A Guide to Using

SEER*Prep

Version 1.1

National Cancer Institute
Cancer Statistics Branch, DCCPS
EPN 343J; MSC 7352
6130 Executive Blvd.
Bethesda, Maryland 20892-7352

Table of Contents

INTRODUCTION TO SEER*PREP **3**

About SEER*Prep 3
About the National Cancer Institute 3
Acknowledgements 3
What You Will Need 3

GETTING STARTED **4**

Before Using SEER*Prep 4
Database Types 4
Overview 4
Moving from SEER*Prep to SEER*Stat 5
SEER*Prep Main Screen 5
SEER*Prep Reports 6
Setting Preferences 6

GETTING HELP **7**

Context-Sensitive Help 7
Technical Support 7

INPUT FILE REQUIREMENTS **7**

File Layouts Supported by SEER*Prep 8
Generating the Input File Description Report 9
Input File Description Report 9

DATABASE DESCRIPTIONS **11**

Opening a Database Description 11
Saving a Database Description 11
Editing a Database Description 11
Components of a Database Description 12
Database Name *12*
Input Case Files *12*
Input Population Files *12*
Population/Count Location *13*
Standard Million Location *13*
Expected Rate Location *13*
Study Cutoff Date *13*
Adding an Input File *14*
Removing an Input File *14*
Database Variables *14*

DATABASE VARIABLES **15**

Variables Table 15
Selecting Database Variables 15
Ordering the Variables Table 15
Adding Database Variables 16
Select Variable(s) to Add Window 16
Removing Database Variables 16
Editing Database Variables 17
Cause of Death Variables 17
Variable Attributes 18
Standard Variables *18*
Derived Variables *18*

Required Variables *18*
Indexed Variables *18*
Variables with Invalid Values *18*
USER-SPECIFIED VARIABLES 19
User-Specified Variable Requirements *19*
User-Specified Variables and Invalid Values *19*
EDIT VARIABLE WINDOW 20
Dealing with Database Variable Indexes *20*
Dealing with Invalid Values in a Variable *20*
Variable Description *21*
Editing Variable Format *21*
Edit Format Window *21*

VERIFY DATA PROCESS 22

WHY VERIFY YOUR DATA 22
VERIFYING YOUR DATA 22
VERIFY DATA WINDOW 23
VERIFY DATA REPORT 23

CREATE FILE PROCESS 25

CREATING A DATABASE FILE 25
RECORD EXCLUSION WINDOW 25
CREATE FILE WINDOW 26
CREATE FILE REPORT 26

ACCESSING YOUR NEW DATABASE IN SEER*STAT 27

TELLING SEER*STAT ABOUT YOUR NEW DATABASE 27
SELECTING YOUR NEW DATABASE IN SEER*STAT 27
DEALING WITH USER-SPECIFIED VARIABLES IN SEER*STAT 27
STANDARD MILLIONS DATABASES 28
EXPECTED RATES TABLES 28

DATABASE MANAGEMENT UTILITY 29

DATABASE MANAGEMENT WINDOW 29
RENAMING AN EXISTING USER DATABASE 29
DELETING AN EXISTING USER DATABASE 30

Introduction to SEER*Prep

About SEER*Prep

SEER*Prep is an efficient and flexible tool for converting ASCII data files to the SEER*Stat database format. Once created, these databases may be accessed in SEER*Stat to produce case listings, frequencies, rates, and survival statistics. Both systems are Microsoft Windows 95/NT applications which provide the functionality and intuitive interface commonly found in a Windows 95/NT environment.

SEER*Prep and SEER*Stat were developed by Information Management Services, Inc. in consultation with the Cancer Statistics Branch of the National Cancer Institute.

About the National Cancer Institute

One of the major roles of the National Cancer Institute, which commissioned SEER*Prep, is to collect, analyze and disseminate data useful in the prevention, diagnosis, and treatment of cancer. In support of this mission, the National Cancer Institute collects information on cancer incidence and survival through the Surveillance, Epidemiology and End Results (SEER) program.

The SEER program is composed of a set of population-based cancer registries across the United States which collect cancer data on a routine basis. The cancer data collected is submitted to the National Cancer Institute annually for analysis.

For more information about the SEER program, please refer to the SEER Web page:

www.seer.cancer.gov

Acknowledgements

SEER*Prep uses the zlib 1.1.3 library written by Jean-loup Gailly and Mark Adler used for decompressing gzipped files. This library is free and can be obtained at the following web address.

www.cdrom.com/pub/infozip/zlib/

What You Will Need

To use SEER*Prep you will need:

- 486- or Pentium-based personal computer (Pentium recommended)
- Microsoft Windows 95/NT with a text editor such as NotePad installed
- SEER*Stat version 1.1 or later installed
- Support for long filenames by PC and LAN (if utilized)
- 16 MB application RAM
- 2 MB hard disk space

Getting Started

Before Using SEER*Prep

SEER*Prep requires input data files to have certain characteristics. It is important to prepare your raw input data before using it with SEER*Prep. To produce documentation for the file layouts supported by SEER*Prep, see Generating the Input File Description Report. In general, input files must follow these guidelines.

- The input files must be in one or more ASCII text files (no tabs or control characters) with one line per record.
- All input files must have the same layout, which is based on a file layout supported by SEER*Prep.
- All input files must have the file extension .TXD or TXD.GZ for SEER*Prep to recognize them.

Database Types

SEER*Prep allows users to create four different types of databases to be accessed in SEER*Stat for use in the production of case listings, frequencies, rates, and survival statistics.

Incidence and Mortality Databases

Incidence and mortality databases created using SEER*Prep will be listed on the Data Tab for use in any SEER*Stat session. Case listings, frequencies, and survival statistics may be produced using these databases. To calculate rates or trends, a database must be created with the appropriate population files.

Standard Million Databases

A Standard Millions database created using SEER*Prep will appear on the Data Tab of SEER*Stat's Frequency/Rate and Case Listing sessions. Standard million data are used to calculate age-adjusted rates and are available for use with any incidence or mortality database in SEER*Stat.

Expected Rates Tables

Expected Rates Tables created in SEER*Prep will be listed on the Data Tab of SEER*Stat's Survival and Case Listing sessions. They are used to calculate survival statistics.

Overview

SEER*Prep allows you to create a database from your own data and begin using it in SEER*Stat in four easy steps.

Step 1: Set Up

To create a database from your own data, you must prepare a description of the new database for SEER*Prep. You should choose a database description template (.DD file) that is based on the same file layout as your input file(s). Once you have opened the appropriate database description files by choosing Open from the File menu, you must name your database, identify the files you are using, and select and edit the variables

you want included in the database. User-specified variables have been built in to all incidence and mortality database descriptions making it possible to design your own specific variables.

Step 2: Verify

You should verify your data by choosing Verify Data from the Execute menu. Using your data, this process produces a frequency table of the valid and invalid values of each variable in the database description. You may need to make corrections to your data before proceeding.

Step 3: Create

Next create a SEER*Stat database by choosing Create File from the Execute menu. When SEER*Prep creates a database, it creates a subdirectory for the new database in the directory you specified on the Preferences window.

Step 4: SEER*Stat

Once a database has been created in SEER*Prep, it is ready for use in SEER*Stat. During the create file process, the directory listed on the SEER*Stat Preferences window under the heading “Secondary Data Location” is updated by SEER*Prep to the directory where the new database was saved. Once a database has been created in SEER*Prep, you may change the name or remove it through the Database Management option on the SEER*Prep **File** Menu. To make a change to a database once it has been created, edit the database description and recreate. The old database will be over-written with the corrected version.

Moving from SEER*Prep to SEER*Stat

Once you have created a database using SEER*Prep, it is ready to be accessed in SEER*Stat. SEER*Prep automatically updates the Secondary Data Location in the SEER*Stat Preferences to the directory containing the new database during the create file process. Simply open a new session in SEER*Stat and select your database from the Data Tab.

You will not be able to calculate rates or trends for databases created without population files, but frequencies and survival statistics may be calculated for any database created with SEER*Prep. The standard millions and expected rates tables used for some calculations are the same for all SEER*Stat databases.

SEER*Prep Main Screen

From the SEER*Prep Main Screen, the database description may be viewed and edited. Click on any of the listed components for specific help.

- Database Name
- Input Case File(s)
- Input Population File(s)
- Population/Count Location
- Expected Rates Location
- Standard Million Location
- Study Cutoff Date for Survival Calculation

SEER*Prep Reports

SEER*Prep has three reports: Input File Description Report, Verify Data Report, and Create File Report. When a report is created, it is immediately displayed in your default text editor. To change the default text editor, see Setting Preferences. To reopen a report that has been closed in your default text editor, simply select View File Description/Report from the File menu and double-click on the file you want to open.

Input File Description Report

When preparing your input files for SEER*Prep, it may be helpful to have written documentation for the file layout you are planning to use. The Input File Description Report contains details about every variable you have chosen from database description. The report includes the variable name, position, length, settings in effect, and may include valid values. To create the Input File Description Report, see Generating the Input File Description Report.

Verify Data Report

When you choose Verify Data from the Execute menu, the Verify Data Report is produced. Ideally, this diagnostic tool is used to check the database description settings and the data in the input files. The report provides frequency of valid and invalid values in the input files for each variable. For more information about the verify data process, see Verifying Your Data.

Create File Report

When you choose Create File from the Execute menu, in addition to creating the database files, the Create File Report is produced. It provides a record of the database description settings and input files used to create the new SEER*Stat database, including any notes or warnings. As a general practice, keep this report for every active SEER*Stat database created with SEER*Prep. For more information about the create file process, see Creating a Database File.

Setting Preferences

To access the Preferences window, choose Preferences from the File menu.

Directory to Save User-Defined Databases

You must enter a directory where the new SEER*Stat databases created by SEER*Prep can be saved. This directory must allow both read and write privileges. SEER*Prep will not allow you to create a database without specifying a directory. If you do not know the entire path name, click the Browse button to select the path.

Default Editor

You may select a text editor or leave this field blank. When left blank, reports will be opened in the program you have associated with the extension you specify for the report. If you do not have an associated program, NotePad will be used.

Getting Help

SEER*Prep provides help relating to all system options. The help system defines how to create a database to your specifications. There are four ways to access the help system from SEER*Prep:

- Click from the toolbar.
- Choose Contents from the Help menu.
- Click the Help button provided on all windows.
- Press F1 on the keyboard for context-sensitive help.

Contents Tab

Click the Contents tab to browse through topics by category.

Index Tab

Click the Index tab to see a list of index entries: either type the word you're looking for or scroll through the list.

Find Tab

Click the Find tab to search for words or phrases that may be contained in a Help topic.

Context-Sensitive Help

Context-sensitive help is available for each window. To access this help, type the F1 key from your keyboard or click on a provided Help button. A description of the window will appear in a help window.

Technical Support

For an up-to-date list of known problems, frequently asked questions (FAQ's), and resolutions, please consult the SEER*Prep web page at:

www.seer.cancer.gov/ScientificSystems/SEERPrep

For technical support beyond the help system, please send e-mail to IMS at:

seerprep@ims.nci.nih.gov

Input File Requirements

Name

All input files must have a .TXD or .TXD.GZ file extension. If necessary, you can rename your file by adding .TXD to the end. SEER*Prep supports long file names. The .GZ is used when the data set has been gzipped. For information about the gzip utility, please see the readme file.

Layout

All input files must be ASCII text files containing one record per line without tabs or control characters with the same layout based on a file layout supported by SEER*Prep. Therefore, you may not have packed decimals, binary fields, or EBCDIC characters

anywhere in the file. End of line characters (carriage returns and/or line feeds) may appear at the end of each record, but are not required. To obtain file layout documentation from SEER*Prep, see Generating the Input File Description Report.

File Layouts Supported by SEER*Prep

SEER*Prep currently supports the 5 file layouts listed below.

Incidence and Mortality layouts

File layout: **SEER 250**

Record length: 250 characters

Populations by: Registry

SEER250.monyyyy.Dmmddyyyy.DD

SEER250.monyyyy.CUSTOM.RACE.Dmmddyyyy.DD

File layout: **NAACCR**

Record length: 1525 characters

Populations by: State/province/country

NAACCR1525.VERn.Dmmddyyyy.DD

NAACCR1525.VERn.CUSTOM.RACE.Dmmddyyyy.DD

File layout: **MORTALITY**

Record length: 58 characters

Populations by: State/province/country

MORT58.monyyyy.Dmmddyyyy.DD

MORT58.monyyyy.CUSTOM.RACE.Dmmddyyyy.DD

Standard Million layout

File layout: **STANDARD MILLION**

Record length: 11 characters

Populations by: N/A

STDMIL11.APR2000.D04142000.DD

Expected Rates Layout

File layout: **EXPECTED RATES**

Record length: 17 characters

Populations by: N/A

EXPRATE17.DEC1999.D04142000.DD

- monyyyy is the submission date for SEER. VERn is the version number for NAACCR.
- Dmmddyyyy is the date when this .DD file was made available.
- "CUSTOM.RACE" indicates that population and case files are linked for producing rates by a 2-digit user-defined race rather than the SEER 250 Race Recode A.

Notes

- A comprehensive population file layout has been included in the readme.txt.
- Not all variables found in the NAACCR file layout have been incorporated in the database descriptions. Please see the readme.txt for more information.
- Consult www.seer.cancer.gov/ScientificSystems/SEERPrep/ for specifics on file layouts.

Generating the Input File Description Report

When you are preparing your data for SEER*Prep, it may be helpful to have documentation for the file layout. If you do not have this documentation on-hand, you can create an Input File Description Report in SEER*Prep by following these steps.

1. Open a database description (See file layouts).
2. Add or remove variables as necessary and sort the Variables Table by clicking on the column headings.
3. Once the Variables Table has been set up in the same order that you want the Input File Description Report, select Generate Input File Description from the File menu.
4. In addition to the position of the variables in the file, you are given the opportunity to include a list of the valid values for each variable. Click "Yes" to accept or "No" to decline. Listing the variable values may make the report very long.
5. You must name the new report. The Input File Description Report will appear in your default text editor.
   - To obtain a paper copy of the report, print the text file using your default text editor.
   - To view the report after it has been closed, select View File Description/Report from the File menu and double-click the report's file name.

Input File Description Report

The Input File Description Report contains the details about every variable included on the Variables Table. The variables are listed in the same order on the Variables Table. The report includes the variable names, positions, length, and settings in effect.

Specifically, the Input File Description Report has the following components.

- Full path name of the database description file
- Record length of all file

- Month and year chosen for the study cutoff date for survival calculations (if required)
- Name of each variable
- Required or not required status of each variable
- Index or no index, as specified for each variable
- Recode invalid values or exclude invalid values, as specified for each variable
- Indicator for derived variables
- Start column of each variable in the input case file
- Start column of each variable in the input population file
- Length of the variable in the input files
- Variable description
- Valid values and format labels for each variable (optional)
- List of variables used to derive a variable when appropriate

To create the Input File Description Report, see Generating the Input File Description Report.

Database Descriptions

Database Description files contain all of the information to define a database. This includes the file layout, database name, file names, the study cut off date for survival statistics, and the required variables. It may include additional input case files, one or more population files, and other variables defined by the input file layout. To calculate statistics other than frequencies and survival in SEER*Stat, you must have a population file associated with your incidence or mortality database. It is a good idea to save the final version of the database description after you create the file.

Database description files have the file extension .dd. These should not be edited outside of SEER*Prep. The basic database descriptions which are provided with SEER*Prep can be found in the same directory as the SEER*Prep executable (.exe). See File Layouts Supported by SEER*Prep for more information about the basic database descriptions.

Opening a Database Description

SEER*Prep comes with some sample database descriptions (.DD files). When creating your own SEER*Stat database, edit one of these sample .DD files to suit your needs. A list of the sample database description files is given under the heading File Layouts Supported by SEER*Prep.

To open a database description, you may do any of the following.

- Click on .
- Use the File menu command Open.
- Select a recently opened database description from the Recent File List at the bottom of the File menu.

Saving a Database Description

It is a good idea to save the database description for every SEER*Stat database you create. If changes are needed in the database, edit the saved database description and recreate the database for SEER*Stat. To save a database description, you may do either of the following.

- Click on to resave the database description.
- Use the File menu Save or Save As commands.

Editing a Database Description

You must define the settings you want for your new database for SEER*Stat by making changes to a database description. Complete each of the following steps.

- Name the database
- Identify the files that need to be included in the database

- Set the study cut off date for survival calculations (if required)
- Select the variables that you want to be able to use in SEER*Stat
- Edit the selected variables

Be sure to check each of the database description elements for accuracy and verify the input data before creating the database. Remember, only the database name can be changed after the database has been created. If corrections are necessary after the database has been created, you must delete the old database and re-create a new corrected version.

Components of a Database Description

Database Name

The database name identifies your newly created database on the Data Tab in SEER*Stat. Within a directory of databases, the names must be unique. Enter the database name by clicking in the text field to begin typing. The database name can be up to 100 characters long. Some sample names are:

SEER250 w/ Pops by Reg: My Registry Aug98 Submission

SEER250 Female Breast Cancer 1950-2000 for Dr. Smith

Input Case Files

All input case files added to a database must:

- have the file extension .TXD or .TXD.GZ (The .GZ indicates that the file is gzipped. Information about the gzip utility is available in the readme.txt file.)
- be ASCII text files without tabs or control characters (End of line characters such as a carriage return may appear at the end of each record, but are not required.)
- contain 1 record per line
- have a single layout based on one of the file layouts supported by SEER*Prep

If you do not have the file layout documentation on-hand, see Generating the Input File Description Report for directions to produce it.

Input Population Files

Population files are required for the calculation of rates and trends in SEER*Stat. All input population files added to a database must:

- have the file extension .TXD or .TXD.GZ (The .GZ indicates that the file is gzipped. Information about the gzip utility is available in the readme.txt file.)
- be ASCII text files without tabs or control characters (End of line characters such as a carriage return may appear at the end of each record, but are not required.)
- contain 1 record per line
- have the same layout as in the database description you selected

It is a good idea to verify that the populations are by the same geographic area types (registries or states/provinces/countries) in both the input population file and the database description. If you do not have the file layout documentation on-hand, refer to the readme.txt for a file layout and see Generating the Input File Description Report for directions to produce a list of valid values.

Population/Count Location

There is one variable in the Input Population file(s) which is not found in the Input Case file: the population count. The population count variable contains the number of people which lived in a single year and geographic location of the same race, sex, and age group. A blank is not accepted as a valid value for the population count variable. Invalid values are always excluded.

The starting column and length of the population count variable in the Input Population file are listed in the bottom left corner of the main screen and in the readme.txt file when an incidence or mortality database is being created.

Standard Million Location

The standard million variable for age group X contains an integer representing the number of persons per 1,000,000 who are alive for that age group. Across all age groups, the standard million variables should sum to 1,000,000. A blank is not accepted as a valid value and values must be zero filled on the left. Invalid values are always excluded.

The starting column and length of the count variable in the Input Standard Million file are listed in the bottom left corner of the main screen and in the readme.txt file when a standard million database is being created.

Expected Rate Location

The expected rate variable for age X contains an integer representing the number of persons per 1,000,000 that are expected to live from age X-1 for a particular sex, race, and year combination. A blank is not accepted as a valid value and values must be zero filled on the left. Invalid values are always excluded.

The starting column and length of the expected rate variable in the Input Expected Rate file are listed in the bottom left corner of the main screen and in the readme.txt file when an Expected Rate Database is being created.

Study Cutoff Date

SEER*Prep uses the study cutoff date to derive the survival time variables. In particular, it specifies the date when the study ended. No survival time is calculated past this date. Typically the date of last possible diagnosis is used as the study cutoff date, unless it is known that excellent follow-up information was obtained through a particular date. The Study Cutoff Date must be set before your incidence or mortality database can be verified or created. To set the study cutoff date, follow these steps.

1. Click on the Study Cutoff Date Edit button in the bottom left corner of the Main screen. The Study Cutoff Date Window will appear.
2. Select a month for the study cutoff date from the drop down box.
3. Type the year for the study cutoff date in the text box.
4. Click OK.

Adding an Input File

To add an input file follow these steps.

1 Click the Add button beneath either the Input Case File(s) or Input Population File(s) text windows.

2 Select the input file(s) you want included in your database.

3 Click Open.

The full path name of your input case file(s) will appear in the appropriate text window. All files must have the appropriate layouts as defined by the database description.

Removing an Input File

To remove an input file follow these steps.

1 Select the input file that you do not want to include in your database description from the Input Case File(s) or Input Population File(s) text windows.

2 Click the Remove button beneath the text window or press the Delete key on your keyboard.

Database Variables

Database variables listed in the Variables Table will be available for use in SEER*Stat. Only variables defined in the file layout may be selected as database variables. At least one variable must be defined for the database to be created.

- Some variables are required by SEER*Stat for certain statistics. These are marked as “Required” and cannot be removed from the list of database variables.

- Some variables are derived from other variables by SEER*Prep because they are needed in SEER*Stat. These are marked with an asterisk.

For every variable, the creation of an index, the treatment of invalid values, and the variable description may be edited. In user-specified variables, which are identified by the active Edit button at the bottom of the Edit Variables Window, the list of valid values and their labels may also be edited. When it is not possible to edit the labels of the valid variables, you should plan to create user-defined variables in SEER*Stat. Please see the SEER*Stat help documentation for instructions.

Database Variables

Variables Table

Variables selected to be in the database are listed in the Variables Table on the main screen along with the following parameters. Those listed below in italics can be edited.

- Variable name
- Input case file start column
- Input population file start column
- Variable length
- Required by SEER*Stat or not
- *Index created or not*
- *Recode or exclude invalid values*
- *Description*

The variables table can be sorted by the values in the various columns. See Ordering the Variables Table for instructions.

Selecting Database Variables

To select variable(s) in the Variables Table, use any of the following methods.

- Click on the name of a variable.
- Click and hold to the right of a variable name you want to select. Drag a Lasso to the left and up or down to highlight the variable names around it.
- Click on a variable name. Hold down the Shift key and click on another variable name. This selects both variables and all those between.
- Click on a variable name. Hold down the Control key and click on another variable name. This methods allows you to select non-adjacent variable names without forcing you to select those in between.

Ordering the Variables Table

To sort the Variables Table, click on the column headings. Here are examples of sorting methods.

- To order the variables by their column position in the population file, click on the column heading “Pop Start Col.” The variables found in the population file will be listed first in order of their position followed by all variables not in the population file.

- Click “Name” to list the variables alphabetically by name. Notice that the derived variables will be listed first because of the asterisk at the beginning of their name.

Adding Database Variables

To include a variable defined by the file layout in your database, follow these steps. Notice that the Add button is disabled when all possible variables have been selected.

1 Click on the Add button beneath the Variables Table. The Select Variable(s) to Add window will open.

2 Select the variable(s) you want to add to your database by any of the following methods.

- Click on the name of a variable.
- Click and drag over a group of consecutive variables.
- Click on a variable. Hold down the Shift key and click on another variable. This selects both variables and all those between.
- Click on a variable. Hold down the Control key and click on another variable. This selects non-adjacent variables.

3 Click OK.

Select Variable(s) to Add Window

The Select Variable(s) to Add window appears when the Add button beneath the Variables Table is clicked. It is a list of all variables defined by the file layout, but not currently selected . These will not be available in SEER*Stat once the database is created. Once a variable has been selected for the database definition, it is no longer listed in the Select Variable(s) to Add window. To select a variable for the database, click on the variable name and click OK. For detailed instructions, see Adding Database Variables.

Removing Database Variables

Variables removed from the database definition will not be available in SEER*Stat, once the SEER*Stat database has been created. Only variables which are not required by SEER*Stat can be removed. Removed variables can be added later using the Select Variable(s) to Add Window.

There are some instances when removing a particular variable makes sense. For example, you might remove a non-required variable which is always blank. Perhaps you would decide to remove the individual ids to comply with privacy standards.

To remove a variable from the database definition, follow these steps.

1 Select the variable(s) you want to remove using any of these methods.

- Click on the name of a variable.
- Click and hold to the right of a variable name you want to select. Drag a Lasso to the left and up or down to highlight the variables around it.

- Click on a variable. Hold down the Shift key and click on another variable. This selects both variables and all those between.

- Click on a variable. Hold down the Control key and click on another variable. This selects non-adjacent variables.

2 Click the Remove button.

Editing Database Variables

Select the variable from the Variables Table and click the Edit button beneath. The Edit Variable Window will appear. You may select only one variable at a time for editing. The creation of an index, the treatment of invalid values, and the variable description can be edited for any variable.

User-specified variables, registry and custom race, may also have the valid values and their labels edited. Values for these variables are always numeric. You may assign labels to each valid numeric code on the Edit Format Window. You cannot assign a label for a range of valid values in SEER*Prep. Instead, you must assign an individual label to each valid value. After the database has been created, you may make user-defined variables in the SEER*Stat database dictionary which have the groupings you wanted. Please see the SEER*Stat Help for instructions.

Cause of Death Variables

Columns 21-39 of the Mortality database description constitute the cause of death variables. There are three cause of death variables (position 1, position 2, and position 3). Each is followed by a variable which indicates the ICD coding scheme used in the corresponding cause of death variable. In addition, the COD Position Indicator allows you to select position 1, 2, or 3 to be used to derive the Cause of Death Recode variable. Though codes exist for ICD-6 through ICD-10, only ICD-8 and ICD-9 may be used to derive the Cause of Death Recode variable at the present time.

Setting up the COD variables

Typically the primary cause of death values are placed in the first position. The ICD Coding Scheme variable for the first position will indicate the ICD coding scheme used for the cause of death and all other positions will be left blank. Finally, the COD Position Indicator will be coded to 1 for all records.

It is possible that the primary cause of death exists in the data in more than one ICD coding scheme. In this case, you should place the cause of death values in the positions based on the coding scheme. For example, you might put ICD-8 data in the first position columns and the ICD-9 data in the second position. The ICD Coding Scheme variables would be the same for all records, position 1 = “08” and position 2 = “09”. Finally, the COD Position Indicator would be coded to 1 for records with primary ICD-8 data and 2 for records with primary ICD-9 data, based on which was the original of the two for a particular record.

Variable Attributes

Standard Variables

The value of a standard variable is read directly from the input files. A standard variable does not have an asterisk at the beginning of its name in the Variables Table.

Derived Variables

The value of a derived variable is calculated by SEER*Prep based on other variables. Lists of the variables used to do the calculations for derived variables can be found in the Input File Description Report. A derived variable has an asterisk at the beginning of its name in the Variables Table to indicate that it is derived.

Note

If a derived variable is also a population variable, the value of the variable must be in the population file. It is not derived in the population file.

Required Variables

Some variables are necessary in SEER*Stat for the calculation of certain statistics such as survival statistics, trends, and rates. These variables are marked as “Required” and cannot be removed from the list of database variables. To see a list of the required variables, sort the Variables Table by the Required column.

Indexed Variables

Variable indexes are created to speed up the record selection in SEER*Stat. However, creating the indexes requires a small amount of additional processing time in SEER*Prep and more disk space to store the database. It is recommended that most variables in the database be indexed. However, an index may not be necessary for a variable that will rarely be used in SEER*Stat.

Variables that have a “Yes” in the Index column of the Variables Table will be indexed. To change this setting, see Dealing with Database Variable Indexes.

Note

Some variables will not have an index created for them no matter how they are marked in the database description. If a variable has a large number of unique values and there is a large number of records in the database, the index will not be created. Instead a message will appear after the database has been created which lists the variables without indexes.

Variables with Invalid Values

When invalid value(s) are found in the input files, SEER*Prep must either recode them to “Invalid Value(s)” or exclude the record in which an invalid value is found. An invalid value is any variable value found in the input file that is neither a blank nor listed as a valid value on the Edit Variable window. The only exception to this rule is the population count variable which may never be blank. To discover if there are invalid values in your data, run the verify data process.

You may choose to treat invalid values individually for each variable when editing the database description or you may make a global rule when you create the database. Variables that are marked to remove records with invalid values have “Exclude” in the

Invalid Values column of the Variables Table. To change the invalid values treatment for an individual variable, see Dealing with Invalid Values in a Variable. To make a global rule for the treatment of invalid values, see Record Exclusion Window.

User-Specified Variables

User-specified variables are the only variables in which valid values and their labels may be edited. They can be found in incidence and mortality database description files to allow you to design a database which handles your data. In addition to the variables named "User-specified" in the input case files, the registry and custom race variables found in some of the basic database descriptions can also be edited. These two variables affect both the input case files and input population files. Specific requirements for user-specified variables can be found in the section titled User-Specified Variable Requirements.

For directions on editing the values of user-specified variables, see Editing Variable Format.

You should enter meaningful text in the variable description field, since the variable name cannot be changed. See the section titled Variable Description for instructions.

For hints on dealing with user-specified variables once the database has been created and is available in SEER*Stat, see Dealing with User-Specified Variables in SEER*Stat.

User-Specified Variable Requirements

User-specified variables are defined by their position and length in the input files. The input text values for these variables must be numeric. It is important to choose the most compatible user-specified variable for your data. If you have 40 possible values, then you need at least a two-digit code which holds the values 00 - 99. No letters or symbols may be defined for these variables. However, a blank space is considered a valid missing value for all user-specified variables just as it is for all variables in the database description, except population count.

User-Specified Variables and Invalid Values

To check that the format of a user-specified variable will fit the input data, run the Verify Data Process. You should not expect many invalid values from user-specified variables. Therefore if there are several records with invalid values, you should investigate the reason for them. To handle invalid values in a user-specified variable, determine if the invalid values are numeric or not.

⇒ If the invalid values are numeric, you may simply add the invalid values as valid values to the variable format.

⇒ If the invalid values are non-numeric, you may do any of the following.

- From the Edit Variable Window, check the box labeled "Exclude records with invalid values" to set the variable to remove all records with an invalid value in this variable.

- From the Edit Variable Window, clear the box labeled "Exclude records with invalid values" to set the variable to recode all invalid values to a single value labeled "Invalid Value(s)."

- In an application outside of SEER*Prep, recode non-numeric values to numeric codes which may be added to the variable format.

Edit Variable Window

Each variable is defined by the following parameters found on the Edit Variable Window. Changes to a variable are made on this screen. Only items listed below in *italics* can be changed. Valid values and their labels may only be changed in user-specified variables, including registry and custom race.

- Variable name
- Input case file start column
- Input population file start column
- Length of variable text
- Required by SEER*Stat or not
- *Index created or not*
- *Recode or exclude invalid values*
- *Description*
- *Valid values with labels as they will appear in SEER*Stat*
- Derived variable indicator

Dealing with Database Variable Indexes

To change whether an index will be created or not for a variable, follow these steps.

1. Select the variable you wish to change.
2. Click the Edit button beneath the Variables Table.
3. On the Edit Variable Window, click the check box labeled “Index variable” to change the index status. When checked, an index will be created at the same time the database is created.
4. Click OK.

Dealing with Invalid Values in a Variable

To change how invalid values are handled for an individual variable, follow these steps. If you would like to exclude invalid values for all variables or not exclude for any of them, see Record Exclusion Window.

1. Select the variable you wish to change.
2. Click the Edit button beneath the Variables Table.
3. On the Edit Variable window, click the check box labeled “Exclude Records with Invalid Values” to change the invalid values status. When checked, records with invalid values will be excluded.
4. Click OK.

Variable Description

The variable description is displayed in the database dictionary in SEER*Stat. Any text may be entered as the variable description. It is particularly important to write meaningful descriptions for the user-specified variables since the variable name cannot be changed. To edit the variable description for a particular variable, follow these steps.

1 Select the variable from the Variable Table and click Edit.

2 The Edit Variable Window will appear.

3 Make changes to the description text.

4 When you are finished, click OK.

Editing Variable Format

The Edit button at the bottom of the Edit Variable Window is available when the valid values and their labels may be edited. Text values for user-specified variables are always numeric. Click the Edit button to change the list of valid numeric codes and their labels on the Edit Format Window.

You must assign one label per valid value. All possible values not included in the list will be considered invalid by SEER*Prep. You cannot assign a label for a range of valid values in SEER*Prep. Instead, assign individual labels to each of the valid values while in SEER*Prep and create the database. Once you are in SEER*Stat, create a user-defined variable in the dictionary of the newly created database which makes the groupings you wanted and has a sensible name. Please see the SEER*Stat Help for instructions.

Edit Format Window

The Edit Format Window displays the variable name and its length at the top of the window. The rest of the window consists of a text box containing one valid value, an equal sign, and a label per line. For example:

1=Brown

2=Blue

3=Green

Valid values must be numbers from 0 to 9 or 99 or 999 depending on the length of the variable. In the input files, these numeric fields must be padded with zeroes on the left. However, zeroes are not required in the valid values list.

Labels may have any meaningful text including numbers, letters, and special characters. They must also be a reasonable length. If the label text wraps to the next line, an error will occur.

Value order is not important to the SEER*Prep create file process. However, the order of the values in SEER*Prep will also be the order in SEER*Stat. For example, ordering the list above alphabetically ensures that the labels will appear alphabetically in SEER*Stat without changing the data in your input files.

2=Blue

1=Brown

3=Green

Verify Data Process

Once you have selected the input files, the variables, and the study cutoff date for your SEER*Stat database, you are ready to verify your data. Although verifying your data is not required before creating a database, it is highly recommended. During the verify data process, SEER*Prep checks the contents of every variable you have defined for your database in the input files and creates a report which is displayed in your default text editor.

Why Verify Your Data

You should verify your data to eliminate possible problems before they happen. The following is a generic list of possible problems. You may find others that are specific to your input data.

- The input files are not set up according to the file layout.
- Invalid values are numerous and need investigation.
- High number of blanks in key fields need investigation.
- Variable values in the case files and population files do not match.

All of these possible problems can be detected by either looking through the Verify Data Report or by comparing the Verify Data Report to frequencies created outside of SEER*Prep.

Verifying Your Data

To verify your data, follow these steps.

1. Verify that your database description includes all input files and variables you want in your database. Only variables which are selected for the database will be checked by SEER*Prep.
2. Select Verify Data from the Execute menu or click .
3. Once the Save Report As window appears, enter the file name for the Verify Data Report and click Save.
4. The Verify Data window will appear. Since you cannot disturb SEER*Prep while the verify data process is taking place, you may want to minimize the application. SEER*Prep uses a lot of system resources, so your system may perform poorly in other applications while the data is being verified.
5. When the verify data process is complete, SEER*Prep will be restored and the Verify Data Report will appear in your default text editor.
6. Check the Verify Data Report.

- Are all the input files you want to use listed in the report?
- Do the frequency counts for certain key variables agree with your expectations?
- Are there any invalid values? Should there be any?
- Are there more blanks than you expected?
- Are there demographic values in the case file that do not appear in the population file or vice versa?

7 Once you are satisfied that the input files will produce the database that you need, you are ready to create the database.

Verify Data Window

While the verify data process is executing, the Verify Data Window appears. One or two progress meters are provided to track the completion of the executable steps which make up the verify data process. The Current Operation progress meter tracks the percent of records in the case or population file completed. The Overall progress meter indicates what percent of the entire verify data process has been completed.

It is easy to misinterpret the progress meters, for they are not based on time. Instead, the Current Operation progress meter reports the percent of records processed and the Overall progress meter reports the percent of completed steps in the verify data process, not time passed. For example, if the Overall progress meter indicates that 50% of the job steps have completed 10 seconds after you began the verify data process, this does not mean that there is only 10 seconds left in the job. Actually there could be either more or less time left. You cannot tell the time with the Overall progress meter.

Verify Data Report

The Verify Data Report can be used as a diagnostic tool to check the database description settings and the data in the input files. In the heading, it lists the following information.

- Full path name of the database description
- Month and year of the study cutoff date
- Full path name of all input case files
- Number of case records processed
- Full path name of all input population files
- Number of population records processed

After the heading, the report continues with a frequency of both valid and invalid values for each database variable. Remember that records with invalid values are either excluded from the database, or all invalid values are recoded to “Invalid value” depending on how your database is set up. Blanks are not considered invalid values, except in the population count variable. The database variables are presented alphabetically within this order.

1 Population file only variables

2 Population and case file standard variables

3 Population and case file derived variables

4 Case file only standard variables

5 Case file only derived variables

Create File Process

Once you have set up your database description and verified the accuracy of your data, you are ready to create a SEER*Stat database. During the create file process, SEER*Prep converts your input files into a single database which is compatible with SEER*Stat and creates a report which is displayed in your default text editor. Once the database has been created, you can use it in SEER*Stat.

Creating a Database File

To create your database for use with SEER*Stat, follow these steps.

1 Select Create File from the Execute menu or click on .This will bring up the Save Report As window.

2 Enter the file name for the Create File Report. Click Save. The Record Exclusion Window will appear.

3 You must choose whether to treat invalid values on a variable by variable basis or globally. When invalid value(s) are found in the input files, SEER*Prep will either recode the them to “Invalid Value(s)” or exclude the entire record. To discover if you have invalid values in your input files, exit the create file process and select Verify Data from the Execute menu.

4 Once your selection has been made, SEER*Prep will begin creating your database. Since you cannot disturb SEER*Prep while the create file process is taking place, you may want to minimize the application. SEER*Prep uses a lot of system resources, so your system may perform poorly in other applications while the database is being created.

5 When the create file process is complete, SEER*Prep will be restored and the Create File Report will appear in your default text editor.

6 You may now use your database in SEER*Stat.

Record Exclusion Window

After initiating the create file process but before actually creating the database, you must choose whether to treat invalid values on a variable by variable basis or globally on the Record Exclusion Window. Select one of the following choices.

Exclude as specified for individual variables

By choosing this option, each variable will be handled on an individual basis. Variables that are marked to remove records with invalid values have a “Yes” in the Invalid Values column of the Variables Table. To edit the setting of any variable, see Dealing with Invalid Values in a Variable

Exclude for all invalid values

This option causes all records with an invalid value for any variable to be excluded from the database no matter how the variables were set up in the database description.

Do not exclude for any records

When this option is chosen, SEER*Prep creates an additional value for each variable which has invalid values named “Invalid Value(s).” All invalid value(s) in the database are recoded as an “Invalid Value(s)” regardless of specifications made in the database description for any one variable. The population count variable is the one exception, since all records with invalid population count are automatically excluded.

Create File Window

The Create File window has two progress meters which track the completion of the numerous steps in the create file process. The Current Operation progress meter names the current job step and reports the percent completed. The Overall progress meter indicates what percent of the overall job steps have been completed.

It is easy to misinterpret the Overall progress meter, for it is not based on time as one might assume. Instead, the Overall progress meter reports the percent of job steps completed, not the time any one of the jobs will take. For example, if the Overall progress meter indicates that 50% of the job steps have completed 10 seconds after you began the create file process, this does not mean that there is only 10 seconds left in the job. Actually there could be either more or less time left. You cannot tell the time with the Overall progress meter.

Create File Report

The Create File Report lists the following items about the newly created database.

- Name and path of the database
- Month and year of the study cutoff date
- Notes or warnings concerning the database
- Full path name of the input case file and the input population files
- Number of case records and population records processed
- Invalid values exclusion option which was chosen
- Number of case and population records written out
- Number of records with invalid values for each variable and whether they were excluded (Notice that once a record has been excluded, it will not show as invalid for additional variables.)

Pay attention to the notes or warnings.

If a population and case file variable has values found in one but not the other, your rates in SEER*Stat could be wrong. For example, if in your case file you have only female breast cancer patients, but you are using a standard population file, the rate calculations may be low because they include men in the population. To avoid this particular situation, it is recommended that you create a subset of the population file containing only women to be associated with your case file in SEER*Prep. It is possible to select females alone on the Selection Tab in SEER*Stat every time you do statistics with the database, but this method leaves you open to mistakes.

Accessing Your New Database in SEER*Stat

Telling SEER*Stat About Your New Database

SEER*Stat accesses databases from two locations: Primary Data Location and Secondary Data Location. The path for each of these data locations is specified in the SEER*Stat Preferences window.

SEER*Prep Preferences includes a "Directory to Save User-Defined Databases". New databases are created in this directory. SEER*Prep will automatically set SEER*Stat's Secondary Data Location to this directory when a new database is created. Therefore, when you create a database in SEER*Prep you will immediately be able to use that new database in SEER*Stat (if SEER*Prep and SEER*Stat are used on the same computer). If you have saved the database on a shared drive and would like to use the database in SEER*Stat from another computer, simply set SEER*Stat's Secondary Data Location appropriately.

Note

Every time you change SEER*Prep's "Directory to Save User-Defined Databases" and create a database, the Secondary Data Location in SEER*Stat will be changed. If you store databases in multiple locations, you will need to set SEER*Stat's Secondary Data Location to the directory containing the database that you wish to analyze.

Refer to SEER*Stat's help system for more information on setting preferences in SEER*Stat. To change the "Directory to Save User-Defined Databases" on the SEER*Prep Preferences window, please see help on Setting Preferences.

Selecting Your New Database in SEER*Stat

Once you create a database using SEER*Prep, it is immediately ready for analysis in SEER*Stat. SEER*Prep automatically updates the Secondary Data Location in SEER*Stat Preferences when a database is created. To use your new database, simply begin a new session in SEER*Stat and select it on the Data Tab. You may use the standard millions or expected rates with any database in SEER*Stat.

Note

If a population file is not selected before the database is created, only frequencies and survival statistics can be calculated by SEER*Stat. In order to produce rates and trends, a population file must be included in the database description.

Dealing with User-Specified Variables in SEER*Stat

User-specified variables can be treated just like any other variable in your SEER*Stat database, but it is suggested that you create a SEER*Stat user-defined variable to make it easier to use. When you create a user-defined variable, you will be able to assign a meaningful variable name and create groupings of valid values for your variable.

For instruction to create a user-defined variable in SEER*Stat for your user-specified variable, refer to the SEER*Stat help system under the heading “Edit Variable Window”.

Standard Millions Databases

Standard million data are used to calculate age-adjusted rates in SEER*Stat. Standard Millions databases may include data for many sets of standard million data. For example, the Standard Millions database provided with the SEER*Stat software includes world standard data, 1991 Canadian standard data, and several sets of data for U.S. populations (1940, 1950, 1960,...,2000).

You may create a second Standard Millions database using SEER*Prep. Only two Standard Millions databases are available in SEER*Stat frequency/rate sessions, the one provided with the software and one database that you create with SEER*Prep.

Edit the “year” variable in the Standard Millions database to define up to 100 sets of data. Modify the format to include the values and labels of your data. SEER*Stat will add “ - User” to the end of a label when it is displayed in SEER*Stat frequency/rate sessions.

Expected Rates Tables

Expected Rates Tables are used to calculate survival statistics. The list of Expected Rates tables in SEER*Stat includes all Expected Rates Tables databases distributed with the software and those created with SEER*Prep. You may also view the data in the Expected Rates tables using Case Listing sessions.

Database Management Utility

When SEER*Prep creates a database, the information about the database is saved in a couple of places within the directory listed on the Preferences Window. Rather than having to find and edit all references to the your created database, SEER*Prep provides a database management utility.

In the database management system, you can either rename a database or delete it. To avoid problems accessing your data in SEER*Stat, you should always rename or delete your databases through the SEER*Prep database management utility.

To access the database management utility, select Database Management from the File menu.

Database Management Window

The Database Management window lists the databases you have created within the directory listed on the Preferences window. The database files in this directory can be made available to SEER*Stat if the Secondary Data Location on the SEER*Stat Preferences Window shows the same directory. To change to a different directory, see Setting Preferences. When you select a database from the list you may either rename or delete it. When you are finished editing your database list, you should close the window.

Renaming an Existing User Database

There are times when it is useful to change the name of a SEER*Stat database which was created by SEER*Prep. For example, suppose you are in the process of creating a database for SEER*Stat in SEER*Prep named “Current Data.” (By the way, this is not a very good name because it isn’t very descriptive.) The last database you made was called “Current Data” as well, but now the old database is not the most current data. Since you are not ready to delete the old database, you may instead change the name to “July Data.”

Any saved SEER*Stat sessions which use a particular database will continue to use that database no matter the name change. In addition, user-defined variables saved to the database will still be available to that database alone, regardless of the name.

To rename an existing User Database, follow these steps.

1 Select Database Management from the File menu.

2 Click on the database you want to rename.

3 Click the Rename button.

4 Enter the new name for the database. The new name must be unique within the directory of databases.

5 Click OK.

Deleting an Existing User Database

You should delete a SEER*Stat database created by SEER*Prep when you don't need it any longer or when corrections need to be made. If corrections are necessary, you must delete the current version of the database and replace it with a corrected version.

To delete an existing User Database, follow these steps.

1 Select Database Management from the File menu.

2 Click on the database you want to delete.

3 Click the Delete button or press the Delete key on your keyboard.

4 Click OK.

Index

.

.dd file 4, 11
.txd file 4, 8, 12

A

About SEER*Prep 3
About the National Cancer Institute.... 3
Acknowledgements 3
add button 14, 16
Adding an Input File....14
Adding Database Variables 16
age-adjusted rates....4, 8, 12, 27
ASCII data....4, 8

B

Before Using SEER*Prep 4
binary data 8

C

calculations - SEER*Stat 4, 11, 12, 27
case file....12
case listing 3, 4, 27
case starting column....10, 15, 20
Cause of Death Variables....17
COD....17
Context-Sensitive Help 7
Create 15, 18, 20
Create File Process 24
Create File Report 26
  viewing 6
Create File Window 25
created database.... 4, 11, 24, 26, 27, 28
Creating a Database File 24
custom race variable 8, 19

D

data....12, 27
  preparation 4, 8, 17
  problems....22, 23
database creation process....4, 24, 26
database description 4, 11
  components 5, 8, 11, 12, 13, 14
  editing....4, 5, 11
  opening....8, 11
  saving....11
  templates....4, 8, 11
Database Description Files....11
Database Descriptions....11
database directory....6, 12, 26, 27
database management....28, 29
Database Management Utility....28
Database Management Window 28
Database Name 12

Database Types 4
Database Variables 14
Dealing with Database Variable Indexes 20
Dealing with Invalid Values in a Variable 21
Dealing with User-Specified Variables in SEER*Stat 27
default text editor 6
Deleting an Existing User Database 28
Derived Variables 18
description 11, 19, 21
documentation - input files 9, 10, 12

E

EBCDIC data 8
edit button 13, 17, 20, 21
Edit Format Window 21
Edit Variable Window 20
  description 21
  format 19, 21
  indexed 18, 20
  invalid values 19, 21, 22, 25
Editing a Database Description 11
Editing Database Variables 17
Editing Variable Format 21
editor - default text 6
e-mail address 7
errors 7
  database name 12, 28
  directory 6, 8, 12
  input file 8, 11, 12, 13
  record length 4, 8, 12
  study cutoff date 11, 13, 14
  variables 14
excluded records 13, 18, 21, 24, 25, 26
execute menu 4
  create file 24, 25
  verify data 22, 23
Expected Rate Location 13
expected rates table 5, 8
Expected Rates Tables 27

F

file documentation - input 9, 10
file format 8, 16, 17
  editing 19, 21
File Layouts Supported by SEER*Prep 8
file menu 4
  database management 28
  generate input file description 10
  open 5, 11
  preferences 6
  save/save as 11
  view file description/report file 6, 10, 24, 26
files - input 8, 12, 17

G

Generating the Input File Description Report 9

Getting Help 7

H

hardware requirements 3
help 7, 8

I

ICD Coding 17
Indexed Variables 18
Input Case Files 12
Input File Description Report 10
  components of 10
  creating 9
  viewing 6
Input File Requirements 8
input files 4, 8
  adding 14
  directory of 6, 8, 12, 14
  documentation for 6, 9, 10
  layout of 4, 8, 10, 17
  listing values of 6, 10, 24
  name requirement for 4, 8, 12
  removing 14
Input Population Files 12
invalid values 10, 15, 18, 19, 21
  detecting 22, 23, 24
  expected rate variable 13
  population count variable 13
  standard million variable 13
  treatment of 15, 20, 21, 24, 25
  user-specified variable 19, 21, 25

M

main screen 5, 15
Mortality file layout 8, 17
Moving from SEER*Prep to SEER*Stat 5

N

NAACCR file layout 9
name of database 12, 28
National Cancer Institute (NCI) 3

O

Opening a Database Description 11
Ordering the Variables Table 15
output database 6, 11, 24, 26, 27, 28
Overview 4

P

packed decimal data 8
population count variable 13, 19
population file 4, 5, 8, 10, 12, 13
  calculations 4, 5, 11, 27
  count variable 13
  documentation 8, 10, 13, 15
  errors 22, 23, 24, 26

requirements....8, 12
population start column....12, 15, 20
Population/Count Location....13
preferences....6, 26
progress meter....23, 25

R

race - custom variable....8, 19
recode invalid values....13, 15, 18, 19, 20, 21, 22, 23, 24, 25, 26
record exclusion....13, 19, 20, 21, 22, 23, 24, 25, 26
Record Exclusion Window....25
registry variable....8, 9, 19, 21, 27
remove button....14, 16
Removing an Input File....14
Removing Database Variables....16
Renaming an Existing User Database....28
Reports....6, 9, 10, 24, 26
required....10, 14, 15, 18
Required Variables....18

S

sample database descriptions....8, 11
Saving a Database Description....11
SEER 250 file layout....8
SEER program....3
SEER website....3
SEER*Prep....3, 4, 5
SEER*Prep Main Screen....5
SEER*Prep Reports....6
SEER*Stat....3, 4, 5, 8, 18, 26, 27
SEER*Stat database....5, 26
absent variables in....16
accessing....4, 5, 27
directory of....4, 5, 6, 28
editing....4, 11, 28
renaming....11, 12, 28
user-defined variables in....27
Select Variable(s) to Add Window....16
Selecting Database Variables....15
Selecting Your New Database in SEER*Stat....27
Setting Preferences....6
standard million....4, 5, 8, 18
Standard Million Location....13
Standard Millions Databases....27
Standard Variables....18
Study Cutoff Date....13
survival calculations....4, 8, 13, 27
system resources....3, 24

T

Technical Support....8
Telling SEER*Stat About Your New Database....26
text editor....6, 7
toolbar buttons....3
create....24, 25
help....7

open 11
save 11
verify 23

U

user-defined variables - SEER*Stat 27
user-specified variable 14, 19, 21, 27
editing 17, 20, 21
making user-defined 27
name 21
requirements 9, 10, 19
User-Specified Variable Requirements 19
User-Specified Variables 19
User-Specified Variables and Invalid Values 19

V

variable - 13, 14, 15, 16, 17, 18, 19, 20
adding 16
availability 16, 27
components of 15, 20
derived 10, 13, 14, 17, 18
description 10, 21
editing 14, 15, 17, 20, 21
indexed 10, 15, 18, 20
removing 16
required 10, 14, 15, 18
selecting 15, 16
sorting 15
standard 18
user-specified 14, 17, 19, 21, 27
Variable Description 21
Variables Table 15
adding variables 16
selecting variables 15, 17
sorting variables 15
variables values 9, 10, 17, 20, 24
blanks 13, 18, 19
creating groupings 21, 27
editing valids 19, 21
invalid 10, 13, 18, 19, 21, 25
listing 10, 23, 24
Variables with Invalid Values 18
Verify Data Process 22
Verify Data Report 24
checking 22, 23
viewing 6
Verify Data Window 23
Verifying Your Data 23
view file description/report 6, 10

W

web sites 3, 7
What You Will Need 3
Why Verify Your Data 22
window 5, 11, 15
create file 24, 25, 26

database management 28
default text editor 6, 9, 10, 24, 26
edit format.....17, 19, 21
edit variable 17, 20, 21
expected rate location 13
population count location.....13
preferences.....6, 28
record exclusion..... 18, 21, 24, 25
select variables to add 16
standard million location.....13
study cutoff date 13
verify data.....22, 23, 24

Z

zlib 3