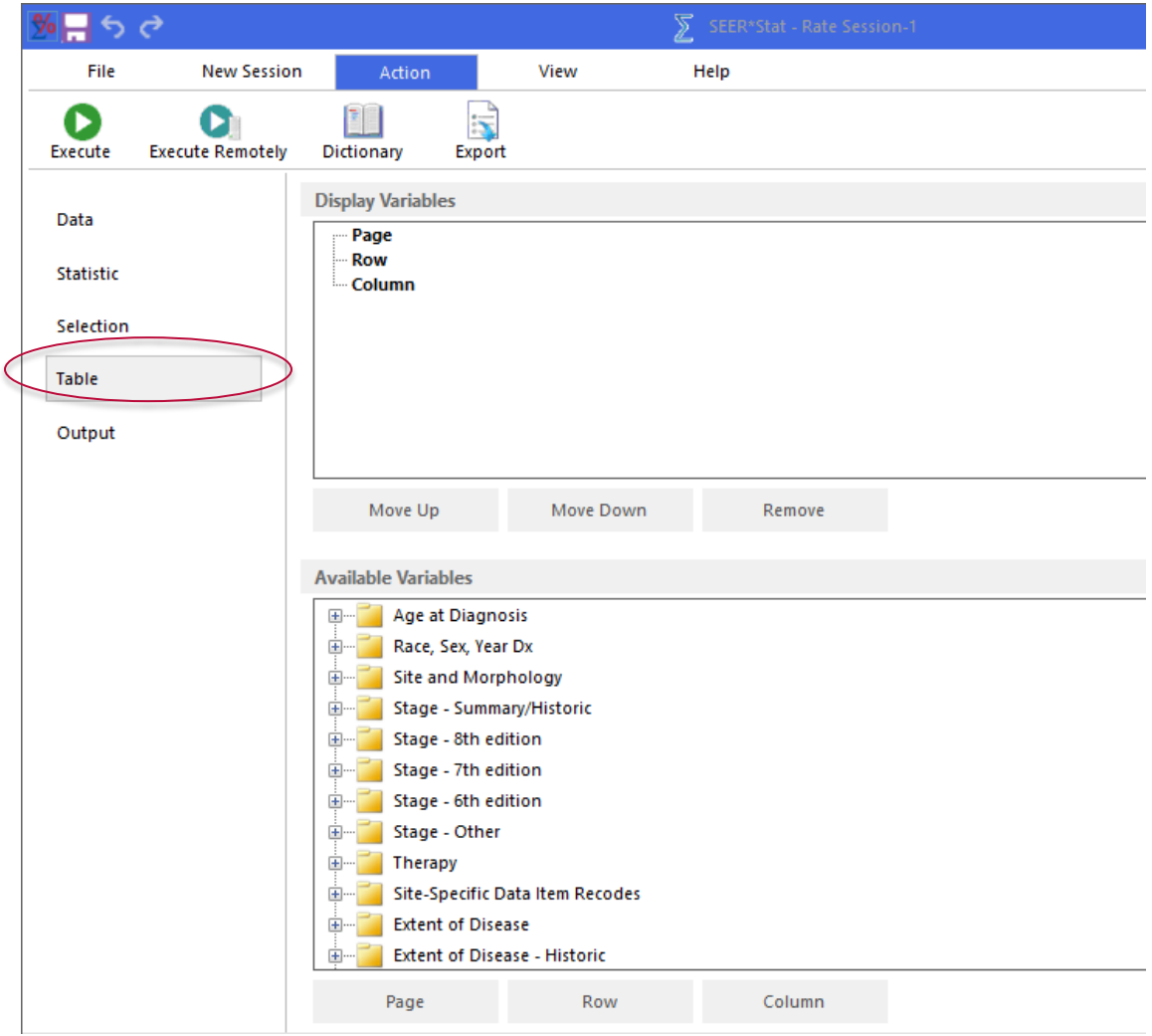


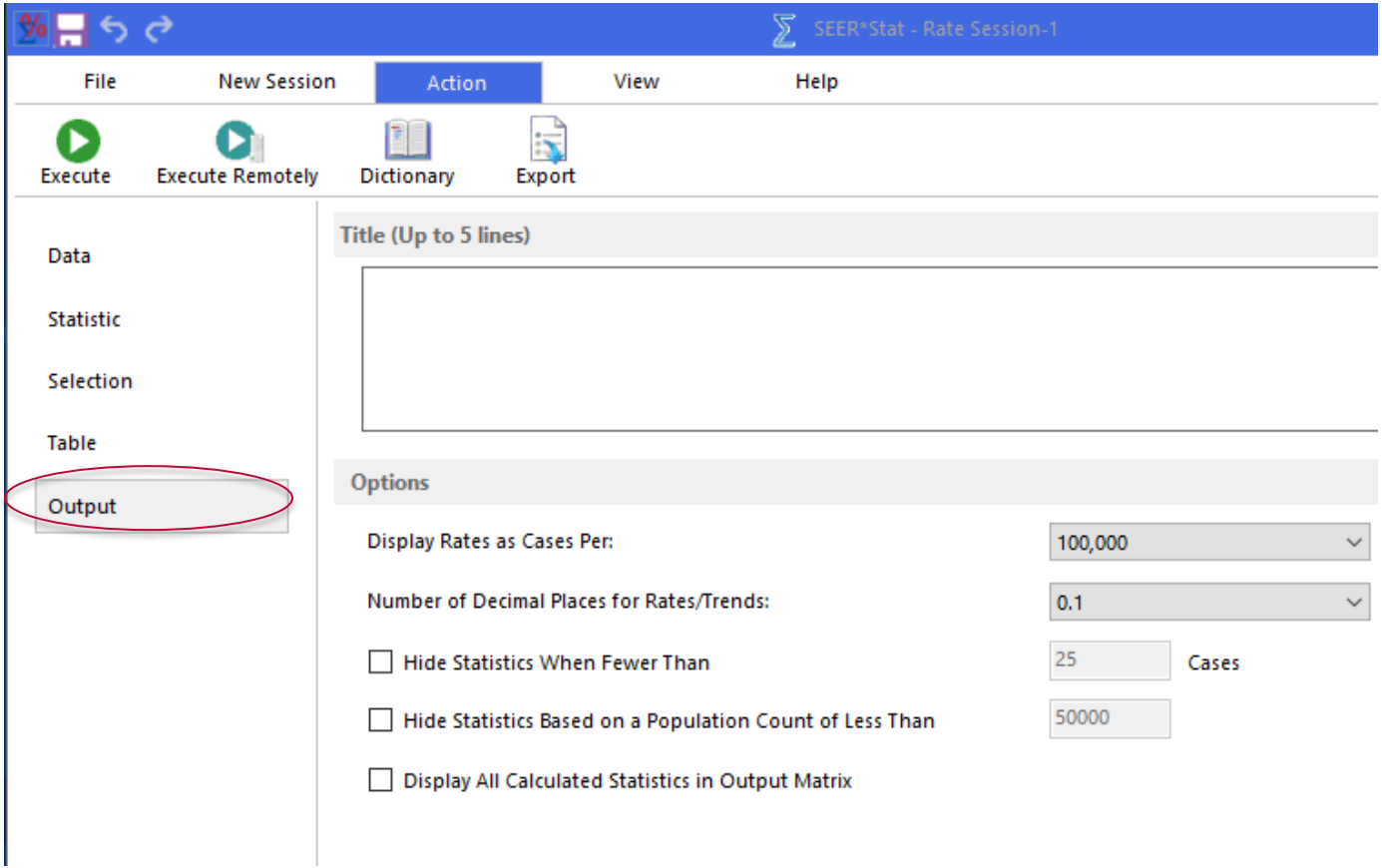
Table Tab

- Define stratifying variables for output
- Variables selected from Dictionary (organized in folders)



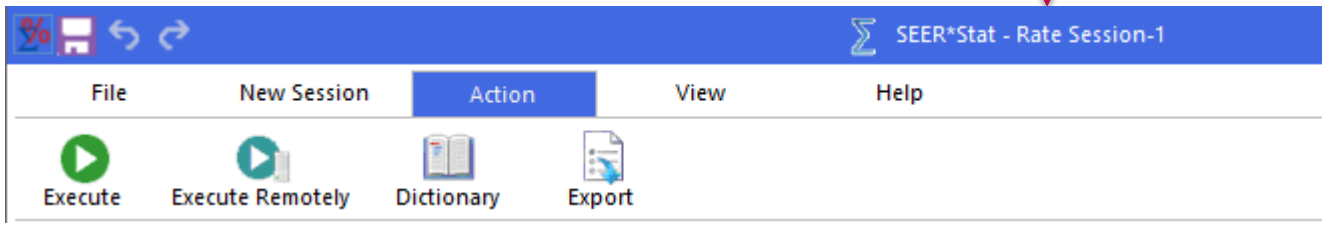
Output Tab

- Set titles and session-specific output options



Other Functionalities

Indicates the session



Run SEER*Stat

View and Edit Variables in the Data

Getting SEER*Stat

SEER*Stat software and associated research data are available, free of charge, on the SEER Web site:

http://seer.cancer.gov

Access to SEER Research Data requires a signed data use agreement form

If interested in trying new (Alpha version), please email seerstat@imsweb.com and request link to installation program.

http://seer.cancer.gov/seerstat

- SEER*Stat and SEER*Prep
- Tutorials
- Frequently Asked Questions (FAQs)
- Contact Information
- Manuals and Documentation
- Links to Related Software

Contact Information

- SEER*Stat Technical Support
  - seerstat@imsweb.com (please attach session file, if applicable)
- SEER Data Questions
  - seerweb@imsweb.com
- Ask a SEER Registrar
  - http://seer.cancer.gov/registrars/contact.html

www.cancer.gov

www.cancer.gov/espanol