LTAS Manual

Table Of Contents

Chapter 1. Introduction & Acknowledgements

The Life Table Analysis System (LTAS) was developed at the National Institute for Occupational Safety and Health (NIOSH) during the 1970's (Waxweiler R, et al., 1983). The original LTAS software was developed on IBM mainframe computer systems. A subsequent version of LTAS, known as PC LTAS, was released for MS-DOS based PC's.

The current version of LTAS, known as LTAS.NET, has been developed for use on Windows 98/NT/2000/XP compatible PCs. This version of LTAS is more interactive and provides more user options than prior versions.

LTAS.NET was developed using Microsoft® Visual Studio® .NET and Microsoft® SQL Server Desktop Edition.

Chapter 2. Overview

Overview

The purpose of this chapter is to provide an overview of the NIOSH Life Table Analysis System and an introduction to the statistics used for analysis. For a more detailed look at the statistics used, please refer to Appendix B - Statistics in LTAS.

Life table analyses originated as a form of survival analysis in which survival times are grouped into intervals. Hazard or incidence rates are calculated for each interval. Intervals are typically constructed by age, race, sex, calendar time, and either duration of exposure or level of exposure (ET Lee, 1980). Rates for the cohort under observation may then be compared with external rates for some large (generally unexposed) population to obtain an estimate of the relative survival of the cohort compared with the external population. The system also allows for internal comparisons of the cohort under study via stratified analysis.

The NIOSH Life Table was created primarily to analyze cohorts defined by occupational exposures. It therefore requires input of a History file in order to calculate cumulative exposure or duration of employment or exposure. The system also requires a Person file containing demographic information and an Outcome file containing data on observed events (deaths or disease incidences). LTAS then generates rates by age, race, sex and calendar time, where rates are calculated as observed events divided by person-time at risk in the interval. In addition, LTAS allows for cross stratification by time since first exposure/employment (TSFE) to account for latency, and by either duration of exposure or cumulative level of exposure to examine exposure-disease relationships. With the updated version of LTAS, (LTAS.NET), the user may opt to stratify on additional user-defined variables in the calculation of rates.

In the calculation of LTAS rates, person-time at risk of death begins when exposure begins and continues for each individual until death, the study end date, or the individual is lost to follow-up, whichever comes first. An individual is "lost to follow-up" if the individual is not deceased and the date last observed occurs before the study end date. Person-time for deceased subjects is truncated on the date of death. Study subjects lost to follow-up may be considered censored or withdrawn, and person-time for that individual truncated at that point. Optionally, the user can choose to extend person-time until the study end date for individuals lost to follow-up.

Observed rates for the cohort under study are compared with the rates from an unexposed or referent population via indirect standardization and optionally via direct standardization. Indirect standardization calculates standardized mortality ratios (SMRs), while direct standardization calculates standardized rate ratios (SRRs). LTAS also has the capability to calculate proportionate mortality ratios (PMRs), which compare the proportion of deaths from a specific cause in the exposed population to the proportion of deaths from that specific cause in the unexposed population. (For both SMRs and PMRs, the U.S. population is typically used as a referent "unexposed" population).

Further information on the NIOSH life table may be found in Waxweiler R, et al., 1983 and Steenland K, et al., 1990; a useful overview of occupational life table analysis is presented in Checkoway H, 1989.

Standardized Mortality Ratios

Indirect standardization compares observed deaths within each stratum (age, race, sex, calendar time, etc) with expected deaths, where expected deaths are computed by multiplying the referent population death rates (stratified appropriately) by the observed person-years at risk in the stratum. The observed and expected deaths are then summed across all strata, and the standardized mortality ratio (SMR) is the ratio of total observed deaths to total expected deaths for the cause of death category considered.

The formula to calculate the SMR is presented below:

$$SMR = \frac{\sum_i W_i R_{i1}}{\sum_i W_i R_{i0}}$$

where the variables are defined as follows:

- R_{i1} = the stratum-specific rate in the observed cohort (the exposed cohort)
- R_{i0} = the stratum-specific rate in the unexposed population
- W_i = stratum-specific person years in the exposed cohort.

Note that the ratio of observed to expected deaths is equivalent to a ratio of sums of weighted rates, in which the weights for each stratum are the person-years in the exposed group. The numerator is a weighted sum of the observed rates in the exposed population and the denominator is a weighted sum of the rates in the unexposed population.

Two or more indirectly standardized rates are not mutually comparable due to differences in the weighting (resulting primarily from differences in age structure of the populations). This makes comparison of SMRs for different groups, whether cumulative level of exposure groups or duration of exposure groups, unsatisfactory and provides the motivation for using direct standardization.

Standardized Rate Ratios

Direct standardization is used in LTAS for an internal analysis, to compare low duration or low cumulative level of exposure groups to higher ones within the study population. The directly standardized rate ratio (SRR) is a ratio of sums of weighted rates in which the weight for each stratum-specific rate is the combined person-years for the observed cohort across all duration (or cumulative level of exposure) categories. LTAS provides the SRR for each duration (or cumulative level of exposure) group compared with the referent group. The cutoff points for the categories must be specified by the user (for example, 0-100 ppm-years might define the referent group, 100-200 ppm-years might define the next group, etc.). Taylor-series-based confidence intervals (Rothman KJ, 1986) are given for each specific SRR. Trend test results are also provided, where the trend is calculated in a regression of directly standardized rates according to the formulas presented in Rothman KJ, 1986. This trend test uses the midpoints of the categories as the independent variable. The cutoff point plus 50% is used for the midpoint of the highest category.

Proportionate Mortality Ratios

LTAS also calculates proportionate mortality ratios (PMRs), which are ratios of the proportion of deaths from a specific cause in the exposed versus the comparable ratio in the unexposed (often the U.S. population). For example, the proportion of lung cancer deaths among all deaths in the exposed cohort could be compared to the proportion of lung cancers deaths among all deaths in the unexposed cohort. SMRs and SRRs are preferred measures of effect; however, the PMR is useful when death data are only available among an exposed population (the incomplete enumeration of deaths among the entire cohort prevents the calculation of rates). A typical situation might arise when death certificates are available from a union, but no complete list of all union members is available. For PMR analyses to be valid, the available observed deaths must be representative of all deaths in the base population (Checkoway H, 1989).

Adjustment for age, race, sex, and calendar time is accomplished (as in SMRs) by stratification and indirect standardization (a weighted average of proportions is calculated for exposed and unexposed across strata of the stratification variables, with the weights being the observed deaths in the exposed population in each stratum). Hypothesis testing and confidence intervals are calculated as in the SMRs, based on an assumed Poisson distribution. Standard proportion rates for the U.S. population (stratified like the rate files by age, race, sex, calendar time) are available with LTAS. PMR runs are chosen in LTAS via selection of proportion rates instead of standard rates.

Note that cause-specific PMRs have the disadvantage of being mutually dependent, unlike SMRs. An elevation of the proportion of death from one cause in the exposed population must in turn result in a decrease in the proportion of deaths from some other cause. This interdependence of cause-specific results may be overcome by an internal analysis in which the data are analyzed as a case-control study, calculating mortality odds ratios (MORs) (Miettenin O and Wang J, 1981). While LTAS does not do this directly, MOR may be calculated easily using the stratified data on the observed deaths in the exposed population.

Some investigators may wish to calculate proportionate cancer mortality ratios (PCMR) which are the ratios of observed cancers to expected cancers, where the expected cancers are calculated based on the proportion of cause-specific cancers among all cancer deaths, rather than all deaths as in the PMR. Cause-specific PCMR can be calculated from the corresponding cause-specific PMR by dividing the cause-specific PMR by the all-cancer PMR. For example, suppose there are 20 observed lung cancers in the cohort, the PMR is 2.0, and the all-cancer PMR is 0.80. Then the lung cancer PCMR is 2.0/0.8=2.5. Exact or approximate confidence intervals for this PCMR can be derived based on the Poisson distribution. A test-based confidence interval for this PCMR can be derived using the normal approximation to the Poisson distribution. First, determine the expected cancers from the PCMR and the observed cancers. For example, the expected lung cancers in our example are 20/2.5=8. Then derive the approximate lower and upper confidence limits for the observed lung cancers via the method described in Vandenbroucke P, 1982,

$$limits = [\sqrt{observed} \pm (1.96 * 0.5)]^2.$$

These limits are in turn divided by the expected cancers to get the lower and upper limits to the PCMR. In our example, these are 1.52, 3.71.

Lagging Exposures

Another feature of LTAS is the ability to "lag" exposures. A lag assumes that an exposure requires a minimum induction period (latency) before it can cause disease or death from disease, which is termed a "lag" period. An investigator may wish to impose a lag such that more recent exposure does not contribute to the cumulative level of exposure.

The lag is a moving exposure blackout window which discounts any exposure occurring for a specified amount of time prior to the time point being considered. In the stratify step, when the user specifies a nonzero lag time, this length of time is used to lag the exposure. As a worker moves through time, at any given point of time x, exposure which occurs during a specified lag (e.g., 10 years) prior to time x is ignored when calculating cumulative exposure at time x. Such a worker (and all their person-time) would be considered unexposed until he/she had been followed for 10 years after first exposure, at which time his/her exposure would start being accumulated. Note also that at the end of follow-up for this worker a 10 year lag would mean that no exposure in the 10 years prior to his/her end of follow-up would be counted in his/her final cumulative exposure.

For more information on how LTAS handles lagging of exposures, please refer to Appendix A - LTAS Date Handling.

Chapter 3. Getting Started

This section is intended to help prepare and plan for the installation and use of LTAS.NET.

System Requirements

Your computer must meet the hardware and software requirements before you attempt to install LTAS.NET. While these requirements are more significant than those of PC-LTAS, most desktop PCs in use today should be capable of running LTAS.NET.

Hardware Requirements

This table shows the hardware requirements for installing and running LTAS.NET.

Hardware	Minimum requirements
Computer	Intel Pentium II or compatible, 166 MHz or higher. Intel Pentium 4 or higher is recommended.
Memory (RAM)	512 MB
Hard disk space	300 MB

Operating System Requirements

To use LTAS.NET, you must have one of the following operating systems installed:

- Windows Server 2003 Standard Edition, Windows Server 2003 Enterprise Edition, Windows Server 2003 Datacenter Edition.
- Windows 2000 Server, Windows 2000 Advanced Server, Windows 2000 Datacenter Server.
- Windows NT Server 4.0, SP5 or later; Windows NT Server 4.0, Enterprise Edition, SP5 or later;
- Windows NT Workstation 4.0, SP5 or later.
- Windows XP Professional, Windows XP Home Edition.
- Windows 2000 Professional. Windows Millennium Edition.
- Windows 98. Windows 98 Second Edition is required if the computer does not have a network card.

Note: The Windows Vista operating system is not currently supported.

Performance

There are three processes involved in running LTAS.NET: importing data, stratifying data, and analyzing data.

The **Import** and **Stratify** processes take more time to run than the **Analyze** process in LTAS.NET.

Because LTAS.NET requires both a large amount of CPU (computer processing time) and I/O (input from and output to the disk drive), the faster the CPU and hard drive, the better the performance of the system will be.

The performance of the **Import** process depends on the size of the import files. LTAS requires three import files to analyze study data: a person file (containing demographic data), a history file (containing exposure data), and an outcome file (containing death or disease incidence data). Generally, larger import files increase the run-time of LTAS. The history file potentially has the largest impact, due to the number of records and the number of stratifiers defined.

The performance of the **Stratify** process also depends largely on the history file. Again, the number of history records and the number of stratifiers have the largest impact on process time.

Both the **Import** and **Stratify** processes have a progress dialog to allow the user to gauge run time. The dialog shows the percent completion, time elapsed, number of input records processed, and processing rate which is displayed in persons per second.

The performance of the **Analyze** process is subject to the size of the report being generated, which is a function of the number of major/minor category selections made.

First Time Installation Requirements

Administrative Rights

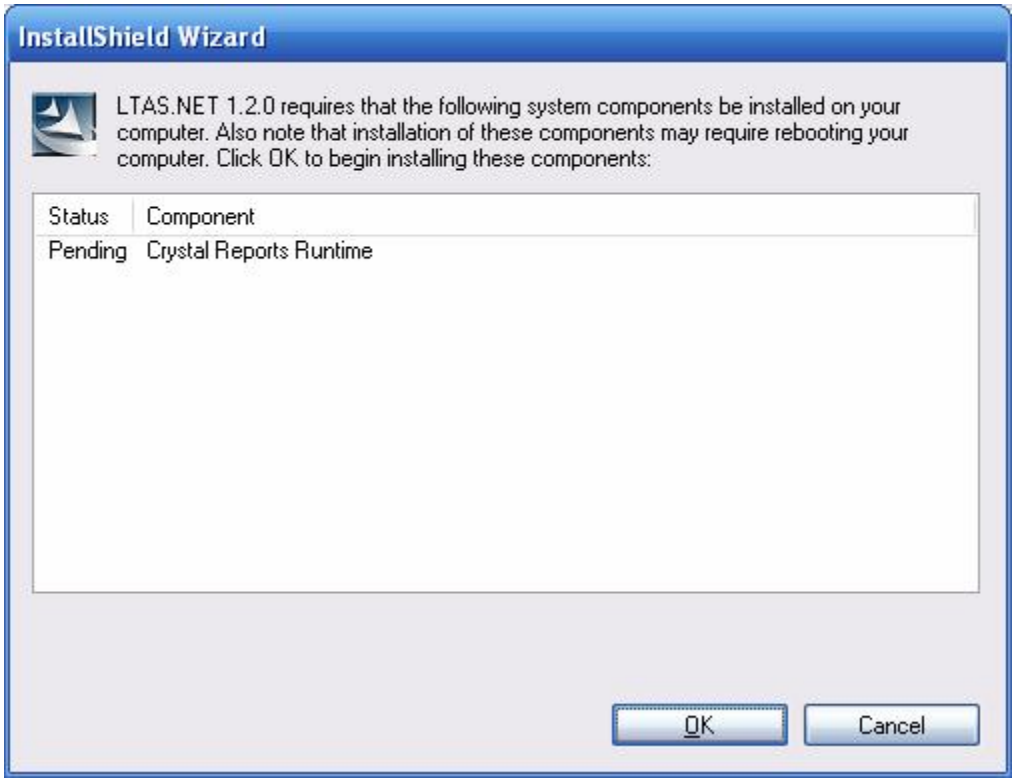
Installing LTAS.NET for the first time requires administrative rights to the computer on which it is being installed. If installing LTAS.NET for use by a non-administrative account, then that account must be temporarily granted administrative rights for the initial LTAS.NET install. Once the initial installation is complete, those rights can be removed from the account. Subsequent installations and upgrades to LTAS.NET on the same computer will not require administrative rights.

Installation of System Components

LTAS.NET depends on several system components that must also be installed on the computer. The first time that you run the LTAS.NET installer, it will check and make sure that these components are installed:

- Microsoft Installer 2.0
- .NET 2.0 Framework
- MS SQL Server 2000 Desktop Engine
- Crystal Reports Runtime

If any of these components are missing, the LTAS.NET installer will install them for you by first displaying the following dialog:



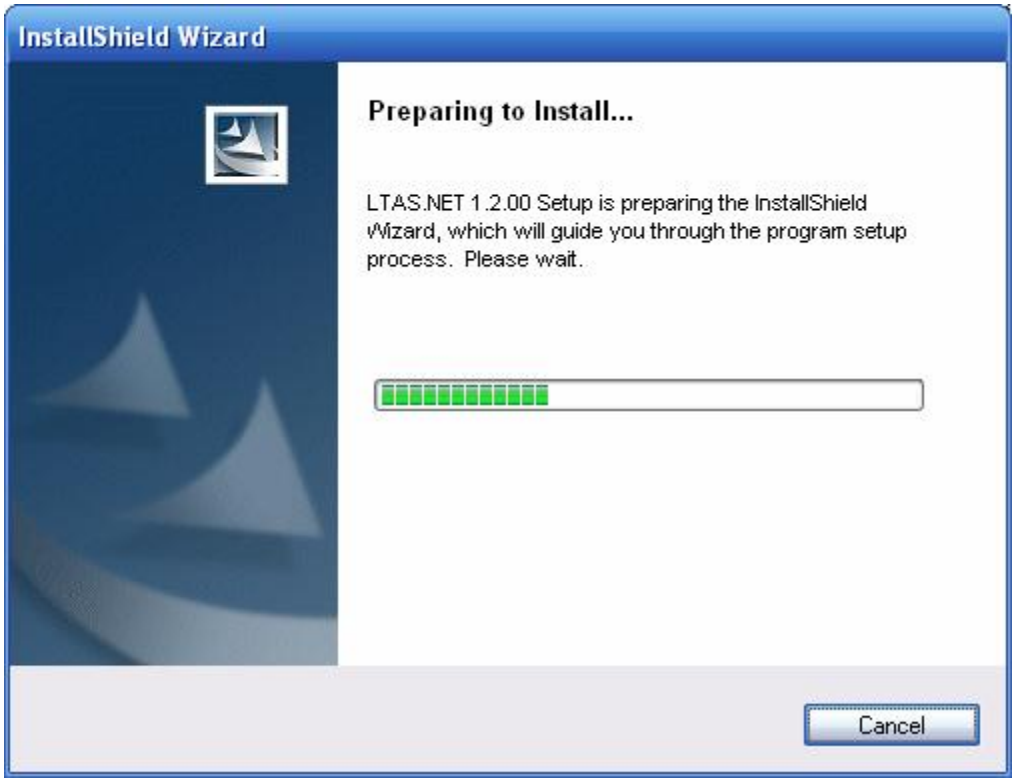
Click the 'Install' button to begin installation of the missing components.

Installation of these components may require one or more reboots of your computer. After a reboot completes, the LTAS.NET installer will continue with the installation.

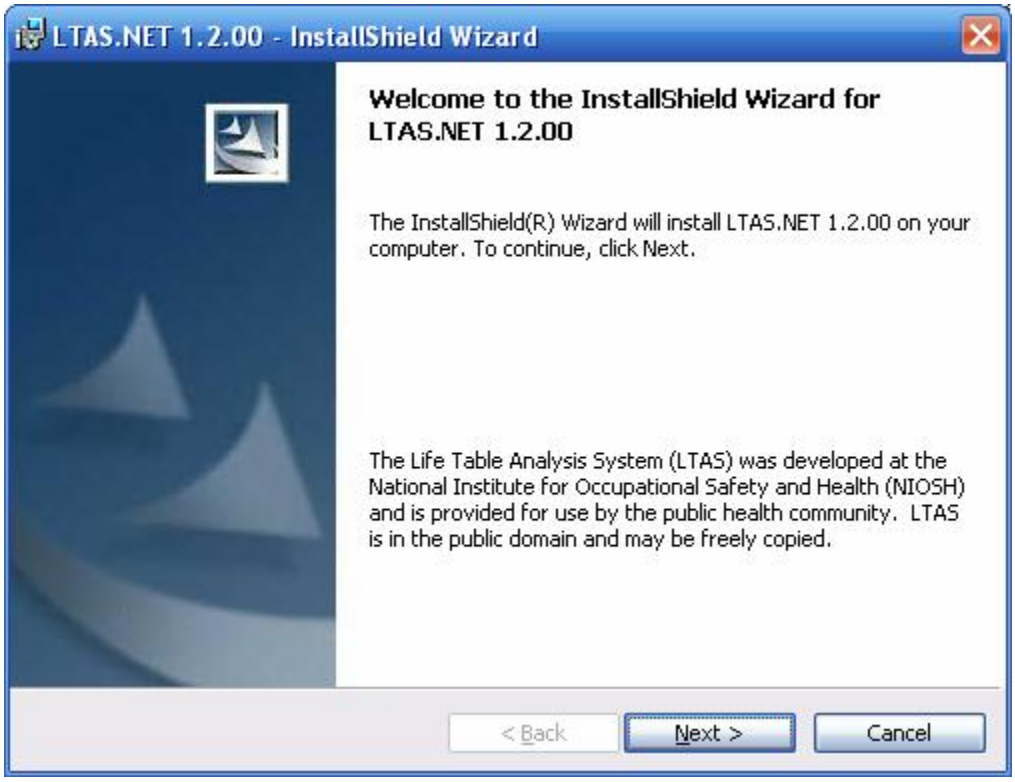
Once these components are installed, the LTAS.NET installer can begin installing LTAS.

Installation

To start the LTAS.NET install, double-click on the LTASsetup.msi file located in the installation folder. As the LTAS.NET installer loads the following dialog will be displayed.



If this is the first time that LTAS.NET has been installed on the computer, additional packages may be required to be installed first. Once these requirements have been met, the LTAS.NET installer uses a wizard interface to walk the user through the rest of the installation process. Below is the display for the first step.



To continue to the next step, click the 'Next' button. Later steps will allow the user to control how LTAS.NET will be installed, for example the selection of the folder where LTAS.NET will be installed. Although not required, we recommend using the defaults given for each of these steps. After the final step, LTAS.NET will be installed.

If you change your mind and do not wish to install LTAS.NET, you can click the 'Cancel' button to exit the installer.

Chapter 4. Creating LTAS.NET Import Files

Understanding your study data in relation to LTAS import files

The user must create import files of the study data to run an LTAS analysis. Required variables for LTAS are date of birth (for calculation of calendar year and age), race and sex. These stratifiers are used for rate standardization. However, the user can opt to include additional stratifiers to control for potential confounding.

There are 3 types of LTAS import files which house the study variables: the Person file, the History file, and the Outcome file. At a minimum, the person file contains data on age, race, and sex (required variables), and any fixed user-defined covariates of interest. The data for any time-dependent variable(s) of interest are contained in the History file (required), while the Outcome file (required) contains the study data related to deaths or disease incidences in the cohort.

An important feature in LTAS is the ability to control for confounders other than age, race, sex and calendar time in analysis. To determine which import file is appropriate to hold the information for a particular confounder, the user must first consider whether the confounder is a fixed or temporal variable: the Person file accommodates fixed stratifiers, while the History file accommodates temporal stratifiers.

For more information on LTAS variables, see the Variable Specification section found later in this chapter.

Import File Formats

Each of the three import files requires a personal identifier to allow for linkage of the files. A detailed description of these import files and their contents can be found in the following sections:

- Person Import File
- Outcome Import File
- History Import File

Variable Specification

The independent variables for a study may be classified for the purposes of LTAS as intrinsic or extrinsic. Variables may be further characterized as fixed or temporal, where the latter are subcategorized as either level or categorical. Additionally, categorical temporal stratifiers can be broken down into 2 groups: personal or global.

Intrinsic vs. Extrinsic Stratifiers

Intrinsic

Intrinsic stratifiers are the required stratifiers needed to import data into LTAS and to perform an analysis. These built-in variables are age, race, sex and calendar period. Note that referent rates are also stratified on these variables. If the user has referent rates that contain other stratifiers (e.g., smoking information), then these variables would also be considered intrinsic stratifiers.

Extrinsic

Extrinsic stratifiers are those that are not required or built-in to the system. Extrinsic stratifers are study-specific variables that are user-defined.

Fixed vs. Temporal Stratifiers

Fixed

Fixed stratifiers are those that do not change over time. For example, sex and race are considered fixed intrinsic stratifiers (they are intrinsic because they are required by the system). Examples of extrinsic (user-defined) fixed stratifiers are age-at-first-exposure and ethnicity. Note that variables that are not fixed by nature can be treated as fixed if the data include only one unchanging value per person. For example, the study data may contain information as to whether a person has ever smoked tobacco, but no indication of when he/she started or stopped smoking. Such a variable could be considered an extrinsic fixed stratifer.

Temporal

Temporal stratifiers are those that change during a person's lifetime. Among the intrinsic (built-in) stratifiers, age and calendar period are temporal stratifiers. The user may also define additional time-dependent variables specific to the study (i.e. extrinsic temporal stratifiers).

Temporal variables can be further characterized as level or categorical. Temporal categorical variables can be even further classified as personal or global. These relationships are defined in more detail below.

Extrinsic Temporal Stratifier	Description
Level	Level temporal stratifiers are those that change categories when a specified quantity has been accumulated. The quantity can be a number of days as in Age or Duration of Exposure ("monotonic" accumulators), or a quantity of exposure units as in Level of exposure ("weighted" accumulators). Level temporal stratifiers are calculable based on a reference date and a rate of accumulation (e.g. age is calculable based on Date of Birth; Time-Since First Exposure is calculable from date first exposed)
Categorical	Categorical stratifiers are temporal stratifiers that change categories at an unpredictable or incalculable date. Date(s) for categorical stratifiers may be person-specific (personal), or global.

The two kinds of categorical stratifiers are defined below:

Categorical Stratifier	Description
Personal	Personal temporal categoricals change at different dates for each person in the cohort. Marital Status would be an example, as would State of Residence and Job Title. Smoker Status might be a personal temporal categorical stratifier, if the study data include dates of starting and stopping smoking but not quantity measures such as pack-years.
Global	Global categoricals change on the same date for everyone in the cohort. Perhaps a new manufacturing process or technology was introduced, a plant relocated, new management began or a new regulation took effect. Global categoricals could also be used to group or un-group calendar periods (eg. change from five year categories (the default) to one year or ten year groups).

Import Files and Formats

In LTAS, the Import process transforms your cohort data into the LTAS data model from the following required import files:

Import File	Description
Person	Each record in this import file represents one person in the cohort. May contain extrinsic (i.e., those that will not be used to standardize) fixed stratifiers.
Outcome	Each record in this import file represents a ‘failure’ (death or diagnosis) in the cohort.
History	Each record in this import file represents a person's exposure to a suspected disease agent or other extrinsic temporal stratifier during a specific period of time. Employment period is a common example of an extrinsic temporal (time-dependent) stratifier within the History file.

LTAS contains a mapping feature in the import wizard to allow for definition of various file layouts as long as the data are provided in one of the following two file formats:

File Format	Description
Fixed ASCII	The file contains only text characters. Variables are of a fixed length and always begin at the same position within the file. Typically, but not necessarily, the file name extension is TXT or ASC.
Delimited ASCII	The file contains only text characters. Variables are separated by a delimiting character such as a semicolon, comma or tab. This format can often be produced by exporting data from a spreadsheet (such as Microsoft Excel) or database program. The file name extension is commonly CSV.

Person Import File

Each record in the Person File represents one person in the cohort. ID, Gender, Race, Vital Status, and DOB are always required fields. Risk Begin Date is conditionally required depending on your import options. Name, Date Last Observed, and Fixed Statifiers are optional fields. All dates must be a string with a 4-digit year field. The field is either 8 or 10 characters in length, depending on whether separators are included. Thus the format is either MMDDYYYY, or MM/DD/YYYY. (If the field is 10 characters long, the 3rd and 6th character will be considered a separator, regardless of which character is used.)

The Person File must be sorted by ascending ID before importing.

Below are descriptions of the Person File fields.

Field	Description
Person ID	Contains a unique code identifying each person. Often this is an employee ID or other unique identifier. LTAS maintains up to 15 characters for this ID. This field is required.
Name	Contains a string of up to 45 characters that contains an identifier of the person. If included blank names are allowed. This field is optional.
Gender	Contains a code indicating the sex of the person. A single unique code must represent each of the genders in the study cohort. These codes can be any character(s): male could be represented by '0', '1', 'M', 'Male', '99' or any single code distinct from the code for female. During the import process, these codes will be translated to the code used internally by LTAS to represent genders. This field is required.
Race	Contains a code indicating the person's race. Like gender, a single unique code must represent each of the races in the study. NIOSH standard rates include the race groups 'White' and 'All Other Races', so to use the provided rate files the Person File must include unique codes for 'White' and 'All Other Races'. This field is required.
Vital Status	Contains a code indicating whether the person is alive or deceased. Like gender and race, a single unique code must represent 'Living' and another code must represent 'Deceased'. For PMR-type studies, all persons in the cohort that are not deceased will be rejected from the study. This field is required.
DOB	Contains a date indicating the person's date of birth. This field is required.
Risk Begin Date	Contains a date that determines when to begin accumulating time at risk for each person in the cohort. See the Time At Risk section for more information. This field is conditionally required.
Date Last Observed	Contains a date that indicates the date of death for deceased persons in the cohort. For living persons it is the date the person was last confirmed alive or lost to follow-up. This field may be left blank for living persons if the "Stop risk at DLO" option is not used. This field is required to be mapped even if all records are for living persons and have a blank value.

Fixed stratifier	Contains a character string representing the value of a user-defined fixed stratifier. To determine which values are distinct, the LTAS import process first strips leading and trailing spaces and performs a case insensitive comparison. Leading and trailing zeroes and any embedded blanks will be considered significant. LTAS maintains up to 15 characters. This field is optional. Note: you may specify more than one fixed stratifier.

Below is a suggested data layout for the Person File (fixed ASCII) which may reduce time spent verifying field mapping; however, the mapping function may be used to accept alternate variable specifications.

Field	Fixed Columns	Comments
Person ID	1-9	Required.
Name	11-30	Optional.
Gender	32	Required.
Race	34	Required.
Vital status	36	Required.
DOB	38-47	Required.
Risk Begin Date	49-58	Conditionally required. Field not necessary for PMR type analyses.
Date Last Observed	60-69	Required field, but may be blank for living persons unless using the option to end risk at DLO.
Fixed stratifier	71-80 82-91	Optional.

History Import File

The History File contains information linking persons in the Person File to temporal exposure data and optionally other time-dependent variables. A record in the History File represents a period of time during which the values of the temporal stratifiers remain constant. Once the value of a stratifier changes, a new history line is needed to document the date of change in the exposure profile. A given person's records in the History File begin at the earliest date of the first exposure and continue through Date Last Observed.

All dates must be a string with 4-digit year field. The field is either 8 or 10 characters in length, depending on whether separators are included. Thus the format is either MMDDYYYY, or MM/DD/YYYY. (If the field is 10 characters long, the 3rd and 6th character will be considered a separator, regardless of which character is used.)

The History File must include the required variables listed (Person ID, Begin Date and End Date) plus at least one Level or one Category variable.

Finally, the records in the History File must be sorted by ascending ID and ascending Begin Date before importing.

Below are descriptions of the History Import File fields:

Field	Description
Person ID	Contains the same values as the Person ID field in the Person File acting as a link between the two files. LTAS requires that the records in the History File be in ascending order by this variable and Begin Date. This field is required.
Begin date	Contains the beginning date for an exposure period. A person may have more than one record in this file, but the dates may not overlap. All records for a given person must be sorted by Begin Date (earliest first). This field is required.
End date	Contains the end date for the exposure period. This date must be on or after the Begin Date of the same record. This field is required.
Level	For each exposure agent, a numeric string indicating the average daily rate of accumulation for that exposure during the time period defined by the Begin and End date. If exposure level data are not available, use a level of '1' to indicate that the period is exposed, or a level of '0' to indicate the period is unexposed. For example, if only employment dates are available, then indicating a level of 1 for date ranges associated with active employment would produce an analysis based on employment duration. If exposure level data are available, the daily level is located in this variable. Exposure levels are "unitless" positive numeric values which can be represented as integers, floating point or scientific notation. Import will provide a mechanism for identifying which Level variable is associated with which exposure agent. This field is required for each exposure agent.
Category	If your data include any categorical temporal stratifiers, the values must be included in the Exposure History File. For example if you wish to stratify on the extrinsic categorical stratifier Smoking Status, the Exposure History File must include the Smoking Status values for each line of history. This field is required for each categorical stratifier.

Below is a suggested data layout for the History File (fixed ASCII) which may reduce time spent verifying field mapping; however, the mapping function may be used to accept alternate variable specifications.

Field	Fixed Columns	Comments
Person ID	1-9	Required.
Begin date	11-20	Required.
End date	22-31	Required.
Level	1 beyond previous field (End Date or Level), 10 cols wide.	Required for each exposure. A single blank column separates Level from the prior variable.
Category	1 beyond previous field (End Date or Level), 10 cols wide.	Required for each temporal categorical stratifier. A single blank column separates Category from the prior variable. Import assumes the first Category to be in the ten columns beginning one beyond the last column of the previous variable.

Outcome Import File

The Outcome File contains information on all 'failures' (deaths or diagnoses) in the cohort. Each outcome record is linked to the Person File by Person ID. A record in the Outcome File represents a single outcome on a single date. An Outcome File may include any or all of the following types of outcomes: disease incidence code, underlying cause of death code or multiple cause of death/disease code.

Depending on the type of rates selected for the study, Import must sometimes determine which types of outcomes are appropriate to be retained. When this is the case, Import must know which outcomes are terminal (resulting in deaths) and which are underlying (direct cause of death/diagnoses). To capture this information, there are two separate fields: terminal and underlying.

All dates must be a string with 4-digit year field. The field is either 8 or 10 characters in length, depending on whether separators are included. Thus the format is either MMDDYYYY, or MM/DD/YYYY. (If the field is 10 characters long, the 3rd and 6th character will be considered a separator, regardless of which character is used.)

The Outcome Import File must be sorted by ascending Person ID and descending Date before importing.

Below are descriptions of the Outcome Import File fields:

Field	Description
Person ID	Contains the same values as the Person ID field in the Person File acting as a link between the two files. LTAS requires that the records in the Outcome File be ordered by this field. This filed is required.
Date	Contains the date of occurrence. If a person has more than one outcome record in the file, those records will be sorted by date within Person ID, earliest first. For studies employing death (non-diagnosis) rates, any outcomes where date is not equal to the person's DLO will be excluded. This field is required.
Disease code	Contains the disease code for any deceased or diagnosed persons. Outcomes in LTAS are generally deaths, but can also be diagnoses if the selected rates also include diagnoses. LTAS translates Outcome codes to LTAS disease categories according to the scheme associated with the selected rates. Outcome codes are typically ICD codes, but LTAS also supports other coding schemes; however, each rate set is linked to a specific disease coding scheme. All the rates provided by NIOSH assume outcomes are coded to ICD specifications. In contrast to PC-LTAS, in LTAS.NET decimals are used for ICD codes where applicable. This field is required.
Terminal	Contains a code indicating whether this outcome is a death or diagnosis. The value 'T' represents terminal (death) outcomes, 'F' represents non-terminal (diagnosis) outcomes. This field is conditionally required.
Underlying	Contains a code indicating whether this outcome is an underlying or multiple cause. 'T' represents underlying and 'F' represents multiple. This field is conditionally required. If a multiple cause of death analysis is requested, the underlying field is required. If both multiple and underlying cause analysis is desired using the same Outcome file then the underlying cause must be flagged 'T' and all multiple causes must be flagged 'F' (i.e., the underlying cause will not be used in a multiple-cause analysis).
ICD Revision	Contains a number indicating the ICD revision used to code this diagnosis. Allows each outcome in the file to be coded to a specified revision. Other ICD revision options include coding all outcomes to a single revision (to be specified by the user during the Import step), or coding each outcome to the revision in effect at the time of occurrence. This field is conditionally required.

Below is a suggested data layout for the Outcome File (fixed ASCII) which may reduce time spent verifying field mapping; however, the mapping function may be used to accept alternate variable specifications.

Field	Fixed Columns	Comments
Person ID	1-9	Required.
Date	11-20	Required.
Disease code	22-26	Required.
Terminal	28	Conditionally required.
Underlying	30	Conditionally required.
ICD Revision	32-33	Conditionally required.

Sample Import Files

The following example illustrates the files that the Import step requires. They are shown here as "word processing tables", but Import expects either fixed or delimited text. That is, the header lines of each table are not included in the files, and the columns here would be separated either by zero or more characters (fixed) or by the delimiter (comma, tab or semicolon). It is not necessary that the fields adhere to the order illustrated in this example, as the order may be defined using the Import Wizard.

Sample Person Import File

ID	Name	Gender	Race	Vital Status	DOB	Risk Begin Date	Date Last Observed
000000001	Adams	1	1	1	10211924	09121943	12311990
000000002	Everett	2	1	2	02231931	01101952	10011989
000000003	Davidson	1	2	2	12111900	06211930	08131970
000000004	Rodgers	2	2	2	04221897	10251916	04141955

Sample Outcome Import File

ID	Date	Cause	Terminal	Underlying
000000002	07111971	090.1	F	F
000000002	10201982	001.1	F	F
000000002	10011989	123.4	T	T
000000002	10011989	122.1	T	F
000000003	02131966	121.1	F	F
000000003	08131970	521	T	T
000000003	08131970	522.4	T	F
000000003	08131970	031.6	T	F
000000003	08131970	010	T	F
000000004	04141955	162.1	T	T

Notes:

1. This Outcome File includes codes for diagnoses, underlying causes of deaths, and multiple causes of deaths.

2. Deaths are identified by the Terminal field. Where Terminal is equal to ‘T’ the outcome is a death, where Terminal is equal to ‘F’ the outcome is a diagnosis.

3. Outcomes where the Underlying field is equal to ‘T’ are underlying causes of deaths. All others are either diagnoses or multiple causes of deaths. An Outcome File of this type could be imported into LTAS regardless of the type of rates selected.

Examples of how to match Different Rates:

1. Underlying cause death rates: Import would skip all records except underlying deaths. You could instruct Import to identify deaths as those records where Terminal is equal to ‘T’, or ignore the Terminal variable and accept all outcomes where Date is equal to DLO as deaths. To identify underlying causes, you could specify that Import use the Underlying variable, or use the first terminal cause for each person.

2. Multiple cause death rates: Import would skip underlying deaths and diagnoses records (which could be identified as Terminal equal to ‘F’, Terminal not equal to ‘T’, Underlying is blank, or Date not equal to DLO).

3. Diagnoses and Underlying cause death rates: Import would skip contributing cause of death records, identified as those where Underlying is not equal to 'T', or a death record is not the first for a given person.

4. Proportions Import will skip all non-terminal records, attempting to include multiple causes of death for a multiple cause study, or only underlying causes of death for a study using underlying cause proportions.

Sample History Import File

Multiple exposures can greatly expand a typical small work history and add significantly to the record count. Suppose the following "raw" job, exposure and marital data exist for subject Adams.

Job History data

ID	Begin Date	End Date
00001	09121943	12311952
00001	11211955	03121960
00001	09191961	12311990

Exposure 1 data

ID	Begin Date	End Date	Level
00001	01011943	12311952	2.5
00001	01011955	12311975	3.33

Exposure 2 data

ID	Begin Date	End Date	Level
00001	01011944	12311953	1
00001	09011956	12311984	1

Marital Event data

ID	Event Date	Marital Event
00001	06151953	Married
00001	02101961	Divorced

To import these data to LTAS, you should provide a History Import File that looks something like this:

ID	Begin Date	End Date	Employment	Level1	Level2	Marital Status
00001	01011943	09111943	0	2.5	0	Single
00001	09121943	12311943	1	2.5	0	Single
00001	01011944	12311952	1	2.5	1	Single
00001	01011953	06141953	0	0	1	Single
00001	06151953	12311953	0	0	1	Married
00001	01011954	12311954	0	0	0	Married
00001	01011955	11201955	0	3.33	0	Married
00001	11211955	08311956	1	3.33	0	Married
00001	09011956	03121960	1	3.33	1	Married
00001	03131960	02091961	0	3.33	1	Married
00001	02101961	09181961	0	3.33	1	Divorced
00001	09191961	12311975	1	3.33	1	Divorced
00001	01011976	12311984	1	0	1	Divorced
00001	01011985	12311990	1	0	0	Divorced

Data Validation & Exceptions

In order for Import to populate the LTAS Study database the input files must contain complete and valid data. Errors and omissions in the data result in Import taking corrective action, which can be to reject a person, outcome, or history from the cohort, correct the data in a field and / or issue a warning message.

The Import Wizard ensures that you define all the required fields within a file prior to commencing the import process. LTAS of course cannot ‘know’ that the proper data are actually in the columns defined as a field, only that a field has been defined. If the field has been ill-defined in all likelihood this will result in field or record level exceptions.

When a requirement is not met, LTAS issues an exception of one of the following action types:

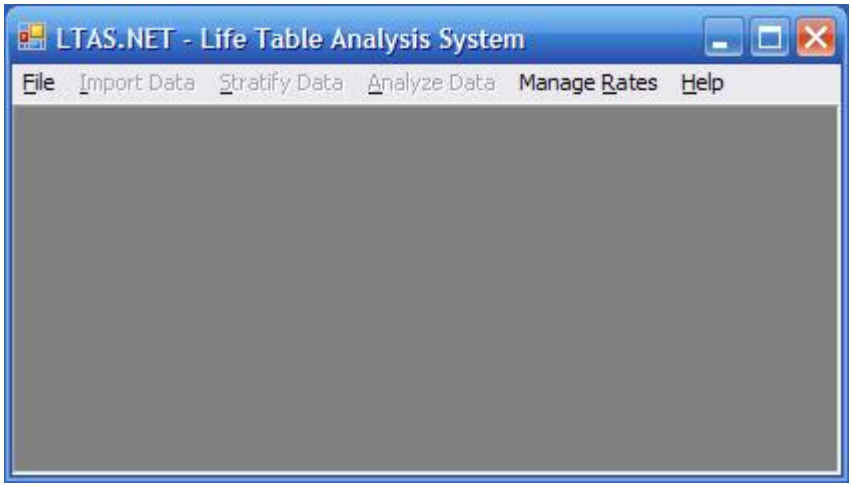
Action Type	Description
Termination	Import has found an error in the data such that the program cannot continue. LTAS displays an error dialog and control is returned to the Import Wizard or the main LTAS window.
Rejection	A record from the file in error (person, outcome or history) is eliminated from the cohort due to error. This may result in cascading rejections (e.g. a deceased person whose outcomes are all rejected is also rejected). An error is listed to the Import Exceptions report.
Exclusion	A person, outcome or history is eliminated from the cohort by request. Exclusions are not considered errors; they generally conform to a request such as limiting the cohort to a specified gender or race.
Warning	The data contains values that 'seem' erroneous. LTAS can continue by either using or ignoring the value. A warning is listed to the Import Exceptions report.
Redemption	LTAS has found incomplete or erroneous data but is able to resolve the problem by making some correction or addition, such as changing a missing day within a date to 15. The data is modified and a message is written ot the Import Exceptions report.

For a complete list of requirements, please see Appendix C - Import File Requirements.

Chapter 5. How to Run LTAS

Starting LTAS

LTAS.NET can be started by double-clicking on the LTAS.NET icon on your desktop. The main window of LTAS is shown below:



Stopping LTAS

You can stop LTAS by clicking on the X in the upper right corner of the main window or by selecting the Exit option from the File drop down menu.

File Menu

To begin using LTAS.NET you must first create a new LTAS project or open an existing LTAS project by using one of the options on the File menu. If you would like to create a new project that uses the same settings as an existing project, you must first export the existing project (while the project is open, select "Export project settings" from the File pull-down, and save as an XML file). Then create a new project, and in the first screen of the Import wizard, select the option to "Use options from an exported project".

Once you have opened your project, running LTAS is a 3-step process that includes:

1. Importing Data
2. Stratifying Data
3. Analyzing Data

Import Data Menu

The Import Data menu remains grayed out until you have opened a project. Once a project has been opened or created, you can use the Import Data menu to begin importing data.

Stratify Data Menu

Once you open a project, the Stratify Data menu may remain grayed out. This indicates that data has not yet been imported. Once the import step process is complete, the Stratify Data menu will be enabled and allow you to stratify the data.

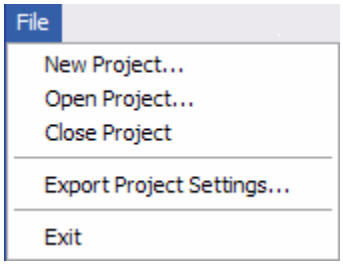
Analyze Data Menu

The Analyze Data menu will remain grayed out until you have completed the import and stratification of your data. Once this has been completed, the Analyze Data menu will be enabled and allow you to begin your analysis. If you re-import your data, the Analyze Data menu will become unavailable until you run Stratify again.

LTAS Project Files

A new LTAS project should be created for each study the user wishes to work on. Each LTAS project is stored as a single file on your computer's hard drive with a .LTAS file extension. The project file is used to store all of your import, stratify and analyze settings as well as your stratified study data.

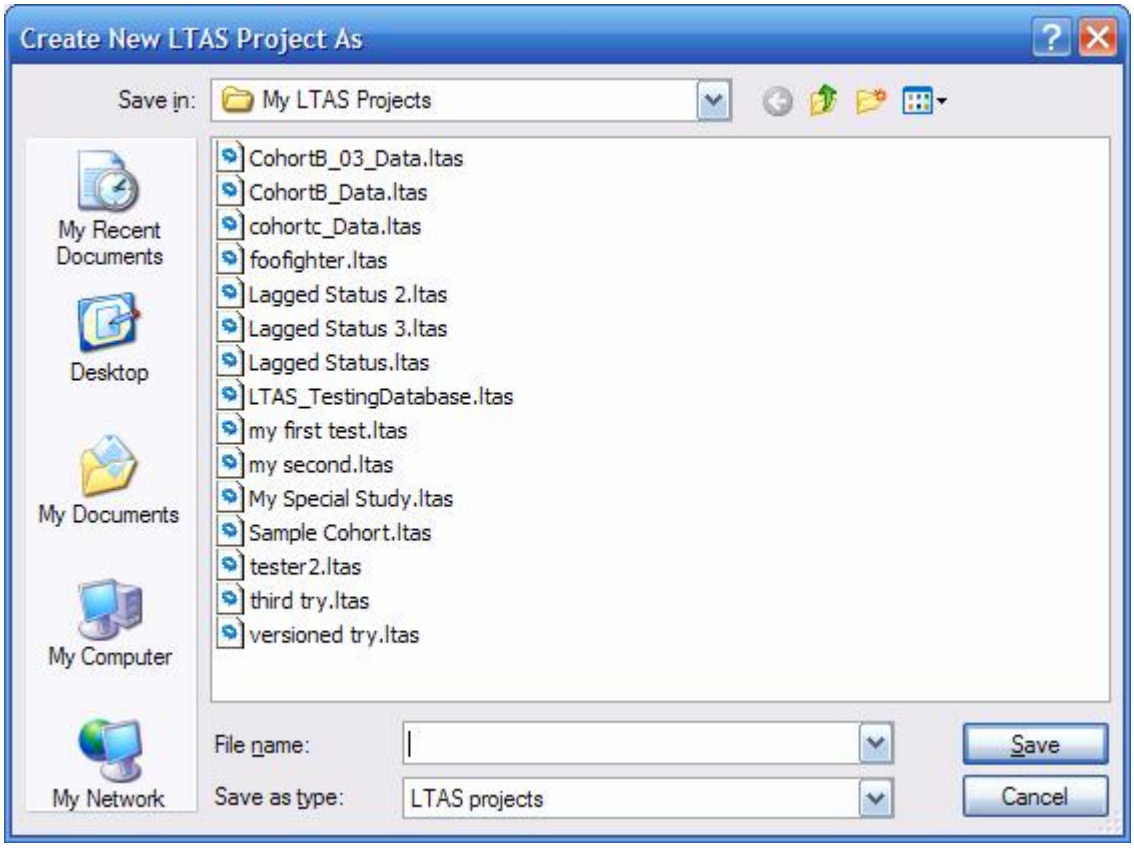
LTAS projects can be accessed through the File menu shown below:



Use the File menu to create new projects, open and close existing projects, perform other project management functions such as copying, deleting or exporting projects, or exit from LTAS.

Create a New LTAS Project

Use the New Project dialog to create a new LTAS Project. From the file menu, select new project. The dialog below appears.



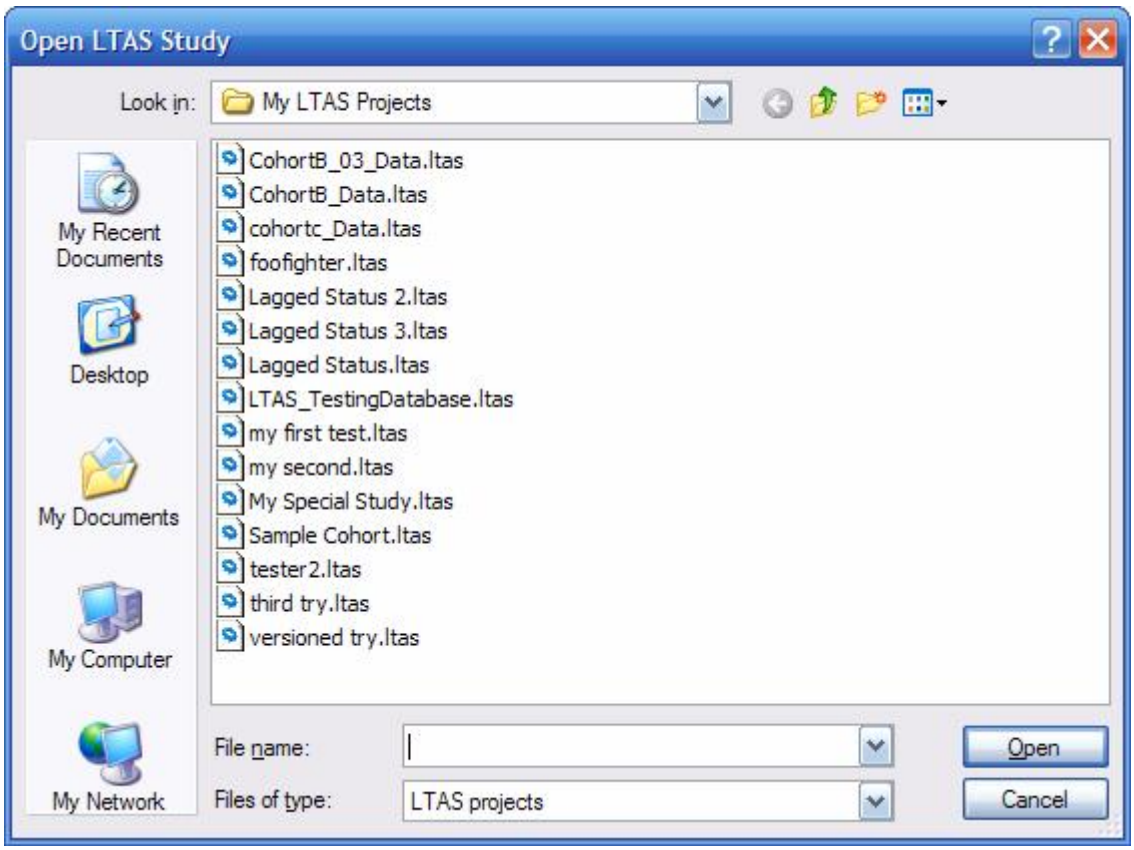
To use this dialog, you must enter a name for the LTAS project file. By default, all LTAS project files are created in the My LTAS Projects folder under My Documents. However, you may wish to place your project files in a different folder or create a folder hierarchy to group your project files. To create your project file in a different folder, simply navigate to the new location.

> **Note:** LTAS.NET does not currently support creating or accessing project files located on network drives.

Once you have selected the name and location of your LTAS project file, click the 'Save' button to create the project. LTAS will then initialize your project file and start the Import Wizard.

Open an existing LTAS Project

To open an existing LTAS project, use the Open Project dialog to select the project from the list and press the 'Open' button.



Managing LTAS Projects

Copying, Backing Up, Deleting and Renaming Projects

You may copy, backup, delete or rename a project file through the Windows Explorer interface using the same operations you use to organize other files on your computer's hard drive. Before attempting the operation make sure you have exited out of LTAS.NET. Windows will not allow you to access the file if LTAS.NET has the project open.

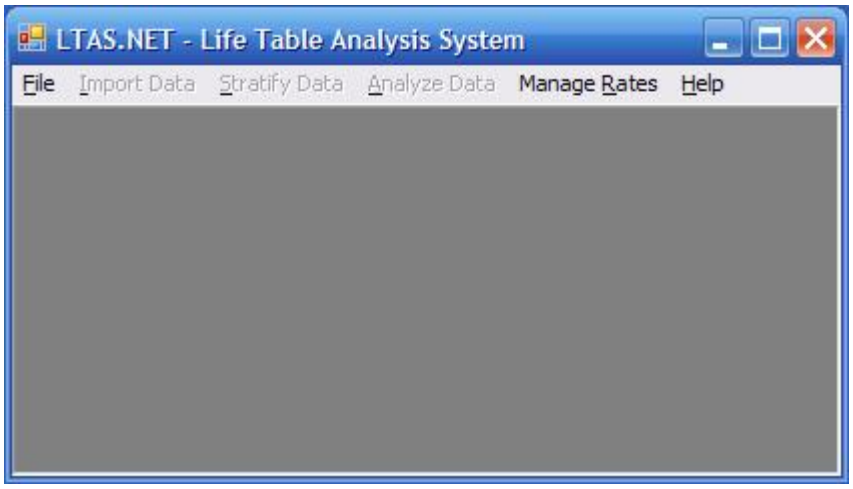
Exporting Projects

Instead of making a full copy of the project file, you can use the 'Export Project' menu option to export just your import, stratify and analyze settings. The export file will not contain any of the project data.

Selecting the 'Export Project' menu option will cause LTAS.NET to prompt for a location to save your current LTAS project options. Performing an Export will not remove your project. This file can then be used to recreate the project with the same import, stratify and analyze settings. You can also use the settings from an exported project as options for a new project, as described previously.

Exiting LTAS

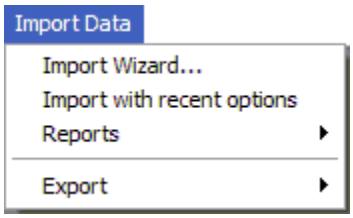
To exit the LTAS system, click the 'x' button at the top right hand corner of the main window; alternatively, LTAS can be closed by selecting the 'exit' option from the file menu.


LTAS.NET - Life Table Analysis System
File
Import Data
Stratify Data
Analyze Data
Manage Rates
Help

Chapter 6. Import Process

In order to perform an analysis using LTAS, it is necessary to provide a collection of data on the cohort under study. These data must include a minimum set of variables containing information on each person who is a member of the cohort, including several demographic and exposure history variables. The Import step of LTAS performs various data validations and transformations, rejects data failing data validation, excludes persons failing to meet study requirements, copies all remaining data to the study directory and reports on its processing.

The Import step can be performed using the Import Wizard accessed from the Import Data menu.



To perform the Import step with the same options used during the last Import, select "Import with recent options" from the Import Data menu.

Import Wizard

The LTAS Import Wizard consists of a series of screens that prompt the user to select from a number of available options. Import will determine the sequence of screens to display according to selections made to previous screens. Each page must be completed with valid information before the Wizard will activate the “Next” button to allow you to proceed to the next page. When all the options on all the pages have been specified, a final screen appears displaying a summary of the selections made and allowing the user to save the set of options for reuse. When the user presses the ‘Execute’ button of this final screen, the Import process begins. A dialog appears informing the user of the progress of the import process and providing an opportunity to cancel the operation.

Retrieve Import Options

The Import Wizard allows you to retrieve import options saved from prior Import Wizard executions. Retrieving a saved session provides all the options in the wizard saving you from having to select your options over again. The Import Wizard provides three different ways to retrieve import options using system defaults, they are:

Retrieval Method	Description
Use system defaults	Resets all of the import options to the system defaults.
Use PC-LTAS defaults	Resets all of the import options to the PC-LTAS defaults.
Use options from this project	Retrieves the import options last used on this project.

Use options from an exported project	Retrieves the import options from a project that has been exported as an XML file using the "Export project settings" within the "File" pull-down.

Project Options

Study Description

Enter up to 80 characters of text to describe the study. This description will be used to throughout the LTAS software and printed reports to identify the study.

Reference Rates

Rate sets for unexposed or referent population are chosen from this list. The most common are "92 Underlying cause U.S. Death Rates", which begin in 1940, and "119 Underlying cause U.S. Death Rates", which begin in 1960. See Appendix E - Rate Files for a description of the rate sets which come with LTAS.

The choice of rate set depends on the type of study. Proportion rate sets must be chosen for PMR studies when only deaths are available in your data. For cohorts with both living and dead, standard rate sets must be chosen. For cohorts with deaths as endpoints, mortality rate sets must be chosen. For cohorts with cancer incidence endpoints, SEER cancer incidence rates may be selected.

Study Begin Date

Workers ending employment before this date will be excluded from the analysis. For many studies this may be set at a very early date prior to any of the work histories (such as 01/01/1900). This cannot be left blank.

Study End Date

End date is the last date of the study, or end of follow-up. See **Study Begin Date**.

Global Categoricals

If you wish to include any Global Categorical stratifying variables, enter a names and optional descriptions here. Remember that global categoricals do not have any data associated with them in your cohort files - their categories are defined in the Stratify process simply as dates when categories change for all the cohort simultaneously.

In LTAS, the Import process transforms your cohort data into the LTAS data model from the following required import files:

Import File	Description
Person	Each record in this import file represents one person in the cohort. May contain extrinsic (i.e., those that will not be used to standardize) fixed stratifiers.
Outcome	Each record in this import file represents a 'failure' (death or diagnosis) in the cohort.
History	Each record in this import file represents a person's exposure to a suspected disease agent or other extrinsic temporal stratifier during a specific period of time. Employment period is a common example of an extrinsic temporal (time-dependent) stratifier within the History file.

LTAS contains a mapping feature in the import wizard to allow for definition of various file layouts as long as the data are provided in one of the following two file formats:

File Format	Description
Fixed ASCII	The file contains only text characters. Variables are of a fixed length and always begin at the same position within the file. Typically, but not necessarily, the file name extension is TXT or ASC.
Delimited ASCII	The file contains only text characters. Variables are separated by a delimiting character such as a semicolon, comma or tab. This format can often be produced by exporting data from a spreadsheet (such as Microsoft Excel) or database program. The file name extension is commonly CSV.

Optional Person File Fields

Include Name

Importing a Name for each Person is optional. If your Person import file includes Name data and you want to import that data, check this box.

Include Risk Begin Date

Importing a Risk Begin Date for each Person is optional depending on the type of study and how you wish to determine the Risk Begin Date. For example, if cohort members have exposure that begins before the start of follow-up, a risk begin date corresponding to the date the person entered follow-up could be denoted for each cohort member in the Person import file. If the follow-up for each cohort member began at the same date as the exposure, then a risk begin date is not necessary (see "Time At Risk Options" below).

Include Fixed Stratifiers

If you wish to include any fixed stratifying variables (i.e., those that do not change over time), enter a name and optional description here. Each stratifying variable that you specify here must be included as a column in your Person import file.

Recoding of Person File Fields

Import will convert the codes for Gender, Race and Vital Status to the internal codes used by LTAS to represent the various categories. The default coding scheme is displayed on the Recoding of Person File Fields page of the Import Wizard. If your data are coded differently, enter the codes to represent the categories here. If you want to exclude certain Gender or Race categories from the import process, uncheck the box next to that category.

The Gender and Race categories available for mapping are determined by the rate file being used. Note that the rate files supplied with LTAS.NET contain only two Race categories: White and All Other Races. Users can create their own rate files with expanded categories for race, and with other intrinsic stratifiers besides Gender, Race, Age and Calendar Year, if such data are available for the comparison population.

Both Gender and Race support mapping multiple codes to the same category by typing in a comma-delimited list of codes. If Import encounters a Gender or Race code in your data that is not mapped, the Person being processed will be rejected.

While Vital Status does not support mapping multiple codes to the same category, when Import encounters a code that is not mapped, it will redeem the Person using a Vital Status of Alive.

Time At Risk Options

Begin risk accumulation

For each person in the cohort Import determines a date on which to begin accumulation of time at risk based on your Begin risk accumulation selections. The begin risk date is defined as the latest in a list of dates. These dates include:

Date	Description
Rate Begin Date	Earliest date in selected rates. Always a candidate.
Qualifying Birthday	Date at which person attains the youngest age included in the selected rates. Always a candidate.
1st Exposed Date	Person's earliest date of exposure. If there is more than one exposure variable in your study, the earliest of the first exposed dates is used. Optionally a candidate.
Study Begin Date	The study begin date. Optionally a candidate.
From Person Data	A risk begin date from the person's record in the Person File. Optionally a candidate.
Global Date	A user-specified global date that applies to all persons in the cohort. Optionally a candidate.

End risk accumulation

End of risk accumulation for persons Lost-To-Follow-up is determined by one of the following options:

Date	Description
End of Study	Risk accumulation ends on the study end date.
Earlier of Date Last Observed or End Of Study	Risk accumulation ends on the earlier of person's date last observed or the study end date.

Mapping of Fixed Person Fields

The Mapping of Fixed Person Fields page of the Import Wizard controls how import will access the raw data from your Person import file. From this page your can change the fixed mapping of any Person field.

Mapping of Delimited Person Fields

The Mapping of Delimited Person Fields page of the Import Wizard controls how import will access the raw data from your Person import file. From this page your can change the delimited mapping of any Person field.

Outcome Options

Include Terminal Flag *

The Terminal Flag field is used to indicate if a given outcome is a death or a diagnosis. If you do not include this field in your import file, LTAS will treat all outcomes for deceased persons as deaths and all outcomes for alive persons as diagnoses.

Include Underlying Flag *

The Underlying Flag field is used to indicate if a given outcome is the underlying cause or a multiple cause. If you do not include this field in your import file, LTAS will treat all outcomes as underlying.

Use ICD Revision

LTAS translates the Disease Code in the Outcome file from ICD code to NIOSH (or user-defined) minor. This can be accomplished by selecting one of the following ICD Revision Options:

ICD Revision Option	Description
In Effect at Time of Death	Use the ICD revision in effect at the time of death. Default.
From Outcome File	ICD revision included for each outcome in the Outcome file.
Global	ICD revision to use for all deaths regardless of date.

* These options have been provided to allow a PC LTAS Demographics file to be used as the Outcome Import file.

Mapping of Fixed Outcome Fields

The Mapping of Fixed Outcome Fields page of the Import Wizard controls how import will access the raw data from your Outcome import file. From this page your can change the fixed mapping of any Outcome field.

Mapping of Delimited Outcome Fields

The Mapping of Delimited Outcome Fields page of the Import Wizard controls how import will access the raw data from your Outcome import file. From this page your can change the delimited mapping of any Outcome field.

Optional History File Fields

Exposure Agents

A set of temporal stratifiers can be tracked for each time-dependent exposure in your study: For each exposure the wizard will present a page to select any of these variables, and to enter a name for and optional description of the exposure. The following stratifiers may be selected:

Field	Description
Duration	Cumulative time a person is exposed.
Level	Total level a person accumulates during their exposed time.
TSFE	Time since first exposure.
TSLE	Time since last exposure.
Employment Status	Categorical stratifier indicating employment periods.

Some planning is in order when selecting these variables. Additional stratifiers produce additional levels of stratification, which require more space in the database and more processing time. This is a more critical issue during the Stratify and Analyze steps, where including all possible levels of stratification will likely require much patience, and possibly result in LTAS inability to complete the process due to insufficient resources (computer memory and / or disk space). On the other hand, selecting fewer stratifiers when importing means that you will not be able to stratify or analyze on the unselected variables without re-importing.

Temporal Categoricals

Categorical stratifiers that are associated with the persons in the cohort and can change over time are included in the History File as Temporal Categoricals. If you wish to include any Temporal Categoricals, specify names and descriptions on this page.

Mapping of Fixed History Fields

The Mapping of Fixed History Fields page of the Import Wizard controls how import will access the raw data from your History import file. From this page your can change the fixed mapping of any History field.

Mapping of Delimited History Fields

The Mapping of Delimited History Fields page of the Import Wizard controls how import will access the raw data from your History import file. From this page your can change the delimited mapping of any History field.

Fixed ASCII Mapping

If you have identified Fixed ASCII as your input data format, the wizard will display the Fixed ASCII field mapping screen. As you select a field from the list, the wizard highlights the associated data in the two panels below. The “Raw Data” panel displays the fixed ASCII file that you named previously. Highlighted in the first line are the columns the wizard currently has assigned to the selected field. The “LTAS Field” panel displays the data as it would be imported into the LTAS study database. Only the first 500 lines are displayed, and no data validations are yet performed.

Inspect the fields the wizard displays. If any of the fields are incorrectly defined, you can correct the error by changing the column and size associated with the field. First select the field in question, correct the associated columns, and press the “Apply” button. To change the column definition you can either enter the starting (leftmost) column in the “Column” field and the number of columns in the “Width” field, or highlight the data in the “Raw data” panel by selecting it using the mouse (click and drag) or keyboard (using shift and arrow keys). (NOTE: as you selected text in the “Raw data” panel, the wizard displays the current column in the “Column” field. This alone does not change the field definition; you must press the “Apply” button. To abandon the change and again see the current definition of the field’s columns simply reselect the variable from the drop-down list.)

When all the fields have been correctly assigned to their associated columns in the raw data, press “Next” to continue. If the “Next” button appears dimmed, one or more of the fields lacks a valid definition. The button will reactivate when all fields are properly defined.

Delimited ASCII Mapping

If you are using Delimited ASCII as your input data format, the wizard will display the Delimited ASCII field mapping screen. A single panel displays the data from the delimited ASCII file. As you select a field from the list, the wizard highlights the associated data in the panel. The entire file is displayed, and no data validations are yet performed.

If any of the fields is incorrectly defined, you can correct the error by changing the column associated with the field. First select the field in question, then select the correct column from the “LTAS Fields” panel and press the “Redefine field” button. The wizard will then re-label that column with the name of the selected field, and remove the label from the incorrectly defined column. Generally this means that a field is now left without a definition. Continue to select and define fields until all have been properly associated, then press the “Next” button to continue. If the “Next” button appears dimmed, one or more fields lack a valid definition.

Import Reports

Several reports are available as result of the Import process. These reports can be accessed from the Import Data -> Reports menu once the Import process has been completed. The following Import reports are currently available:

Report	Description
Import Options	Provides a detailed listing of all of the Import options selected the last time that data was imported into the project.
Cohort Summary	Provides a summary of the import record counts and person demographics for the data last imported into the project.
Exceptions Summary	Provides a summary of the import exceptions that resulted from the last import.
Exceptions by Person	Provides a detailed listing of the import exceptions grouped by person.
Exceptions by Type	Provides a detailed listing of the import exceptions grouped by record type, action and exception message.

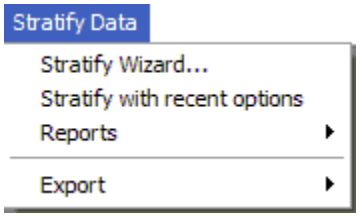
Exporting Import Results

You can export the exceptions that occurred during the import process to a comma-delimited or XML formatted data file using the Import Data -> Export menu option. The comma-delimited file includes the headings for each column as the first line and can easily be loaded into spreadsheet programs like Excel.

Chapter 7. Stratify Process

Stratify Process

Once you have imported your cohort data into LTAS, you will need to stratify the data before completing the analysis.



To do this you select the stratifiers of interest and define category thresholds as necessary. Stratify assigns sums of time and events into these categories and writes the summations to the study database in preparation for the Analyze step.

To begin the Stratify process, select the 'Stratify Wizard' option from the 'Stratify Data' menu. This will start the Stratify Wizard.

Be aware that the more stratifiers and categories you request for stratification, the more work you are asking Stratify to do. Large cohorts with many stratifiers and categories require lengthy stratification. Consider reducing the number of stratifier-category combinations to reduce Stratify times.

To perform the Stratify step with the same options used during the last Stratify, select "Stratify with recent options" from the Stratify Data menu.

Stratify Wizard

The LTAS Stratify Wizard consists of a series of screens that prompt the user to select from a number of available options. The wizard will determine the sequence of screens to display according to selections made to previous screens. These screens include:

- Stratifier Selection
- Lag Time
- Categories

On the final screen of the wizard, the user presses the ‘Finish’ button to start the Stratify Process A dialog appears informing the user of the progress of the stratify process and providing an opportunity to cancel the operation.

Stratifier Selection

The Stratify Selection screen lists all the known stratifiers. This includes intrinsic, extrinsic, fixed and temporal. The intrinsic stratifiers are those that are included in the selected rate set (on which standardization is performed). Typically these include Gender, Race, Age and Calendar Period. Since the categories for these stratifiers are defined for the rate set, these categories cannot be changed in Stratify. Because these are standardizing variables, you must always stratify your data on these intrinsic stratifiers and categories. This is why intrinsics are shown as “read-only” (checked and grayed) – you have no option but to include these intrinsic stratifiers and their categories in your analysis.

Extrinsic stratifiers are those are not included in the rate set. These stratifiers were identified during the Import process, either associated with an exposure agent, defined as input fields in the raw history file ,or specified as global categoricals. You may choose to stratify on any of these variables by checking the box next to its name.

Define Lag Time

Lag Time

For each exposure agent that you choose for stratification, you may specify a lag time value. Lag time values allow you to offset the exposure accumulation by a fixed period of time (see Appendix A for more detail). You may specify a lag period by entering a number in the “Lag” column and choosing "Years" or "Days" for the "Units" column.

Include a separate category for lag time

Checking this option will cause the time during the lag to be placed in a separate category. If unchecked, the time during the lag will be included in the lowest category.

The separate category created by checking this option will be labeled as 'Lagged Time' and is available for inclusion in any of the Analysis reports. If you plan on performing SRR analysis please note that this category is not included in either the SRR or Rothman Trend calculations. Only the person time and observed deaths will be reported for this category.

Edit Categories

The Edit Categories screen is displayed for each accumulating stratifier. This screen is used to define the categories to be used for stratification.

Threshold Values Represent

Specifies the units of measure that the threshold values (cutpoints) for each category represent. The unit of measure for exposure level threshold values can be specified by entering the description, while threshold values for global categoricals always represent dates. The threshold values for the other types of accumulating stratifiers including duration, TSFE and TSLE can be represented as days or years.

Categories

The categories grid is used to define the categories to be used for stratification. Use the 'Examples' button to populate the grid with a sample set of categories. Use the 'Clear' button to clear the grid.

Add a new category by entering a new label and threshold value in the bottom row of the grid. LTAS.NET will automatically resort the grid so that the categories remain ordered by the threshold value.

Delete an existing category by highlighting the row containing the category to be deleted and then press the 'Delete' key.

Each category is defined by a label and a corresponding threshold value. The label can be any description up to 30 characters in length. The threshold value must be a valid number, zero or higher.

The first category in the grid defines the minimum required threshold. Any person that fails to accumulate to this minimum will be rejected during the stratification process. Note that a minimum cannot be specified if a lag is selected.

Stratify Reports

Several reports are available as result of the Stratify process. These reports can be accessed from the Stratify Data -> Reports menu once the Stratify process has been completed. The following Stratify reports are currently available:

Report	Description
Stratify Options	Provides a detailed listing of all of the Stratify options selected the last time that data was stratified.
Stratify Summary	Provides a summary of the stratified cohort including the numerator (observed outcomes) and denominator (person years).
Rejections	Provides a detailed listing of the persons rejected for failing to satisfy minimum exposure.
Personal Cumulative Dose	Provides a detailed listing of each person in the study along with cumulative totals for each exposure agent.

Exporting Stratify Results

You can export the stratify results to a comma-delimited or XML formatted data file using the Stratify Data -> Export menu option. The following stratify data is currently available for exporting:

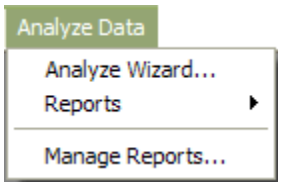
Export Data	Description
Cohort	Exports a detailed listing of the stratified cohort.
Rejections	Exports a detailed listing of the rejections.
Personal Cumulative Dose	Exports a detailed listing of each person in the study along with cumulative totals for each exposure agent.

The comma-delimited file format includes the headings for each column as the first line and can easily be loaded into spreadsheet programs like Excel.

Chapter 8. Analyze Process

Analyze Process

The analyze process produces reports about your study using report options selected using the Analyze Wizard. The analyze wizard and reports that have already been defined can be accessed from the Analyze Data menu.



Analyze Wizard

The LTAS Analyze Wizard consists of a series of screens that prompt the user to select from a number of available options. The wizard will determine the sequence of screens to display according to selections made to previous screens. These screens include the following:

- Report Style
- Category Selection
- Report Options

On the final screen of the wizard, the user presses the 'Finish' button to start the Analyze Process.

Select Report Style

The Select Report Style screen is the first step of the Analyze Wizard and is used to select the style and layout of the Analyze report that you want to create.

Report Style

Style	Description
Distribution of Person Years	Reports the distribution of person years for each combination of selected categories.
Standardized Mortality Ratio	Reports the observed deaths, expected deaths and standardized mortality ratios (SMRs) for each combination of selected categories.
Standardized Rate Ratio	Reports the distribution of observed deaths (or other events), person years at risk, standardized rate ratios (SRRs), and Rothman trend test statistics (if a grid report is selected for a temporal exposure variable) for each combination of selected categories.

Report Layout

Layout	Description
Grid	Formats each page of the report with each combination of categories from two selected stratifiers by using a two-dimensional table.
List	Formats each page of the report as a (one-dimensional) list of categories for one selected stratifier.

Category Selection

Stratifier Categories

The Category Selection screen displays all of the stratified categories that are available for analysis. These categories are displayed in a tree with a checkbox beside each stratifier group and category.

From this screen you can either click on the checkbox beside a stratifier group to select all of the categories available for that stratifier or expand the tree and click on individual categories to include in the analysis.

In addition to controlling which categories are included in the analysis, you can include 'All categories combined' summary of a stratifier by toggling the checkbox beside the stratifier group so that a black checkmark appears. By default, a gray checkmark will be displayed beside the stratifier group indicating that analysis will only be performed on the selected categories individually.

Outcome Causes

For SMR/PMR and SRR reports, the Category Selection screen also allows you choose specific outcome causes. The major and minor categories for all causes are displayed in a tree with a checkbox beside each cause.

From this screen you can either click on the checkbox beside a major cause to select all of the minor causes underneath, or expand the tree and click on individual minor causes to include in the analysis. You can include a combined summary for a major by toggling the checkbox beside the major cause so that a black checkmark appears. By default, a gray checkmark will be displayed beside the major indicating that analysis will only be performed on the selected minors individually.

Report Options

The Report Options screen is used to define several options that control how the report is generated. A more detailed description of these options can be found below.

Variable for Each Row of List

For List style reports, you must choose one of your category variables to represent each row of the report. The values of all other category variables will be displayed in the page header of the report.

Variable for Vertical Axis of Grid / Variable for Horizontal Axis of Grid

For Grid style reports, you must choose one of your category variables to represent the vertical axis of the grid and another variable to represent the horizontal axis of the grid. The values of all other category variables will be displayed in the page header of the report. Note that SRR reports perform calculations across the horizontal axis. The variable that you select for this axis will determine which categories are available as the reference group.

Confidence Interval

For SMR and SRR reports, you must choose the type of confidence interval that you would like calculated for the report.

Reference Group

For SRR reports, you must choose one of the categories from your horizontal variable to be the reference group. All SRR calculations will be calculated relative to this group.

Column Width

Grid reports are available in three different column widths: narrow, medium and wide. Selecting a narrow column width can help squeeze more columns onto a single page, while selecting a wider column width may be necessary for your data to fit within each column.

Font Size

Grid reports can be printed with different font sizes to allow more data to fit within a given column size. You may decide to increase both the font size and column width to make a report more readable, or decrease the font size and column width in order to fit more columns onto a single page.

Page Orientation

Use page orientation to switch between Portrait and Landscape page layouts. Consider using Landscape orientation to fit more columns onto a single page.

Save Report As

By checking this box and typing in a name, LTAS will save all of your selections as a report definition in your study database. This will allow you to run the report again without running the Analyze Wizard. Instead, you can instantly access all saved report definitions from the Analyze Data -> Reports menu. You can also edit saved reports using Analyze Data->Managing Reports, as described below.

Analyze Reports

When you create an Analyze report using the Analyze Wizard and choose to save that report, that report automatically becomes available as an option under the Analyze Data -> Reports menu. Selecting that option will run the report again without having to use the Analyze Wizard.

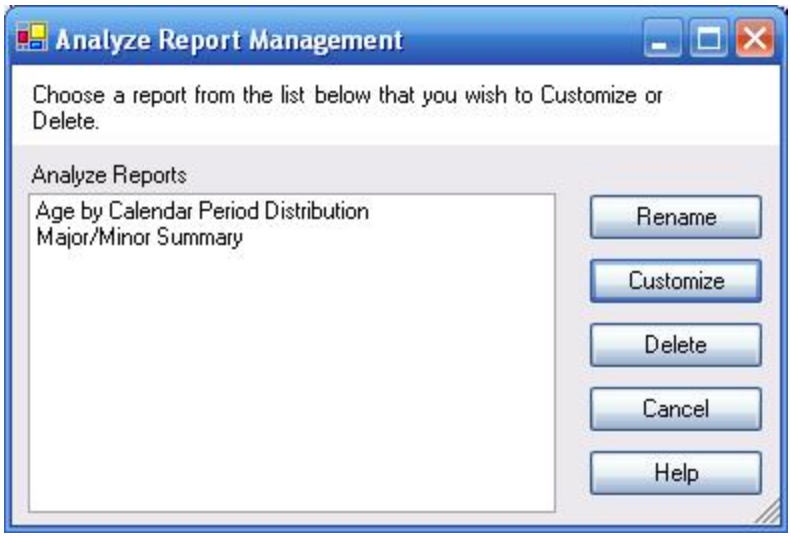
When you create a new project and choose to use options from an existing project, any reports currently defined in the existing project will be copied over to your new project. If you choose to use the system defaults, the following reports will be automatically defined within your project:

Report	Description
Age by Calendar Period Distribution	Provides a table of person year totals distributed by age category and calendar year.
Major/Minor Summary	Provides a detailed listing of observed and expected deaths for each disease category along with SMR calculations.

Under certain conditions, selecting a saved report to run may cause the Analyze Wizard to appear. This will happen if the report contains references to variables and categories that are no longer defined. This can be caused by upgrading LTAS.NET, changing rate files, or stratifying with new variables or categories. If this happens to one of your reports, step through the Analyze Wizard, verify all of your selections and save the report.

Managing Reports

Analyze reports created using the Analyze Wizard as well as the default reports included with each project can be renamed, modified or removed by using the Analyze Data -> Manage Reports menu option. Selecting this option causes the following dialog to appear:



Rename a Report

You can rename a report using the Manage Reports dialog by selecting the report that you wish to rename and clicking the Rename button. You will then be prompted with an edit box containing the name of the report.

Modify a Report

You can modify a report using the Manage Reports dialog by selecting the report that you wish to modify and clicking the Customize button. This will load the report definition and take you to the Analyze Wizard where you can modify any of the report options.

Delete a Report

To delete a report, select the report you wish to delete and click on the Delete button. When you are prompted to confirm that you want to delete the report, you must click on the Yes button to complete the process.

Column and Row Totals

When you create a report with a Grid layout, a total column and total row are automatically generated including a grand total in the bottom right. The values displayed in these cells of the grid are not a simple summation across or down the grid. The table below describes the values that are displayed in these cells and how they are calculated.

Value	Description
Observed	Sum of observed deaths (or other outcomes) across each category in the row or column. Since these are exact values, the grand total value also matches both the sum of the row total values and the sum of the column total values.
Expected	Sum of expected deaths across each category in the row or column. While the report only displays two digits of precision to the right of the decimal point, expected values are represented in the rate databases with much greater precision. Since the total values are the sum of these higher precision values, they do not always match the sum of the report values. This is also true of the grand total value.
Person Years	Sum of person years across each category in the row or column. While the report only displays two digits of precision to the right of the decimal point, person year values are calculated with much greater precision. Since the total values are the sum of these higher precision values, they do not always match the sum of the report values. This is also true of the grand total value.
SMR/PMR	These ratio values do not represent summed values. These values are calculated from the total observed and total expected values.
SRR	These ratio values do not represent summed values. For the row total, the values are calculated by applying the SRR calculation over each category for the horizontal variable while ignoring the vertical variable. For the column total, the values are calculated by applying the SRR calculation over the non-reference columns.
CI	The confidence interval values displayed in the total column and total row do not represent summed values, they represent the CI of the SRR, SMR or PMR value calculated for that cell.

Chapter 9. Appendices

Appendix A - LTAS Date Handling and Lag

Introduction

The purpose of this appendix is to describe the logic used for all date calculations within LTAS. Generally, the date handling logic for LTAS.NET was designed to be compatible with PC-LTAS. LTAS.NET differs from PC-LTAS in a few situations which were deemed to be inconsistent in PC-LTAS. These situations are described in more detail throughout the document. This section also describes procedures used to lag exposures, to account for the expected latency between exposure and disease incidence or mortality.

Date calculations show up in LTAS in the following areas:

1. History Gaps and Overlaps
2. Transition Dates
3. Age Calculations
4. First Qualifying Birthday
5. Calendar Period Calculations
6. TSLE Calculations
7. TSFE Calculations
8. Exposure Level and Duration Calculations
9. Exposure Lag

Many date calculations are based on calculating the number of days between two events. Because events are recorded as only a date without the exact time, LTAS.NET always treats events that start something as occurring at the beginning of the day and events that end something as occurring at the end of the day.

Begin Events	End Events
Date of Birth	Risk End Date
Birthdays	Date of Death
Risk Begin Date	Date Last Observed
Study Begin Date	Study End Date
History/Job Begin Date	History/Job End Date
Calendar Year/Period Begin Date	Calendar Year/Period End Date
Transition Dates	

Based on this treatment of event dates, LTAS.NET will calculate the number of days between two events based on the following logic:

If First Event is a ...	And Second Event is a ...	The number of days is ...
Begin Event	End Event	(Second Event - First Event) + 1
Begin Event	Begin Event	(Second Event - First Event)
End Event	End Event	(Second Event - First Event)
End Event	Begin Event	(Second Event - First Event) - 1

Note: There are currently no calculations in LTAS.NET where the first event is an End Event and the second event is a Begin Event, but this could be conceptualized as the first being a work history end date and the second being a subsequent work history begin date. The time between these dates would be appropriately calculated using the above formula.

The date calculations within PC-LTAS are not as consistent as in LTAS.NET. This has resulted in several differences noted throughout this document.

History Gaps and Overlaps

LTAS.NET and PC-LTAS will fill gaps in the history with Unexposed history records. For instance, the following import file:

```
000000031 10/01/1956 10/02/1956  2E-6
000000031 01/01/1961 12/20/1961  0.03629944
```

will actually produce the following three history records:

```
000000031 10/01/1956 10/02/1956  2E-6
000000031 10/03/1956 12/31/1960  0
000000031 01/01/1961 12/20/1961  0.03629944
```

The middle record is automatically inserted with a zero exposure value to provide a complete set of history records for the person.

Now consider the following two history records:

```
000000032 10/01/1956 10/02/1956  2E-6
000000032 10/02/1956 12/20/1961  0.03629944
```

In this example, the end date for the first history record is equal to the start date for the next history record. In this case, the history records actually overlap by 1 day. While PC-LTAS permitted 1 day overlaps, this will cause an overlapping history exception in LTAS.NET and a rejection of the person. Overlaps of 2 or more days will cause an overlapping history exception in both PC-LTAS and LTAS.NET and a rejection of the person.

Transition Dates

LTAS.NET and PC-LTAS calculate transition dates for all temporal stratifiers and categories as part of the stratify process. A transition date is defined as a day that one or more temporal stratifiers fall into a new category compared to the previous day.

The stratify process uses transition dates to distribute person time across cells where a cell is defined as the time between two transition dates when all categories remain the same. Transition dates are considered begin events because they mark the beginning of each cell. The person time allocated to that cell is the difference between the two transition dates.

For example, if we have a cell between two transition dates, 06/01/1985 and 06/05/1985, the person time allocated for that cell would be 4 days because both days represent begin events. The same result is calculated in LTAS.NET and PC-LTAS.

Age Calculations

Transition dates for Age categories are calculated by LTAS.NET by adding the number of threshold years defined for each category to the person's birth date. The categories and threshold values depend on the rate database being used. The age categories used by rates provided with LTAS.NET are shown below:

Age Category	Threshold (years)
15 - < 20	15
20 - < 25	20
25 - < 30	25
30 - < 35	30
35 - < 40	35
40 - < 45	40
45 - < 50	45
50 - < 55	50
55 - < 60	55
60 - < 65	60
65 - < 70	65
70 - < 75	70
75 - < 80	75
80 - < 85	80
85 +	85

Using the table above for an example, the transition date for the 20 - < 25 category for a person born on 08/01/1943 would be 08/01/1943 + 20 years = 08/01/1963.

While PC-LTAS used similar logic for calculating transition dates for Age categories, the threshold values used in the calculations were in days instead of years. As a result, the transition date calculated by PC-LTAS may have fallen on the days before or after the actual birthday. In LTAS.NET, the transition date is on the actual birthday. The only exception is the case where a person was born on February 29, in which case, the transition date would be March 1 for non-leap years.

Any transition dates that fall outside of the time at risk are discarded by both LTAS.NET and PC-LTAS.

First Qualifying Birthday

Both LTAS.NET and PC-LTAS allow the Risk Begin date to be based on different events, including the first birthday at which a person enters a valid age category for the rates being used. This date is calculated by adding the threshold value of the lowest age category in the rate file to the persons birth date. Due to the inconsistent nature of date handling and the use of days for threshold values, PC-LTAS may incorrectly calculate a date that is 1-2 days after the persons actual birthday. In LTAS.NET, the first qualifying birthday date is on the actual birthday. The only exception is the case, as noted earlier, where a person was born on February 29, in which case the first qualifying birthday will be on March 1 for non-leap years.

Calendar Period Calculations

Transition dates for Calendar periods are set to the exact value of the begin date defined for each calendar period category. The categories and their begin date values depend on the rate database being used. The calendar periods used by rates provided with both LTAS.NET and PC-LTAS are shown below:

Calendar Period	Begin Date
1940 - 1944	1/1/1940
1945 - 1949	1/1/1945
1950 - 1954	1/1/1950
1955 - 1959	1/1/1955
1960 - 1964	1/1/1960
1965 - 1969	1/1/1965
1970 - 1974	1/1/1970
1975 - 1979	1/1/1975
1980 - 1984	1/1/1980
1985 - 1989	1/1/1985
1990 - 1994	1/1/1990
1995 - 1999	1/1/1995
Etc.	Etc.

Any transition dates that fall outside of the time at risk are discarded.

TSLE Calculations

Accumulation

PC-LTAS does not include support for time since last exposure (TSLE) calculations. LTAS.NET defines TSLE to be a function of the date of the last exposure in a person's exposure history. TSLE is defined to be zero for all dates up to and including the date of the last exposure. TSLE begins accruing the day after exposure ceases.

As an example, consider the following last exposure record:

```
000001023 01/05/1985 06/01/1985 0.2343
```

In this case, LTAS.NET will calculate TSLE as zero before exposure begins and prior to the end of the last exposure period. TSLE will take on positive values starting on 06/02/1985. TSLE will accumulate through the appropriate Risk End Date. For instance, a TSLE of zero would be assigned to time periods occurring between the Risk Begin Date and 06/01/1985. If the person died on 06/03/1985, the calculated TSLE would be 2 days (06/03/1985 - 06/01/1985) as of that date.

It is important to note that gaps (or unexposed periods) in the exposure record will be ignored in calculating TSLE. The relevant index date for calculating TSLE is always the last date in the last exposure period for the worker (except when applying a lag, as discussed below).

Transition Dates

LTAS.NET determines transition dates for TSLE by adding the threshold value for each category to the day after the last exposure end date. For example, if the first TSLE category has a 3 day threshold and the last exposure end date is 06/01/1985, the first transition date would be 06/05/1985 ((06/01/1985 + 1) + 3).

TSFE Calculations

Accumulation

Both PC-LTAS and LTAS.NET provide support for time since first exposure (TSFE) calculations. LTAS.NET defines TSFE to be a function of the date of the first exposure in a persons exposure history. TSFE is defined to be zero for all dates up to the day before the date of the first exposure. TSFE begins accruing the day exposure begins, and is calculated as [End Date (for time period under consideration) - First exposure Begin Date]+1.

As an example, consider the following first exposure record:

```
000002023 01/01/1945 01/02/1945 0.375
```

In this case, LTAS.NET will calculate TSFE starting on 01/01/1945. TSFE will accumulate through the appropriate Risk End Date. For instance, if the study ends on 01/03/1945, the calculated TSFE would be 3 days (01/03/1945 - 01/01/1945 + 1). TSFE is calculated dynamically, as are other time-dependent covariates. Thus, for each time period under consideration, TSFE is calculated as the date at the end of the period minus the first exposure begin date, plus one.

TSFE values calculated by PC-LTAS are always 1 day less than TSFE values calculated by LTAS.NET because PC-LTAS defined TSFE as (Risk End Date - First Exposure Begin Date).

Transition Dates

LTAS.NET and PC-LTAS determine transition dates for TSFE by adding the threshold value for each category to the first exposure begin date. For example, if the first TSFE category has a 1 day threshold and the first exposure begin date is 01/01/1945, the first transition date would be 01/02/1945 (01/01/1945 + 1).

Exposure Level and Duration Calculations

Calculations for exposure duration are essentially the same as those used for exposure level where the exposure value is always 1.0.

Accumulation

LTAS.NET and PC-LTAS accumulate exposure by multiplying the exposure value by the number of days in the exposure period (i.e., the difference between the dates plus one). The total exposure is calculated by summing the calculated value for each exposure record.

As an example, consider the following exposure record:

```
000002023 01/01/1945 01/02/1945 0.375
```

In this case, LTAS.NET and PC-LTAS will calculate the exposure as ((01/02/1945 - 01/01/1945) + 1) * 0.375 = 2 * 0.375 = 0.75.

Transition Dates

LTAS.NET and PC-LTAS determine transition dates for exposure by calculating the first day that the accumulated exposure (at the beginning of the day) is greater than or equal to the threshold value for each category.

As an example, consider the following exposure history:

```
000000031 10/01/1956 10/02/1956  2E-6
000000031 01/01/1961 12/20/1961  0.03629944
```

and the following exposure category thresholds:

```
1E-5
10.0
```

To determine the first transition date, LTAS.NET begins accumulating exposure. Accumulated exposure for the first record is 4E-6 ((10/02/1956 - 10/01/1956 + 1) * 2E-6). Since this value is less than the first threshold, LTAS.NET moves on to the next exposure record.

Accumulated exposure after the next record would be 4E-6 (from the first record) plus 12.85 ((12/20/1961 - 01/01/1961 + 1) * 0.03629944). This is well above the first category threshold, so LTAS.NET calculates the transition date by subtracting the accumulated exposure (not including the current record) from the threshold value and dividing by the current record exposure level value: days to transition = (1E-5 - 4E-6) / 0.03629944. The result is rounded up to the nearest whole number, which is 1 day in this case.

The actual transition date is calculated by adding the calculated days to transition to the history begin date = (01/01/1961 + 1) = 01/02/1961.

LTAS.NET then calculates the transition date for the next category: days to transition = (10.0 - 4E-6) / 0.03629944. The result rounded up to the nearest whole number is 276.

Again, the transition date is calculated by adding the calculated days to transition to the history begin date = (01/01/1961 + 276) = 10/04/1961. PC-LTAS uses similar logic to calculate the same results.

Exposure Lag

Introduction

Exposure lagging is commonly used when a researcher believes that there is a minimum empirical induction period for disease or death (i.e., a minimum period between the time that an exposure occurs and a disease will be detected or result in death). The concept of exposure lagging contains several components (described in Checkoway et al. 2004, pp. 168-173). The simplest definition of exposure lag period consists of a period of time during which exposure is discounted and does not contribute to cumulative exposure. It is conceptualized as a moving exposure blackout window which discounts any exposure occurring for a specified amount of time before the time point being considered.

To incorporate exposure lag into a life table analysis, at each unit of calendar time, a unit of person-time is added to cells formed by the cross-classification of several variables (e.g., age, calendar time), one of which is cumulative exposure. When a lag is used, the exposure cell is not the cell for the current cumulative exposure, but the cell for the cumulative exposure achieved a specified amount of time (equal to the lag) earlier.

Another common method for taking into account induction and latency is to ignore person-time and events that occur during the initial period of cohort follow-up time corresponding to a selected length of time (Checkoway et al. 2004, pp. 169), under the presumption that any disease occurring during that time could not have been caused by exposure. This is accomplished by specifying a minimum value for time since first exposure. For example, if a minimum of five years is selected, the first five years of follow-up and any events occurring during that period are not considered.

While both methods (specifying a lag for exposure or specifying a minimum value for time since first exposure) described above are available in PC-LTAS, the latter approach may be problematic if workers in a study have exposure before the start of follow-up.

Table 1a. Description of method of handling lags in PC-LTAS, for various exposure metrics

Variable	Method of handling lags in PC- LTAS	Method of handling lags in LTAS.NET
Cumulative exposure, events and person-time	Exposures received during the lagged time period are not counted towards the cumulative exposure for a given time-at-risk. If a person's entire exposure and follow-up time falls in the lag period, all their person-time and any events are placed in an unexposed group. Total number of observed and expected events in the unexposed group are output at the bottom of the SMR table.	Same as PC-LTAS, except in the Stratify wizard (screen 2) the user is given the option to separate person-time and events occurring during the lag period into a separate group. By default, LTAS.NET will combine the person-time with other unexposed person-time and events into the lowest exposure group.
Time since first exposed	Calculated based on the lagged exposure (calculated as time since date x_i where date x_i is the actual date first exposed plus the length of the lag period). Persons with negative TSFE are lagged out.	Calculated based on the first lagged exposure (calculated as time since date x_i where date x_i is the actual date first exposed plus the length of the lag period, or zero, whichever is greater).
Time since last exposure	Not applicable	Calculated based on the last lagged exposure (calculated as time since date x_l where date x_l is the actual date last exposed plus the length of the lag period, or zero, whichever is greater).
Minimum exposure option	Minimum exposure option is disabled if the user specified a nonzero lag	Same as PC-LTAS
Handling of multiple exposures	Not applicable	The user is given the option to specify a unique non-zero lag for each time-dependent exposure variable (i.e., for each variable in the History file).

Table 1b. Computation of lags in PC-LTAS and LTAS.NET

Construct	Under a specified lag of *n* days: [a]	
	PC-LTAS	LTAS.NET

Lagged cumulative exposure	For day *d* in the interval [risk.begin.date, risk.begin.date + *n* days), lagged.cumulative.exposure(*d*) = 0. For days *d* ≥ risk.begin.date + *n* days, lagged.cumulative.exposure(*d*) = cumulative.exposure(*d*-*n* days).	For day *d* in the interval [risk.begin.date, risk.begin.date + *n* days), lagged.cumulative.exposure(*d*) = 0. For days *d* ≥ risk.begin.date + *n* days, lagged.cumulative.exposure(*d*) = cumulative.exposure(*d*-*n* days).
Events	Events occurring in the interval [risk.begin.date, risk.begin.date + *n* days) are assigned to an unexposed group which is treated separately from the lowest exposure group. Otherwise, events occurring on day *d* ≥ risk.begin.date + *n* days, are assigned to an exposure group based on the lagged cumulative exposure on day *d*.	Events occurring in the interval [risk.begin.date, risk.begin.date + *n* days) may be assigned to the lowest cumulative exposure group (default), or the user may assign them to an unexposed group treated separately from the lowest exposure group. Under this option, the unexposed group is not used for SRR or trend calculations. [b] Otherwise, events occurring on day *d* ≥ risk.begin.date + *n* days, are assigned to an exposure group based on the lagged cumulative exposure on day *d.*
Person-time	Person-time occurring in the interval [risk.begin.date, risk.begin.date + *n* days) is assigned to an unexposed group which is treated separately from the lowest exposure group. Otherwise, person-time on day *d* ≥ risk.begin.date + *n* days, is assigned to the exposure group on day *d*-*n*.	Person-time occurring in the interval [risk.begin.date, risk.begin.date + *n* days) may be assigned to the lowest cumulative exposure group (default), or the user may assign it to an unexposed group treated separately from the lowest exposure group. Under this option, the unexposed group is not used for SRR or trend calculations. Otherwise, person-time on day *d* ≥ risk.begin.date + *n* days, is assigned to the exposure group on day *d*-*n*.

Lagged time since first exposure	TSFE is calculated as the length of time since the date.first.exposed + *n* days. Persons with negative TSFE at their risk.end.date are considered to be lagged out which means that all of their person- time and any observed events are assigned to the unexposed group.	TSFE is calculated as the length of time since the date.first.exposed + *n* days or zero, whichever is greater.
Lagged time since last exposure	Not available in PC-LTAS.	TSLE is calculated as the length of time since the date.last.exposed + *n* days, or zero, whichever is greater.

a Note that the lag period in PC-LTAS can be specified in terms of a number of days, months or years while the lag period in LTAS.NET must be in terms of days or years.

LTAS.NET and PC-LTAS calculate cumulate exposure with a non-zero lag in a similar manner (see Table 1). In the Stratify step, the user has the option of specifying a non-zero lag time which is used to calculate a lagged cumulative exposure. For a specified lag period of *n* days, the lagged cumulative exposure on day *d* is equal to the unlagged cumulative exposure on day *d* - *n*. Thus, for a specified lag period of *n* days, exposure that occurs during the *n* days prior to day *d* is ignored when calculating cumulative exposure on day *d* and person-time and events (death or disease) on a given day are assigned to the cumulative exposure group achieved *n* days earlier. As a worker moves through calendar time, on any given day *d*, exposure that occurs during the specified lag (e.g., *n* days) prior to day *d* is ignored when calculating cumulative exposure at day *d*. As a consequence, exposure that occurs in the *n* days prior to the end of follow-up will not contribute to the total lagged cumulative exposure. For example, if a 10-year lag period is specified, exposure during the 10 years prior to *d* is not considered when the lagged cumulative exposure on day *d* is determined and exposure in the 10 years (approximately 3653 days) prior to the end of follow-up will not count toward the total lagged cumulative exposure. In LTAS.NET, lags may be specified (and are calculated independently) for each time-dependent exposure of interest.

When a lag is specified, TSFE and TSLE (LTAS.NET only) are calculated based on the lagged cumulative exposure values. That is, for TSFE (or TSLE), person-time is grouped into the categories for the length of time elapsed since the first (or last) exposure was received under the assumed lag. For example, if a ten year lag period is specified, TSFE would be calculated as zero for the first ten years since first exposure (and before exposure begins), and would not begin to be counted until the first day after the ten-year lag period had expired. Also note that in this situation, TSLE would be calculated as zero until ten years had lapsed following the last lagged exposure. Both LTAS.NET and PC-LTAS do not allow the user to specify a minimum exposure if a non-zero lag is specified.

One important difference between the lagging methods used in the two life table programs is that, for LTAS.NET, the default setting is for lagged-out events and person-time (i.e., those that occur within the lag period of time after follow-up begins), to accrue to the lowest exposure category and not to be separated into a unique zero exposure category as in PC-LTAS (Steenland et al. 1990). The reason for making this change was to permit the lagged-out events and person-time to be treated as other unexposed events and person-time, which would accrue in the lowest exposure category (as no exposure minimums are allowed in lagged analyses). This change also makes the lagged analyses in LTAS.NET more compatible with approaches used in other statistical software packages, such as Epicure (Preston et al. 1993). An example of such an analysis is provided in Checkoway et al. (1997), with comparisons between lags of zero and fifteen years for silica exposures and lung cancer and non-malignant respiratory disease (see also Checkoway et al. 2004, p. 170). However, if the user desires to assign these events and person-time to a separate, unexposed, group (as in PC-LTAS), a box in screen 2 of the Stratify wizard permits this option. In this case, the events and person-time in this separate unexposed category are not included in SRR or trend test calculations.

Simple examples of the allocation of person-time and events after application of lags in LTAS.NET (using the default option for handling lagged-out events and person-time) are provided below. In these examples,

- cumulative exposure is stratified using threshold values of 0, 5 and 1400 unit-days resulting in cumulative exposure groups of 0-<5 unit-days, 5-<1400 unit-days, and 1400+ unit-days;

- TSFE is stratified into 5-year groups like 0-<5 years, 5-<10 years, 10-<15 years, et cetera; and

- TSLE is stratified into 5-year groups like 0-<5 years, 5-<10 years, 10-<15 years, et cetera.

Example 1a: Worker exposure and follow-up all occur within the first ten years of observation, with a lag of zero years.

				Exposure information			
Worker	DOB	VS	DLO	Risk begin	Start	End	Level (daily)
0009	12/31/1933	D	12/30/1971	12/31/1961	12/31/1961	1/1/1962	2

Solution: In order to properly allocate the person-time and events, transition dates are needed for cumulative exposure, TSFE and TSLE. Cumulative exposure is 4 unit-days, and since the first cutpoint for cumulative exposure is 5 unit-days, there are no transition dates for cumulative exposure; TSFE has a transition date of 12/31/1961 + 5 years = 12/31/1966 - all others are outside the time at risk and are discarded; TSLE has a transition date of 1/1/1962 + 1 day + 5 years = 1/2/1967 - all others are outside the time at risk and are discarded. For worker 0009, time at risk is stratified using these transition dates as below:

Worker	Start		End	PDAR	Cumulative exposure	TSFE	TSLE
0009		12/31/1961	01/01/1962	2	0-<5 unit-days	0-<5Y	0→0-<5Y
0009		01/02/1962	12/30/1966	1824 [a]	0-<5 unit-days	0-<5Y	0-<5Y
0009	*	12/31/1966	01/01/1967	2	0-<5 unit-days	5-<10Y	0-<5Y
0009	*	01/02/1967	12/30/1971	1824 [b]	0-<5 unit-days	5-<10Y	5-<10Y

* denotes a transition date

a PDAR = 364 + 365 + 366 (1964 was a leap year) + 365 + 364 = 1824 days

b PDAR = 364 + 366 (1968 was a leap year) + 365 + 365 + 364 = 1824 days

Person-time: Worker 0009 is at risk for 3652 days: all 3652 days are assigned to the cumulative exposure category 0-<5 unit-days; for TSFE, 1826 days are assigned to 0-<5Y and 1826 days are assigned to 5-<10Y; for TSLE, the two positive exposure days are assigned to 0-<5Y resulting in a total of 1828 days assigned to 0-<5Y and 1824 days are assigned to 5-<10Y.

Deaths: The death is assigned to the 0-<5 unit-days cumulative exposure category, 5-<10Y TSFE and 5-<10Y TSLE.

Example 1b: Similar to example 1a, except exposure begins before follow-up.

				Exposure information			
Worker	DOB	VS	DLO	Risk begin	Start	End	Level (daily)
0010	12/31/1933	D	12/30/1971	12/31/1961	12/31/1959	01/1/1962	2

Solution: The transition dates for cumulative exposure, 1/3/1960 and 11/30/1961, are discarded since they are outside the time at risk; the transition dates for TSFE are 12/31/1959 + 5 years = 12/31/1964 and 12/31/1959 + 10 years = 12/31/1969; the transition date for TSLE is 1/1/1962 + 1 day + 5 years = 1/2/1967. Using these transition dates, time at risk for worker 0010 is stratified as below:

It is important to note that the first transition date for cumulative exposure, although it ends up being discarded, is 1/3/1960, not 1/2/1960. At the end of the day on 12/31/1959, cumulative exposure is 2 unit-days; at the end of the day on 1/1/1960, cumulative exposure is 4 unit-days; and at the end of the day on 1/2/1960, cumulative exposure is 6 unit-days. Transition dates are begin events, so the transition date is the first date on which cumulative exposure (at the beginning of the day) is greater than or equal to the threshold value, so in this example the transition day is 1/3/1960.

Worker		Start	End	PDAR	Cumulative exposure group	TSFE group	TSLE group
0010		12/31/1961	01/01/1962	2	1400+	0-<5Y	0→0-<5Y
0010		01/02/1962	12/30/1964	1094 [a]	1400+	0-<5Y	0-<5Y
0010	*	12/31/1964	01/01/1967	732 [b]	1400+	5-<10Y	0-<5Y
0010	*	01/02/1967	12/30/1969	1094 [c]	1400+	5-<10Y	5-<10Y
0010	*	12/31/1969	12/30/1971	730 [d]	1400+	10-<15Y	5-<10Y

* denotes a transition date

a PDAR = 364 + 365 + 365 (1964 was a leap year) = 1094 days

b PDAR = 1 + 365 + 365 + 1 = 732 days

c PDAR = 364 + 366 (1968 was a leap year) + 364 = 1094 days

d PDAR = 1 + 365 + 364 = 730 days

Person-time: Worker 0010 is at risk for 3652 days: all 3652 days are assigned to the cumulative exposure category 1400+ unit-days; for TSFE, 1096 days are assigned to 0-<5Y, 1826 days to 5-<10Y, and 730 days to 10-<15Y; for TSLE, 1828 days are assigned to 0-<5Y and 1824 days to 5-<10Y.

Death: The death is assigned to the 1400+ unit-days cumulative exposure category, 10-<15Y TSFE and 5-<10Y TSLE.

Example 1c: Same as 1b, except with a lag of ten years.

Solution: Without the lag, the transition date to enter the 5-<1400 unit-days cumulative exposure category is 1/3/1960 and the transition date to enter the 1400+ cumulative exposure category is 11/30/1961. With a ten year lag, the transition dates for cumulative exposure become 1/3/1960 + 10 years = 1/3/1970 and 11/30/1961 + 10 years = 11/30/1971; the transition date for TSFE, (12/31/1959 + 10 years) + 5 years = 12/31/1974, is discarded because it is outside the time at risk; the transition date for TSLE, (1/1/1962 + 10 years + 1 day) + 5 years = 1/2/1977, is discarded because it is outside the time at risk. Using these transition dates, time at risk for worker 0010, with a lag of ten years, is stratified below:

Worker	Start		End	PDAR	Lagged cumulative exposure group	Lagged TSFE group	Lagged TSLE group
0010		12/31/1961	12/30/1969	2922 [a]	0 → 0-<5	0 → 0-<5Y	0 → 0-<5Y
0010		12/31/1969	01/02/1970	3	0-<5	0-<5Y	0 → 0-<5Y
0010	*	01/03/1970	11/29/1971	696 [b]	5-< 1400	0-<5Y	0 → 0-<5Y
0010	*	11/30/1961	12/30/1971	31	1400+	0-<5Y	0 → 0-<5Y

* denotes a transition date

a PDAR = 1 + 365 + 365 + 366 (1964 was a leap year) + 365 + 365 + 365 + 366 (1968 was a leap year) + 364 = 2922 days

b PDAR = 363 + 333 = 696 days

Person-time: Worker 0010 is still at risk for 3652 days but with a lag of 10 years, they are allocated differently: 2922 days are considered unexposed and are included in the 0-<5 lagged cumulative exposure group; three days are assigned to the 0-<5 exposure category; 696 days are assigned to the 5-<1400 exposure category; and 31 days are assigned to the 1400+ unit-days cumulative exposure category. All 3652 days are assigned to 0-<5Y TSFE and 0-<5Y TSLE.

Death: The death is assigned to the 1400+ unit-days cumulative exposure category, 0-<5Y TSFE and 0-<5Y TSLE.

Example 2a: Worker exposure occurs within the first ten years of observation, but follow-up extends beyond ten years, with a lag of zero years.

```
Exposure information
```

```
Worker  DOB         VS  DLO         Risk begin Start      End        Level (daily)
0011    12/31/1933  D   12/30/1981  12/31/1961 12/31/1959 01/1/1962  2
```

Solution: The transition dates for cumulative exposure, 1/3/1960 and 11/30/1961, are discarded since they are outside the time at risk; the transition dates for TSFE are 12/31/1959 + 5 years = 12/31/1964, 12/31/1959 + 10 years = 12/31/1969, 12/31/1959 + 15 years = 12/31/1974, 12/31/1959 + 20 years = 12/31/1979; the transition dates for TSLE are 1/1/1962 + 1 day + 5 years = 1/2/1967, 1/1/1962 + 1 day + 10 years = 1/2/1972, 1/1/1962 + 1 day + 15 years = 1/2/1977. Time at risk for worker 0011 is stratified below:

Worker	Start		End	PDAR	Cumulative exposure group	TSFE group	TSLE group
0011		12/31/1961	01/01/1962	2	1400+	0-<5Y	0→0-<5Y
0011		01/02/1962	12/30/1964	1094 [a]	1400+	0-<5Y	0-<5Y
0011	*	12/31/1964	01/01/1967	732 [b]	1400+	5-<10Y	0-<5Y
0011	*	01/02/1967	12/30/1969	1094 [c]	1400+	5-<10Y	5-<10Y
0011	*	12/31/1969	01/01/1972	732 [d]	1400+	10-<15Y	5-<10Y
0011	*	01/02/1972	12/30/1974	1094 [e]	1400+	10-<15Y	10-<15Y
0011	*	12/31/1974	01/01/1977	733 [f]	1400+	15-<20Y	10-<15Y
0011	*	01/02/1977	12/30/1979	1093 [g]	1400+	15-<20Y	15-<20Y
0011	*	12/31/1979	12/30/1981	731 [h]	1400+	20-<25Y	15-<20Y

* denotes a transition date

a PDAR = 364 + 365 + 365 (1964 was a leap year) = 1094 days

b PDAR = 1 + 365 + 365 + 1 = 732 days

c PDAR = 364 + 366 (1968 was a leap year) + 364 = 1094 days

d PDAR = 1 + 365 + 365 + 1 = 732 days

e PDAR = 365 (1972 was a leap year) + 365 + 364 = 1094 days

f PDAR = 1 + 365 + 366 (1976 was a leap year) + 1 = 733 days

g PDAR = 364 + 365 + 366 = 1093 days

h PDAR = 1 + 366 (1980 was a leap year) + 364 = 731 days

Person-time: Worker 0011 is at risk for 7,305 days: all days are assigned to the cumulative exposure category 1400+ unit-days; for TSFE, 1096 days are assigned to 0-<5Y, 1826 days to 5-<10Y, 1826 days to 10-<15Y, 1826 days to 15-<20Y and 731 days to 20-<25Y; for TSLE, 1828 days are assigned to 0-<5Y, 1826 days to 5-<10Y, 1827 days to 10-<15Y and 1824 days to 15-<20Y.

Death: The death is assigned to the cumulative exposure category 1400+ unit-days, 20-<25Y TSFE and 15-<20Y TSLE.

Example 2b: Same as 2a, except with a lag of ten years.

Solution: With a ten year lag, the transition dates for cumulative exposure are 1/3/1960 + 10 years = 1/3/1970 and 11/30/1961 + 10 years = 11/30/1971; the transition dates for TSFE are (12/31/1959 + 10 years) + 5 years = 12/31/1974 and (12/31/1959 + 10 years) + 10 years = 12/31/1979; and the transition date for TSLE is (1/1/1962 + 10 years + 1 day) + 5 years = 1/2/1977. With a lag of ten years, time at risk for worker 0011 is stratified below:

Worker	Start		End	PDAR	Cumulative exposure group	TSFE group	TSLE group
0011		12/31/1961	12/30/1969	2922 [a]	0 → 0-<5	0 → 0-<5Y	0 → 0-<5Y
0011		12/31/1969	01/02/1970	3 [b]	0-<5	0-<5Y	0 → 0-<5Y
0011	*	01/03/1970	11/29/1971	696 [c]	5-<1400	0-<5Y	0 → 0-<5Y
0011	*	11/30/1961	01/01/1972	33 [d]	1400+	0-<5Y	0 → 0-<5Y
0011		01/02/1972	12/30/1974	1094 [e]	1400+	0-<5Y	0-<5Y
0011	*	12/31/1974	01/01/1977	733 [f]	1400+	5-<10Y	0-<5Y
0011	*	01/02/1977	12/30/1979	1093 [g]	1400+	5-<10Y	5-<10Y
0011	*	12/31/1979	12/30/1981	731 [h]	1400+	10-<15Y	5-<10Y

* denotes a transition date

a PDAR = 1 + 365 + 365 + 366 (1964 was a leap year) + 365 + 365 + 365 + 366 (1968 was a leap year) + 364 = 2922 days

b PDAR = 1 + 2 = 3 days

c PDAR = 363 + 333 = 696 days

d PDAR = 1 + 31 + 1 = 33 days

e PDAR = 365 (1972 was a leap year) + 365 + 364 = 1094 days

f PDAR = 1 + 365 + 366 (1976 was a leap year) + 1 = 733 days

g PDAR = 364 + 365 + 366 = 1093 days

h PDAR = 1 + 366 (1980 was a leap year) + 364 = 731 days

Person-time: Worker 0010 is still at risk for 7,305 days but with a lag of 10 years, the days are allocated differently: 2922 days (approximately 8 years) are unexposed, with zero TSFE and zero TSLE, but are included in the 0-<5 unit-days cumulative exposure group and 0-<5Y TSFE and TSLE groups. Of the 12 years with positive exposure, 3 days are assigned to the cumulative exposure category of 0-<5 unit-days, 696 days to 5-<1400 unit-days, and 3684 days to 1400+ unit-days. For TSFE, 4748 days are assigned to 0-<5Y, 1826 days to 5-<10Y, and 731 days to 10-<15Y. For TSLE, 5481 days are assigned to 0-<5Y and 1824 days to 5-<10Y.

Death: The death is assigned to the 1400+ cumulative exposure category, 10-<15Y TSFE and 5-<10Y TSLE.

Appendix B - Statistics in LTAS

Purpose:

The purpose of this appendix is (a) to provide documentation for key statistics computed by the NIOSH Life Table Analysis System (LTAS) including the standardized mortality ratio (SMR) and the standardized rate ratio (SRR) and (b) to provide details regarding the implementation of the SMR (including confidence intervals and tests of significance) and the SRR (including confidence intervals, tests of significance and trend tests) in both PC-LTAS and LTAS.NET.

Outline:

1. STANDARDIZED RATE

1.1 DEFINITION

A standardized rate (SR) over strata is given by

$$SR = \sum_i W_i R_i / \sum_i W_i \quad [1]$$

where i is an index for the strata, W_i is the weight for stratum i, and R_i is the rate for stratum i (Rothman, 1986, page 44). Since the sum of the weights appears in the denominator of equation [1], it is not necessary that the weights sum to 1.

2. STANDARDIZED MORTALITY RATIO

2.1 DEFINITION

The standardized mortality ratio (SMR) is defined as the ratio of the standardized rate in a cohort to the standardized rate in a reference population, using the person-years at risk (PYAR) in the cohort to weight the rates.

Let i be an index for the strata, A_i be the number of observed deaths in stratum i, T_i be the number of PYAR in stratum i, and $\Re_i$ be the rate for the reference population in stratum i.

The SR in the cohort is given by

$$SR_{cohort} = \frac{\sum_i (T_i)\left(\frac{A_i}{T_i}\right)}{\sum_i T_i} \quad [2],$$

the SR in the reference population is given by

$$SR_{ref} = \frac{\sum_i (T_i)(\Re_i)}{\sum_i T_i} \quad [3],$$

and the SMR is given by, after simplification,

$$SMR = \frac{SR_{cohort}}{SR_{ref}} = \frac{\sum_i (A_i)}{\sum_i (T_i)(\Re_i)} = \frac{D}{E} \quad [4],$$

where D is the observed number of deaths in the cohort and E is the expected number of deaths in the cohort. The SMR can be intuitively understood as the ratio of the observed number of deaths in the cohort to the expected number of deaths in the cohort, if the referent population rates were applied to the PYAR in the cohort (Breslow and Day (1987), page 65).

2.2 CONFIDENCE INTERVAL for the SMR

The SMR statistic, as calculated in equation [4], is an estimate of the true SMR parameter in the population from which the cohort was selected. In the following, SMR is used to denote either the statistic or the parameter, depending on the context. To obtain a $(1 - \alpha)\times 100\%$ confidence interval for the true, but unknown SMR, assume that the observed number of deaths in the cohort is a Poisson random variable with mean μ and use properties of the Poisson distribution to obtain a $(1 - \alpha)\times 100\%$ confidence interval for μ, (μ_L, μ_U). Here, μ_L and μ_U are given by the solutions to the following equations

$$\sum_{x=D}^{\infty} \frac{e^{-\mu_L}\mu_L^x}{x!} = \frac{\alpha}{2} \quad \text{and} \quad \sum_{x=0}^{D} \frac{e^{-\mu_U}\mu_U^x}{x!} = \frac{\alpha}{2} \quad [5]$$

where D is the observed number of deaths in the cohort. Finally, a $(1 - \alpha)\times 100\%$ confidence interval for the SMR is given by (λ_L, λ_U), where $\lambda_L = \mu_L / E$, $\lambda_U = \mu_U / E$ and E is the expected number of deaths in the cohort.

These equations can be difficult to solve for μ_U and μ_L, especially when D is large. Breslow and Day (1987, pages 69-72) provide a formula based on Byar's approximation to the exact Poisson test:

$$\mu_L = D\left(1 - \frac{1}{9D} - \frac{Z_{\alpha/2}}{3\sqrt{D}}\right)^3 \quad \text{and} \quad \mu_U = (D+1)\left(1 - \frac{1}{9(D+1)} + \frac{Z_{\alpha/2}}{3\sqrt{D+1}}\right)^3 \quad [6]$$

where D is the observed number of deaths in the cohort and $Z_{\alpha/2}$ denotes the $100\times(1-\alpha/2)$ percentile of the standard normal distribution. This formula is also presented in Rothman and Boice (1979).

2.3 SIGNIFICANCE TEST for the SMR

To perform a significance test for the true, but unknown, population SMR, the null hypothesis is H_0: SMR = 1. The alternative hypothesis can be one-sided (H_1: SMR < 1 or H_1: SMR > 1) or two-sided (H_1: SMR $\neq$ 1). An exact test of significance for the SMR is based on the Poisson distribution. For level of significance α, the expected number of deaths (E) is compared to critical values which are a function of the observed number of deaths (D) and the level of significance. The critical value for the one-sided test (H_1: SMR < 1) is given by *C* where

$$\sum_{x=0}^{D} \frac{e^{-C}C^x}{x!} = \alpha \quad [7],$$

and H_0 is rejected if E > C.

The critical value for the one-sided test (H_1: SMR > 1) is given by C where

$$\sum_{x=D}^{\infty} \frac{e^{-C}C^x}{x!} = \alpha \quad [8],$$

and H_0 is rejected if E < C.

The critical values for the two-sided test (H_1: SMR $\neq$ 1) are given by C_1 and C_2 where

$$\sum_{x=D}^{\infty} \frac{e^{-C_1}C_1^x}{x!} = \alpha/2 \quad \text{and} \quad \sum_{x=0}^{D} \frac{e^{-C_2}C_2^x}{x!} = \alpha/2 \quad [9],$$

and H_0 is rejected if $E < C_1$ or $E > C_2$.

These equations can be difficult to solve for C, especially when D is large. Breslow and Day (1987, pages 68-69) provide an approximate test of significance for the SMR that is based on an extremely accurate approximation to the exact Poisson test. This test is also presented in Rothman and Boice (1979). The test statistic is given by

$$\chi = \sqrt{9\tilde{D}}\left\{1 - \frac{1}{9\tilde{D}} - \left(\frac{E}{\tilde{D}}\right)^{1/3}\right\} \quad [10]$$

where D is the observed number of deaths, E is the expected number of deaths, $\tilde{D}$ = D if D > E, and $\tilde{D}$ = D + 1, otherwise. The significance of the test can be obtained by comparing the value of the test statistic, χ, to quantiles of the standard normal distribution.

2.4 SMR in PC-LTAS

PC-LTAS calculates the SMR using equation [4]. In general, PC-LTAS uses two gender strata (male, female), two race strata (White, All Other Races), and 15 age strata (15-19 years, 20-24 years, 25-29 years, &, 80-84 years, 85+ years); however the actual number of strata is dependent on the rate file selected. The number of 5-year calendar year strata is dependent on the rate file selected. In PC-LTAS, standardization is performed over the age and calendar year strata for gender-race specific SMRs; over the gender, race, age, and calendar year strata for overall SMRs; and over the gender, race, age, and calendar year strata for stratifier-specific SMRs, such as cumulative exposure and time since first exposure (TSFE) SMRs.

The SMR is undefined when both the observed and expected numbers of deaths are equal to zero. The SMR is infinite when the observed number of deaths is non-zero but the expected number of deaths is zero. PC-LTAS reports the SMR as 0.0000 in both of these situations.

2.4.1 PC-LTAS and the SMR CONFIDENCE INTERVAL

In the Analyze step of PC-LTAS, the user must specify the desired confidence level. The choices are 90% one-sided or 95% two-sided. If 90% one-sided is selected, then PC-LTAS computes 90% (i.e., $\alpha = 0.10$) confidence intervals for the SMRs. It is important to note that these confidence intervals are actually two-sided 90% confidence intervals; however, the end points can be used to manually construct one-sided 95% confidence intervals. If 95% two-sided is selected, then PC-LTAS computes two-sided 95% (i.e., $\alpha = 0.05$) confidence intervals for the SMRs. The method for computing the confidence interval depends on both the observed (D) and expected (E) numbers of deaths.

(a) If D = 0 and E = 0, then PC-LTAS sets both the lower and upper bounds of the confidence interval to zero.

(b) Otherwise, if D = 0 and E > 0, then PC-LTAS computes the confidence interval by setting the lower bound to zero and defining the upper bound as $\lambda_U = 3.00 / E$ for a 90% confidence interval and $\lambda_U = 3.69 / E$ for a 95% confidence interval. Note that $e^{-3.00} \approx 0.0498$ and $e^{-3.69} \approx 0.0250$ (see equation [5]).

(c) Otherwise, if D is in {1, 2, 3, 4, 5}, then an exact confidence interval is calculated using equation [5] as $\lambda_L = SMR / c_L$ and $\lambda_U = SMR / c_U$ where c_L and c_U are values from the look-up table given below:

Observed deaths (D)	90% Confidence Level		95% Confidence Level	
	c_L	c_U	c_L	c_U
1	19.49	0.211	39.53	0.180
2	5.63	0.317	8.26	0.277
3	3.67	0.387	4.85	0.342
4	2.92	0.437	3.67	0.391
5	2.54	0.476	3.09	0.428

(4) Otherwise, if $D \geq 6$, then an approximate confidence interval is calculated using the approximation in equation [6]. Note that PC-LTAS uses $Z_{0.05} = 1.645$ for 90% confidence intervals and $Z_{0.025} = 1.96$ for 95% confidence intervals.

2.4.2 PC-LTAS and the SMR SIGNIFICANCE TEST

PC-LTAS will flag SMRs that are statistically significant. In the Analyze step of PC-LTAS, the user must specify the alternative hypothesis. The choices are 90% one-sided or 95% two-sided. If 90% one-sided is selected, PC-LTAS assumes that the alternative hypothesis of interest is H_1: SMR > 1 and will flag SMRs with a single star (*) to indicate One-Sided P < 0.05 or a double star (**) to indicate One-Sided P < 0.01. If 95% two-sided is selected, then PC-LTAS will flag SMRs with a single star (*) to indicate Two-Sided P < 0.05 or a double star (**) to indicate Two-Sided P < 0.01. The method for determining which SMRs should be flagged as significant depends on the observed (D) and expected (E) numbers of deaths.

(a) If D > 0 and E = 0, then PC-LTAS flags the SMR with a double star (**).

(b) Otherwise, if D is in {0, 1, 2, 3, 4, 5} then PC-LTAS flags the SMR based on critical values of the exact Poisson distribution using a look-up table.

For a one-sided test, PC-LTAS compares E to the critical value (C) from equation [8]. If E is less than C (based on level of significance 0.01) then PC-LTAS flags the SMR with a double star (**); otherwise, if E is less than C (based on level of significance 0.05) then PC-LTAS flags the SMR with a single star (*); otherwise, the SMR is not flagged.

For a two-sided test, PC-LTAS compares E to the critical values (C_1 and C_2) from equation [9]. If E is less than C_1 or if E is greater than C_2 (based on level of significance 0.01) then PC-LTAS flags the SMR with a double star (**); otherwise, if E is less than C_1 or if E is greater than C_2 (based on level of significance 0.05) then PC-LTAS flags the SMR with a single star (*); otherwise, the SMR is not flagged.

When D = 0, equations [7] and [8] are unsolvable for C and equation [9] is unsolvable for C_1. The one-sided test should not be rejected; the two-sided test should only be rejected if the expected number of deaths is large enough. PC-LTAS handles these situations correctly. The actual values used by PC-LTAS to determine significance are given in the table below:

Observed deaths (D)	Confidence level					
	90% - one-sided		95% - two-sided			
	* C	** C	* C_1	* C_2	** C_1	** C_2
0	0.0	0.0	0.0	3.69	0.0	5.30
1	0.0513	0.0101	0.0253	5.57	0.00501	7.43
2	0.0355 [a]	0.149	0.242	7.22	0.103	9.27
3	0.818	0.436	0.619	8.77	0.338	10.98
4	1.370 [b]	0.823	1.09	10.13 [c]	0.672	12.59
5	1.97	1.28	1.62	11.67	1.080 [d]	14.15

a Value was specified incorrectly in PC-LTAS; the correct value is 0.355.

b Value was specified incorrectly in PC-LTAS; the correct value is 1.366.

c Value was specified incorrectly in PC-LTAS; the correct value is 10.24.

d Value was specified incorrectly in PC-LTAS; the correct value is 1.078.

(c) Otherwise, if $D \geq 6$, then PC-LTAS flags the SMR based on an approximation to the Poisson distribution. The test statistic is computed using equation [10]. For a one-sided test, the value of χ is compared to critical values of 1.645 and 2.326 for significance levels of 0.05 and 0.01, respectively; if the value of χ is greater than the critical value, the SMR will be flagged as significant. For a two-sided test, the absolute value of χ is compared to critical values of 1.96 and 2.576 for significance levels of 0.05 and 0.01, respectively; if the absolute value of χ is greater than the critical value, then the SMR will be flagged as significant. *Note that due to a programming error, PC-LTAS did not calculate χ correctly in all situations, resulting in some non-significant p-values being flagged as significant.*

2.5 SMR in LTAS.NET

LTAS.NET computes SMRs using the PC-LTAS algorithm with the following modifications:

(a) When D = 0 and E = 0, LTAS.NET reports a blank for the SMR. The rationale for this change is that the quantity 0/0 is undefined.

(b) When D > 0 and E = 0, LTAS.NET reports infinity for the SMR. The rationale for this change is that, mathematically, for D>0, the quantity D/0 is defined as ∞ (infinity).

2.5.1 LTAS.NET and the SMR CONFIDENCE INTERVAL

LTAS.NET computes confidence intervals for SMRs using the PC-LTAS algorithm with the following modifications:

(c) LTAS.NET will compute 99% confidence intervals for the SMRs in addition to 90% and 95% confidence intervals based on a user-specified option.

(d) When D = 0 and E = 0, LTAS.NET reports blanks for the lower and upper bounds of the confidence interval. The rationale for this change is that when D = 0 and E = 0, the SMR is undefined and therefore the confidence interval for the SMR is also undefined.

(e) When D > 0 and E = 0, LTAS.NET will report infinity for the lower and upper bounds of the confidence interval.

(f) If D = 0 and E > 0, then LTAS.NET computes the confidence interval by setting the lower bound to zero and defining the upper bound as $\lambda_U = 2.995732 / E$ for a 90% confidence interval, $\lambda_U = 3.688879 / E$ for a 95% confidence interval, and $\lambda_U = 5.2983174 / E$ for a 99% confidence interval. Note that $e^{-2.995732} \approx 0.0500000$, $e^{-3.688879} \approx 0.0250000$, and $e^{-5.2983174} \approx 0.005000000$ (see equation [5]).

(g) The look-up table for calculating exact confidence intervals for the SMR was corrected, updated with additional significant figures, and expanded to include all D valus less than 11:

Observed deaths (D)	90% Confidence level		95% Confidence level		99% Confidence level	
	c_L	c_U	c_L	c_U	c_L	c_U

1	19.49572575	0.21079860	39.49789021	0.17948026	199.49958229	0.13458716
2	5.62807153	0.31767242	8.25732198	0.27682858	19.3246869	0.21566151
3	3.66886557	0.38691422	4.84909517	0.34218165	8.87932845	0.27328683
4	2.92757531	0.43699041	3.67017808	0.39056441	5.95055201	0.31760930
5	2.53787838	0.47560006	3.07979176	0.42851026	4.63852770	0.35336290
6	2.29619829	0.50665424	2.72492650	0.45943657	3.90393250	0.38314972
7	2.13069326	0.53239576	2.48724129	0.48534685	3.43585688	0.40855411
8	2.00963480	0.55422197	2.31626773	0.50751151	3.11150540	0.43061163
9	1.91684001	0.57305801	2.18692201	0.52678394	2.87319412	0.45003548
10	1.84318013	0.58954550	2.08533669	0.54376326	2.69039804	0.46733716

2.5.2 LTAS.NET and the SMR SIGNIFICANCE TEST

LTAS.NET will flag SMRs for significance using the PC-LTAS algorithm with the following modifications:

(h) LTAS.NET flags significance based on *two*-sided p-values, regardless of the confidence level requested.

(i) The look-up table for determining significance was corrected, updated with additional significant figures, and expanded to include all D values less than 11:

D (Observed deaths)	**Two-sided significance level**			
	$\alpha = 0.05$ (*)		$\alpha = 0.01$ (**)	
	C_1	C_2	C_1	C_2
0	0	3.6887945	0	5.29831737
1	0.02531781	5.57164339	0.00501254	7.43012950
2	0.24220928	7.22468767	0.10349455	9.27379209

3	0.61867212	8.76727307	0.33786339	10.97747750
4	1.08986537	10.24158868	0.67220654	12.59408979
5	1.62348639	11.66833208	1.07792824	14.14975941
6	2.20189425	13.0594740	1.53691182	15.6596748
7	2.81436305	14.4226754	2.03733748	17.1335933
8	3.45383218	15.7631892	2.57110272	18.5782257
9	4.11537310	17.0848035	3.13240234	19.9984232
10	4.79538870	18.3903560	3.71692213	21.3978275

(j) The approximation formula for determining significance based on 6 or more observed deaths was not implemented correctly in PC-LTAS, resulting in SMRs that were not statistically significant occasionally being flagged as significant. In addition to expanding the look-up table, LTAS.NET implements the approximation formula correctly so that only significant SMRs are identified as such.

3. STANDARDIZED RATE RATIO (SRR)

3.1 DEFINITION

A standardized rate ratio (SRR) is a ratio of two SRs. SRRs can be used to compare the death rates in two groups, standardized to the same set of weights (Rothman 1986, page 229; Rothman and Greenland 1998, page 262).

Let G be the number of exposure groups, g be an index for the exposure groups (g = 0,1,2,3,...,G-1), i be an index for the strata (i.e., sex, race, age, calendar year et cetera), A_{gi} be the number of observed deaths in stratum i for exposure group g, T_{gi} be the number of PYAR in stratum i for exposure group g, $T_{.i}$ be the number of PYAR in stratum i, $T_{..}$ be the number of PYAR in the cohort, and W_i be the weight for stratum i.

The standardized rate for exposure group g is given by

$$SR_g = \sum_i W_i \frac{A_{gi}}{T_{gi}} / \sum_i W_i \qquad [11]$$

and the standardized rate ratio for group b relative to group a is given by

$$SRR_{b\ vs\ a} = SR_b / SR_a = \frac{\sum_i W_i \frac{A_{bi}}{T_{bi}} / \sum_i W_i}{\sum_i W_i \frac{A_{ai}}{T_{ai}} / \sum_i W_i} \quad [12].$$

Weights can be assigned to the strata in several different ways. For example, when calculating the SRR for group b relative to group a, the weight for stratum i could be based on the distribution of PYAR in group a. Then the SRR for group b relative to group a would have the interpretation of the expected number of deaths in group a, if group b rates were applied to group a, divided by the observed number of deaths in group a.

3.2 CONFIDENCE INTERVAL for the SRR

The SRR statistic, as calculated in equation [12], is an estimate of the true SRR parameter in the population from which the cohort was selected. In the following, SRR is used to denote either the statistic or the parameter, depending on the context. One method for obtaining a confidence interval for the SRR is to first obtain a confidence interval for ln(SRR), which involves the estimating the variance of ln(SRR), and then transforming the lower and upper bounds of the confidence interval for ln(SRR) into a confidence interval for the SRR (Rothman and Greenland, 1998, page 263). So, to obtain a confidence interval for the SRR of group b relative to group a, first compute the estimate of the variance of the natural log of the SRR,

$$var[\ln(SRR_{b\ vs\ a})] \approx \frac{\sum_i \frac{W_i^2 A_{bi}}{T_{bi}^2}}{\left[\sum_i \frac{W_i A_{bi}}{T_{bi}}\right]^2} + \frac{\sum_i \frac{W_i^2 A_{ai}}{T_{ai}^2}}{\left[\sum_i \frac{W_i A_{ai}}{T_{ai}}\right]^2} \quad [13]$$

and then a confidence interval for $SRR_{b\ vs\ a}$ is given by

$$CI(SRR_{b\ vs\ a}) = \exp\left(\ln(SRR_{b\ vs\ a}) \pm z_{\alpha/2}\sqrt{var[\ln(SRR_{b\ vs\ a})]}\right) \quad [14]$$

where $Z_{\alpha/2}$ denotes the $100\times(1\alpha/2)$ percentile of the standard normal distribution. It is important to note that equation [13] assumes that the standardized rate in group b varies independently of the standardized rate in group a and that the amount of data is at least moderate (Flanders, 1984). The confidence interval relies on large sample theory. Since clear guidance regarding the amount of data required to ensure a reasonable approximation is lacking, the user is instructed to interpret the confidence interval with caution, particularly when the number of deaths in either group a or group b is low.

3.3 SIGNIFICANCE TEST for the SRR

Rothman provides a test statistic for the Mantel-Haenszel test for person-time data of the following null hypothesis

$$H_0: \frac{E\left[\sum_i A_{bi}\right]}{E\left[\sum_i \left[\frac{A_{bi}+A_{ai}}{T_{bi}+T_{ai}}\right] T_{bi}\right]} = 1 \quad [15]$$

where E[ ] denotes expectation (Rothman and Greenland, 1998, page 274). The test statistic is given by

$$\chi^2_{b\ vs\ a} = \frac{\left[\sum_i A_{bi} - \sum_i \frac{(A_{bi}+A_{ai})T_{bi}}{(T_{bi}+T_{ai})}\right]^2}{\sum_i \frac{(A_{bi}+A_{ai})T_{bi}T_{ai}}{(T_{bi}+T_{ai})^2}} \quad [16].$$

The value of the test statistic in equation [16] can be compared to the chi-square distribution with 1 degree of freedom to obtain p-values. Note that unlike the formula for $SRR_{b\ vs\ a}$ and the formula for $CI(SRR_{b\ vs\ a})$, the chi-square formula does *not* involve W_i, the weight for stratum *i*. The tested hypothesis can be shown to be equivalent to

$$H_0: \frac{E\left[\sum_i \frac{A_{bi}}{T_{bi}}\left[\frac{T_{ai}T_{bi}}{T_{ai}+T_{bi}}\right]\right]}{E\left[\sum_i \frac{A_{ai}}{T_{ai}}\left[\frac{T_{ai}T_{bi}}{T_{ai}+T_{bi}}\right]\right]} = 1 \quad [17].$$

3.4 TREND TEST for the SRRs

Rothman describes a method for estimating the linear component of the trend in the SRs across the exposure categories (Rothman, page 336-341). The trend test is based on a weighted regression of the SRs, where the midpoints of the exposure categories are used as the independent variable, the SRs are the dependent variable, and the regression weight is the inverse of the estimated variance of the SRs.

For the weighted least-squares regression, the dependent variable (Y_g) is given by

$$Y_g = SR_g = \sum_i W_i \frac{A_{gi}}{T_{gi}} \Big/ \sum_i W_i \quad [18];$$

the independent variable (X_g) is defined to be the midpoint of exposure group g (in the highest category, the cutoff value plus 50% is used for the midpoint); and the regression weights are given by

$$\omega_g = 1 / Var(SR_g) \quad [19]$$

where

$$Var(SR_g) \cong \frac{\sum_i W_i^2 \frac{A_{gi}}{T_{gi}^2}}{\left[\sum_i W_i\right]^2} \quad [20].$$

The estimated slope is given by

$$\hat{\beta}_1 = \frac{\sum \omega_g X_g Y_g - \frac{\sum \omega_g X_g \sum \omega_g Y_g}{\sum \omega_g}}{\sum \omega_g X_g^2 - \frac{(\sum \omega_g X_g)^2}{\sum \omega_g}} \quad [21]$$

and the estimated intercept, which is not reported but used in the calculation of the standard error of the slope, is given by

$$\hat{\beta}_0 = \frac{\sum_g \omega_g Y_g - \hat{\beta}_1 \sum_g \omega_g X_g}{\sum_g \omega_g} \quad [22].$$

Finally, the estimated standard error of the slope is given by

$$se(\hat{\beta}_1) = \sqrt{\frac{\sum_g \omega_g (Y_g - \hat{\beta}_0 - \hat{\beta}_1 X_g)^2}{(\sum_g \omega_g (X_g - \bar{X})^2)(n-2)}} \quad [23]$$

where n is the number of exposure groups (including the reference group) and $\bar{X}$ is the weighted mean which is given by

$$\bar{X} = \frac{\sum_g \omega_g X_g}{\sum_g \omega_g} \quad [24].$$

3.5 SRR in PC-LTAS

PC-LTAS will report SRRs in the Exposure by TSFE report when direct standardization is requested. Direct standardization is requested by using the right mouse button (or CTRL-X) to select the cause of death category for analysis which places an X next to the cause of death instead of a check mark. PC-LTAS only performs direct standardization for the total category of TSFE (i.e., the total row). PC-LTAS uses two gender strata (male, female), two race strata (White, All Other Races), and 15 age strata (15-19 years, 20-24 years, 25-29 years, ..., 80-84 years, 85+ years); however the actual number of strata is dependent on the rate file selected. The number of 5-year calendar year strata is dependent on the rate file selected. In PC-LTAS, standardization is performed over the age and calendar year strata for gender-race specific SRRs and over the gender, race, age, and calendar year strata for overall SRRs. In the Stratify step, the user can choose to stratify by either duration of exposure or cumulative exposure, in addition to TSFE; however, PC-LTAS does not consider the stratification by TSFE when calculating SRRs (unless a minimum TSFE is specified in which case the SRR analysis would be limited to PYAR with TSFE greater than or equal to the minimum specified TSFE). In the following, exposure refers to either cumulative exposure or duration of exposure, depending on which variable was selected in the Stratify step. The reference group is defined to be the lowest exposure group based on the exposure strata defined in the Stratify step.

PC-LTAS performs direct standardization using weights defined for each stratum as the total stratum-specific PYAR across the exposure groups divided by the total PYAR. That is, using the above notation,

$$W_i = \frac{\sum_{g=0}^{G-1} T_{gi}}{\sum_j \sum_{g=0}^{G-1} T_{gj}} = \frac{T_{0i} + T_{1i} + \ldots + T_{(G-1)i}}{\sum_j T_j} = \frac{T_i}{T_{..}} \quad [25]$$

where j indexes the strata. Note that, by definition, these weights sum to 1. PC-LTAS calculates, but does not report, a SR for each exposure group using these weights (see equation [11]). PC-LTAS reports a SRR for each of the exposure groups relative to the reference group. That is, for each exposure group g, PC-LTAS computes a SRR relative to the reference group as

$$SRR_{g \text{ vs } 0} = SR_g / SR_0 = \frac{\sum_i W_i \frac{A_{gi}}{T_{gi}} \Big/ \sum_i W_i}{\sum_i W_i \frac{A_{0i}}{T_{0i}} \Big/ \sum_i W_i} \quad [26].$$

The SRR for the reference group is defined to be 1,

$$SRR_{0 \text{ vs } 0} = SR_0 / SR_0 = \frac{\sum_i W_i \frac{A_{0i}}{T_{0i}} \Big/ \sum_i W_i}{\sum_i W_i \frac{A_{0i}}{T_{0i}} \Big/ \sum_i W_i} = 1 \quad [27].$$

In addition, PC-LTAS reports a total SRR obtained by first computing the SR for all of the non-reference exposure groups combined which is then divided by the SR for the reference group. That is,

$$SRR \text{ total vs } 0 = \frac{\sum_i W_i \frac{\sum_{g=1}^{G-1} A_{gi}}{\sum_{g=1}^{G-1} T_{gi}} \Big/ \sum_i W_i}{\sum_i W_i \frac{A_{0i}}{T_{0i}} \Big/ \sum_i W_i} \quad [28].$$

Note that PC-LTAS will report SRRs for each exposure group (and total) for a particular race-gender group (such as white males) or for all race-gender groups combined. For the former, *i* indexes over the 5-year age and 5-year calendar year strata, and for the latter, *i* indexes over the gender and race strata in addition to the 5-year age and 5-year calendar year strata.

When the reference category has zero observed deaths, PC-LTAS reports SRRs in all categories as 0.0000.

The PC-LTAS SRR report contains the observed number of deaths, the expected number of deaths, the SMR, the SRR, a confidence interval for the SRR, and a chi-square test. PC-LTAS also reports the estimated slope parameter and standard error for the Rothman test of linear trend in standardized rates in the last row of the report.

3.5.1 PC-LTAS and the SRR confidence interval

PC-LTAS reports a confidence interval for each SRR using equations [13] and [14] with g and 0 substituted for b and a, respectively. PC-LTAS uses critical values of $Z_{\alpha/2} = 1.96$ and $Z_{\alpha/2} = 1.645$ for 95% and 90% confidence intervals, respectively. PC-LTAS reports a confidence interval for each exposure group, including the reference group; however, the confidence interval for the reference group is not valid and should be ignored.

When either of the reference or exposure cells has zero observed deaths, PC-LTAS reports the confidence interval for the SRR as 0.00 - 0.00; however, this is not valid and should be ignored.

3.5.2 PC-LTAS and the SRR test of significance

PC-LTAS reports a chi-square test of significance for the SRR using the Mantel-Haenszel chi-square formula described above in equation [16]. It is important to note here that the Mantel-Haenszel chi-square statistic provides a test of significance for the Mantel-Haenszel risk-ratio estimator, which is equivalent to the SRR that uses weights equal to $W_{iM\text{-}H} = T_{gi} \times T_{0i} / (T_{gi} + T_{0i})$ (see equation [17]), which differ from the weights used by PC-LTAS (see equation [25]). The rationale for choice of the weights used in the Mantel-Haenszel statistic is that the Mantel-Haenszel weights are equivalent to the inverse of $(1/T_{gi} + 1/T_{0i})$, which is proportional to the inverse of the variance of the rate difference (see Kahn and Sempos (page 89) for a more detailed explanation).

Since the chi-square statistic reported by PC-LTAS is based on the Mantel-Haenszel risk ratio, there may be a conflict between the confidence interval for the SRR and the chi-square test reported PC-LTAS. This has been noted in the past by users of PC-LTAS.

Note that if the SR for the reference group is 0, then PC-LTAS reports a chi-square value of 0.

3.5.3 PC-LTAS and SRR TREND TEST

PC-LTAS reports the results of a trend test for the SRRs obtained by performing a weighted least-squares regression (Neter, Kutner, Nachtsheim and Wasserman, 1996, pages 409, 403 and 208) on the SRs with the midpoints of the exposure categories as the independent variable. PC-LTAS reports the estimated slope from equation [21] and the estimated standard error of the slope from equation [23].

When there are no deaths overall, PC-LTAS reports the estimated slope and standard error for the Rothman test of linear trend as 0 and 0. When there are no deaths in one or more exposure categories, PC-LTAS reports an estimated slope, but reports the estimated standard error as 0.

3.6 SRRs in LTAS.NET

LTAS.NET provides SRRs when the SRR report, a grid type report, is requested. Each cell of the LTAS.NET SRR report contains the observed number of deaths, the person-years at risk, the SRR and a confidence interval for the SRR. LTAS.NET computes SRRs using the PC-LTAS algorithm with the following exceptions:

(a) LTAS.NET performs direct standardization using weights defined by each stratum as the total stratum-specific PYAR across the exposure groups. That is, the weights are the numerator of equation [25]. Thus the weights used by LTAS.NET do not sum to 1, as did the weights used by PC-LTAS; however, it can be shown that the two sets of weights produce equivalent results.

(b) While PC-LTAS appends the SRR results to the SMR report, LTAS.NET reports SMRs and SRRs in separate reports. In LTAS.NET, SMRs are available in both list and grid formats; however, SRRs are only available in a grid format.

(c) To obtain SRRs in LTAS.NET, in the Analyze Wizard, the user first specifies the SRR report (with a grid format), then selects two or more variables to be included in the report (called stratifier variables), and finally specifies which of these variables will be placed on the horizontal axis and which will be placed on the vertical axis. Thus, LTAS.NET offers flexibility in selecting the row and column variables (i.e., reports other than Exposure by TSFE are possible).

LTAS.NET produces a single SRR table if only two variables are selected to be in the report. If more than two variables are selected to be in the report, then LTAS.NET produces a separate table for each combination of the selected variables that were not specified as the horizontal and vertical axis variables. For example, suppose that

(i) exposure1-level, exposure1-TSFE, and gender are selected to be in the report;

(ii) exposure1-level is specified as the horizontal axis;

(iii) exposure1-TSFE is specified as the vertical axis; and

(iv) the male and female categories were selected for gender, in addition to the total gender category;

then LTAS.NET reports three SRR tables: one for Gender = All, one for Gender = Male, and one for Gender = Female. SRRs in the first table are standardized to the gender, race, age, and calendar year stratification variables (assuming that the rate file selected for the study includes these stratification variables). SRRs in the second and third tables are standardized to the race, age, and calendar year stratification variables (once again, assuming that the rate file selected for the study includes these stratification variables). All three tables have categories of exposure1-TSFE on the vertical axis and categories of exposure1-level on the horizontal axis which results in SRRs for each category of exposure1-level relative to the reference category.

(d) LTAS.NET computes SRRs for each row in the grid table, not just the total row. That is, SRRs are reported for each level of the vertical axis variable in addition to the total row for the vertical axis variable.

(e) LTAS.NET allows the user to specify the reference group (i.e., the reference category does not necessarily have to be the lowest or first exposure group).

(f) LTAS.NET allows the user to exclude categories of the selected stratifier variables.

(g) LTAS.NET will suppress all SRRs when the reference category contains zero observed deaths. LTAS.NET will report NR in the place of the SRR and a footnote at the bottom of the report will state NR indicates the statistic is Not Reported due to one or more cells will zero deaths.

3.6.1 LTAS.NET and the SRR CONFIDENCE INTERVAL

LTAS.NET will compute confidence intervals for the SRRs using the PC-LTAS algorithm with the following exceptions:

(h) LTAS.NET computes 99% confidence intervals for the SRRs in addition to 90% and 95% confidence intervals based on a user-specified option. LTAS.NET uses critical values of 1.645, 1.96 and 2.576 for 90%, 95% and 99% confidence intervals, respectively.

(i) The confidence interval for the SRR for the reference group is suppressed. The rationale for the change is that the SRR for the reference group is, by definition, 1, and the formula used by PC-LTAS to estimate the confidence interval is not appropriate in that it requires an assumption that the standardized rate for the numerator varies independently of the standardized rate for the denominator, which cannot be met if the numerator and denominator represent the same group.

(j) LTAS.NET reports the standardized rate for the reference group and, when the standardized rate is based on five or more deaths, the margin of error for a Wald-based confidence interval for the standardized rate. The Wald-based confidence interval is given by

$$SR_{ref} \pm Z_{\alpha/2}\, se(SR_{ref})$$

where $Z_{\alpha/2}$ is the 100(1-α) percentile of the standard normal distribution (1.645, 1.96 and 2.576 for 90%, 95% and 99% confidence intervals, respectively), SR_{ref} can be estimated using equation [11] and $se(SR_{ref})$ is the square root of the variance of the SR_{ref} which can be estimated using equation [20], and $Z_{\alpha/2}$ se(SRref) is the margin of error for the confidence interval. Since the confidence interval is based on large sample theory, it is likely not appropriate when the observed number of deaths is small. LTAS.NET will report NP in place of the margin of error when the reference cell contains fewer than 5 deaths.

(k) LTAS.NET will suppress the confidence interval when there are zero deaths in either the reference or exposure categories. The rationale for this change is that zero deaths in the reference category or the exposure category will result in a division by zero error when calculating the estimated variance of the natural logarithm of the SRR (necessary for estimating the confidence interval for the SRR). LTAS.NET will report NR in the place of the confidence interval for the SRR and a footnote at the bottom of the report will state NR indicates the statistic is Not Reported due to one or more cells will zero deaths.

3.6.2 LTAS.NET and the SRR SIGNIFICANCE TEST

LTAS.NET will not report the Mantel-Haenszel chi-square statistic since it is based on different weights than the reported SRR and will not flag the SRR statistic for significance. The user, however, can perform a test of significance by determining whether or not the confidence interval includes 1.

3.6.3 LTAS.NET and the SRR TREND TEST

LTAS.NET computes the slope and standard error for the SRR trend test using the PC-LTAS algorithm with the following exceptions:

(l) LTAS.NET reports the slope and standard error for the SRR trend test for each row in the grid table in addition to the total row.

(m) Since the user can specify any type of variable for the horizontal axis (e.g., while the horizontal axis is typically an exposure type variable, the user could specify a categorical variable like race or smoking status as the horizontal variable) LTAS.NET only reports the results of the SRR trend test if the horizontal axis variable is a level, duration, TSFE or TSLE type variable.

(n) Since an infinite weight would be assigned to any exposure group (including the reference) with zero deaths, LTAS.NET will not report the estimated slope and standard error for the Rothman test of linear trend when any cell has zero observed deaths. LTAS.NET will report NR in the place of the slope and standard error and a footnote at the bottom of the report will state NR indicates the statistic is Not Reported due to one or more cells will zero deaths.

Appendix C - Import File Requirements

This appendix contains a comprehensive list of the import file requirements that are enforced by LTAS along with the actions taken when a requirement is not met.

Requirements are organized by import file and requirement type. Requirement types identify the scope of the requirement and consist of the following:

Requirement Type	Description
File-level	File-level requirements are those that pertain to the file as a whole: generally sort order and required fields.
Person-level	Person-level requirements define a prerequisite for a group of records referring to a single person. For example the requirement that when using an exposure history file, a person must have at least one exposure history occurring during the study period is a person-level requirement.
Record-level	Record-level requirements validate a single record within the file, and involve more than one field within the record such as ensuring that date orders are logical.
Field-level	Field-level requirements involve only a single field within a record. For example, the Gender field must have one of only 2 values.
Multi-source	Multi-source requirements involve data from more than one file, such as the requirement that terminal outcome dates (outcome file) agree with date last observed (person file).

Person Import File Requirements

File-level Requirements

Requirements of the layout of the file.

ID	Requirement	Description	Action
PF10t	Sort Order	File must be sorted in ascending order by ID	Termination.
PF20e	Required fields defined	Must contain all required fields: ID, Gender, Race, Vital Status, DOB, Date Last Observed.	Enforced.
PF30e	Conditionally Required fields defined	Risk Begin Date required unless selected rates are proportional.	Enforced.

Person-level Requirements

Requirements of the set of records associated with a single person.

ID	Requirement	Description	Action
PP20r	Unique ID	Duplicate values are forbidden.	Rejection. (Persons after 1st)

Record-level Requirements

Requirements of the values of more than one field within a single record.

ID	Fields	Requirement	Description	Action
PR10r	DOB DLO	DLO > DOB	Date last observed must occur after date of birth. Ignored for living persons unless “Stop risk at DLO” option is selected.	Rejection.
PR12w	DOB DLO	Age < 100 Years.	An age at DLO (= DLO - DOB) that is 100 years or above may indicate a date error in your data.	Warning.
PR20r	Risk Begin Date DLO	DLO > Risk Begin	Date last observed must occur after risk begin.	Rejection.
PR30w	Risk Begin Date DOB	Risk Begin > DOB	Risk begin date occurs after date of birth.	Warning.
PR40w	Age @ Risk Begin *Age is derived as RiskBegin - Person.DOB. RiskBegin is “In-Rec” risk begin. Compared to global rate 1st age category threshold.	Age in rates	Age at risk begin (= Derived Risk Begin Date - DOB) should be >= min age category in selected rate set. (It’s assumed that all rate sets max age category is open-ended, so age at DLO is not tested. This assumption should be enforced by the rate system.) Data (person-time, outcomes) are maintained in a < min age category (e.g. “Age < 15").	Warning.

Field-level Requirements

Requirements of the values of individual fields.

ID	Field	Requirement	Description	Action
PC10r	ID	Non-Blank	All person records must contain a value.	Rejection.
PC30r	Gender Race	Valid	All person records must contain a value, which must be a specified code for the variable of interest (gender, race).	Rejection.
PC35r	Gender Race	In rates	Gender or Race indicated by the value must be included in the rate set.	Rejection.
PC40x	Gender Race	Selected	Specified value must be included in study.	Exclusion.
PC50d	Vital Status	Valid	Vital status code in the person file must contain a valid code as specified for alive or deceased. Missing or invalid codes set to alive.	Redemption.
PC60x	Vital Status	Selected (deceased)	For PMR-type studies, all persons must be deceased.	Exclusion.
PC70d	DOB DLO	Complete	Dates must contain month, day and year. (DLO not tested if not used. See PC90r.) Missing month is set to 07 (July). Missing day is set to 15.	Redemption. (Missing year is invalid date - rejection.)
PC80r	DOB	Valid	All person records must contain a value, which must be a valid date. Incomplete dates are first redeemed then tested for validity.	Rejection.
PC90r	DLO	Valid	When ‘end risk at DLO’ option selected, all persons must contain a valid date. When not selected, only deceased persons must contain valid date. Incomplete dates are first redeemed then tested for validity.	Rejection.
PC100w	DLO	In rates	Date Last Observed must be after 1st calendar period in rates. Because PR40w assures Risk Begin is in rates, this requirement is not tested for directly.	Warning.
PC120d	DLO	Before Study End	Date Last Observed must occur on or before the study end date. DLO set to study end. Deceased “reincarnated”.	Redemption.

PC140r	Risk Begin	Valid	If present and requested for import, must be valid date.	Rejection.
PC150x	Risk Begin	Before Study End	Risk begin date must occur before study end date.	Exclusion.
PC170r	Fixed Stratifier	Non Blank	All person records must have a non-blank value for each selected fixed stratifier.	Rejection.

Outcome Import File Requirements

File-level Requirements

Note that many of these requirements are new to LTAS.NET and did not exist in PC-LTAS since the outcomes were in the demographics file.

Requirements of the layout of the file.

ID	Requirement	Description	Action
OF10t	Sort Order	Must be sorted in ascending order by Person ID, and within Person ID by Date.	Termination.
OF20e	Required fields	Must contain all required fields: Person ID, Date, Disease Code, and conditionally required fields: Terminal, Underlying.	Enforced.

Person-level Requirements

Requirements of the set of records associated with a single person.

ID	Requirement	Description	Action
OP10r	Proper number of underlying causes.	Deceased have exactly one underlying cause, alive none.	Rejection.

Record-level Requirements

Requirements of the values of more than one field within a single record.

There are currently no Record-level Requirements for the Outcome File.

Field-level Requirements

Requirements of the values of individual fields.

ID	Field	Requirement	Description	Action
OC09r	Person ID	Non-blank		Rejection.
OC10d	Date	Complete	Contains month and day. Missing months are set to 7 (July) and missing days are set to 15.	Redemption.
OC20r	Date	Valid	Must be a complete and valid date. Incomplete dates are redeemed first.	Rejection.
OC22x	Date	<= study end	Outcome must occur on or before study end	Exclusion.
OC23r	Date	In rates	Outcome must occur on or after rate begin date.	Rejection.
OC30d	Disease Code*	Map-able	The disease code in an outcome record must be defined in the LTAS System database for the selected rate set. This means that the code has been mapped to a minor disease category. Unmapped codes are recoded to the residual category.	Redemption.
OC40r	Terminal and/or Underlying	Valid	Must be either 'T' or 'F'.	Rejection.
OC50r	ICD Revision	Valid	If used, must contain a number corresponding to an ICD revision defined in the LTAS System database for the selected rate set.	Rejection.

*This replaces PC-LTAS "Invalid COD" and "ICD look-up prob" rejections with a non-rejection.

History Import File Requirements

File-level Requirements

Requirements of the layout of the file.

ID	Requirement	Description	Action
EF10t	Sort Order	File must be sorted in ascending order by Person ID. Records with the same Person ID must be sorted in ascending order by Begin Date.	Termination.
EF20e	Required fields	Must contain all required fields: Person ID, Begin Date, End Date.	Enforced.
EF30e	Conditionally required fields	Each record must contain at least one additional field beyond the required fields. Additional fields are either Level fields associated with exposure agents, or Category fields associated with categorical stratifiers.	Enforced.

Person-level Requirements

Requirements of the set of records associated with a single person.

ID	Requirement	Description	Action
EP10r	1 or more histories accepted.	When an Exposure History File is used, all persons should have at least one history record.	Rejection. *Necessary especially when using categoricals.*
EP20w	Fewer than 2 histories rejected.	A 2nd rejected history for a given person triggers this exception.	Warning.
EP30r	History period integrity.	No overlaps allowed in the begin / end dates of history periods for a given person. History n+1 begin date must be after nth history end date.	Rejection.
EP40r	1 or more histories in the study period.	At least 1 day of 1 history must be >= Study begin and <= Study end date.	Rejection.

Record-level Requirements

Requirements of the values of more than one field within a single record.

ID	Fields	Requirement	Description	Action
ER10r	Begin Date End Date	Begin Date before End Date		Rejection.

Field-level Requirements

Requirements of the values of individual fields.

ID	Fields	Requirement	Description	Action
EC10r	Person ID	Non-Blank		Rejection.
EC20r	Begin Date End Date	Valid	Must be a complete and valid date.	Rejection.
EC30x	Begin Date	Before Study End	History must begin before the Study End date.	Exclusion.
EC50d	End Date	On or before Study End	History must end on or before Study End date. Dates after the Study are set to the Study End Date.	Redemption.
EC60r	Level	Valid	Must be valid number (integer, floating point or scientific notation) >= 0.0.	Rejection.
EC70r	Category	Non-Blank	Must contain a non-blank character string.	Rejection.

Multi-source Requirements

Requirements of the collection of all records across files for a single person.

ID	File.Fields	Requirement	Description	Action
M1r	Outcome.PersonID History.PersonID Person.ID	Has master.	A person ID in the Outcome or History file must exist in the Person file.	Rejection (outcome or history).
M10r	Person.Vital Status Outcome.Terminal	Vital Status logic	A deceased person must have at least one terminal record in the Outcome File. An alive person must have no terminal records in the Outcome File.	Rejection.
M20r	Person.DOB History.Begin Date	History Begin after DOB	An exposure history must begin after the person's date of birth.	Rejection.
M30r	Person.DLO History.End Date	History End on or before DLO	An exposure history must end on or before the person's date last observed.	Rejection.
M40r	Person.DOB Outcome.Date	Outcome date > DOB	The outcome (i.e., death or disease) date(s) must be after the date of birth.	Rejection.
M50r	Person.DLO Outcome.Date	Outcome date <= DLO	The outcome date(s) must be before the date last observed. Also for terminal outcomes, test becomes EQ, not LE.	Rejection.

Date Field Requirements

Date fields must be in one of the supported formats, must exist, and must not be a date in the future.

Format - To be considered valid a date must conform to one of the following formats:

- MM/DD/YYYY
- MMDDYYYY

Existence - Dates are also checked for calendar validity (that the date actually existed, i.e. month <= 12 etc).

Not Future - Any date beyond "today" will be considered invalid.

Partial dates - are allowed as noted.

Appendix D - Differences between PC-LTAS and LTAS.NET

Creating a New Project

Compatibility with PC-LTAS Multiple Cause of Death Analysis

The following paragraphs summarize the difference between LTAS.NET and PC-LTAS with respect to a multiple cause of death analysis:

In previous versions of the LTAS, causes of death were considered to be either underlying or contributing. For each decedent, an underlying cause of death (UCOD) analysis would only consider a single underlying cause of death; a multiple cause of death (MCOD) analysis would consider 1 or more (up to 10) contributing causes of death in addition to the underlying cause of death. For example, a death certificate may have specified causes A, B, C, and D, of which cause C was indicated as the underlying cause of death; the UCOD analysis would only consider C, but the MCOD analysis would consider C in addition to A, B, and D. PC-LTAS required the input demographics file to specify the underlying cause of death in columns 65-68 and the contributing causes (if available) in columns 69-72 (1st contributing cause), 73-76 (2nd contributing cause), …, 105-108 (10th contributing cause). If the user selected an UCOD rate file, PC-LTAS only used the UCOD, but if a MCOD rate file was selected, PC-LTAS used both the UCOD and the contributing causes of death.

More recently, and especially for death certificates coded to the tenth revision of the ICD, it is possible for the underlying cause of death to be distinct from the separate causes of death listed on the death certificate. For example, a death certificate may have specified causes A, B, C, and D, but the underlying cause of death, as identified by the coding rules, could be a new cause E which is not necessarily equal to one of A, B, C, and D. A specific example of this is when the death certificate lists alcohol dependence (F10.2) and cirrhosis of the liver (K71.7); ICD10 coding rules would specify the underlying cause of death as K70.3 (alcoholic cirrhosis of the liver). In this case, K10.3 would be used for the UCOD analysis, but F10.2 and K71.7 would be used for the MCOD analysis. The difference here is that the set of multiple causes does not include the underlying cause, per se. Most of the time, the set of multiple causes will include the underlying cause, but it is not required to include the underlying cause (and does not for approximately 1.5% of deaths occurring between 2000 and 2002).

Further complicating this is the issue of how multiple causes are handled in creating the rate files. For LTAS.NET, MCOD rate files will treat the MCOD string as completely separate from the UCOD. If the UCOD is included in the MCOD string it will be included in the MCOD rate file. If the UCOD is not included in the MCOD string it will not be included in the MCOD rate file. Thus, it is important that the user's data be handled similarly so that the same cause of death is not double counted in the user's file. In many instances, forcing the UCOD to be included along with the MCOD string would cause the same cohort death to have two causes included in the same LTAS minor category, whereas this would not be the case in the rate file.

Death information is specified in the outcome file in LTAS.NET. The outcome file includes three required fields: a person ID, the date of occurrence and a character string representing a disease code. Each person ID can have 1 or more records in the outcome file. Optional fields include a terminal flag (true or false, indicating whether the disease code represents a death (true) or a diagnosis (false)), an underlying flag (true or false, indicating whether the disease code represents an underlying cause of death (true) or a multiple cause of death (false)), and an ICD revision (a number specifying the revision of the ICD associated with the disease code). If the user selects an UCOD rate file, LTAS.NET will only use the deaths marked with underlying=true; if the user selects a MCOD rate file, LTAS.NET will only use the deaths marked with underlying=false. This implies that for some decedents, the underlying cause of death disease code will be replicated with underlying=false; but for other decedents, the set of disease codes with underlying=false will not contain the disease code with underlying=true.

Import Processing

Compatibility with PC-LTAS Verify Step

The following is a list of differences between the LTAS.NET Import process and the PC-LTAS Verify step:

- Work History File and Exposure File are combined into a single import file: History.
- A History File is required.
- Demographics File is split into two files: Person and Outcome.
- A mapping function allows users to adjust the import for different input file formats.
- There is no DOD (Date of Death) field. The DLO (Last Observation Date) is used to hold this information for deceased persons.
- Non-cancer COD, PCMR error not issued.
- Too many histories error not issued because the system limit has been removed.
- Default exposure used warning not issued. A 'No history accepted' rejection replaces this warning.
- Invalid DLO for VS = dead warning not issued.
- 2 Histories Rejected error is now just a warning.
- Date validation is more complete, generating exceptions for non-existent dates.
- DLO < Rate File Begin is now just a warning. This is because in an "End of Study" run, subjects with a DLO occurring before the rate begin date should be included -- the program assumes their person-time ends at the end of study date.
- Any Vital Status code other than the Alive or Deceased code is treated as Unknown instead of defining a specific Unknown code. All Person records with Unknown codes are redeemed as Alive.
- Age at Risk Begin Date < Lowest Age Category in Rates is now just a warning. The Risk Begin Date is not redeemed like it used to be.
- COD (Cause of Death) validation operates differently to support non-ICD codes. Codes >= 3 characters no longer generate an exception as long as they are defined in the mapping table. Codes that aren't defined are still mapped to a residual although the process is different. See Disease Code Mapping section for more details.
- History periods can now be 1 day long. PC-LTAS required all periods to be 2 or more days long.

Disease Code Mapping

Similar to PC-LTAS, LTAS.NET translates the cause of death from the outcome file from ICD code to NIOSH (or user-defined) minor. While PC-LTAS uses ICD code ranges, LTAS.NET maps each ICD code individually in order to improve the precision and accuracy of the mapping. As a result, you may notice the following differences when comparing the mapping of your data using PC-LTAS and LTAS.NET:

- In LTAS.NET, ICD codes require a decimal if a fourth digit is present.
- LTAS.NET may detect invalid ICD codes that PC-LTAS does not.
- LTAS.NET may map a specific ICD code to a different and more appropriate minor than PC-LTAS does.

Stratify Processing

Compatibility with PC-LTAS Stratify Step

The following is a list of differences between the LTAS.NET Stratify process and the PC-LTAS Stratify step:

- Stratification can be performed on additional confounders (not just age, race, sex and calendar time, as in PC-LTAS)
- Single day history gaps are now filled with unexposed person time. PC-LTAS only fills history gaps of 2 or more days.
- A zero value can be used for the first stratification category. PC-LTAS required all values to be greater than zero.
- Stratification now supports Time Since Last Exposure (TSLE).
- Units may be specified for exposure levels.
- There are several difference in how date calculations are performed for calculating person time and handling exposure lagging. See Appendix A - LTAS Date Handling for more details.

Analyze Processing

Compatibility with PC-LTAS Analyze Step

The following is a list of differences between the LTAS.NET Analyze process and the PC-LTAS Analyze step:

- LTAS.NET does not offer a chi-square test as did PC-LTAS
- There are more report options and more report flexibility in LTAS.NET
- SRR results are now presented in a separate report than the SMRs rather than reported as an additional row of data in the SMR table as in PC-LTAS
- LTAS.NET will compute only 2-sided p values

Appendix E - Rate Files

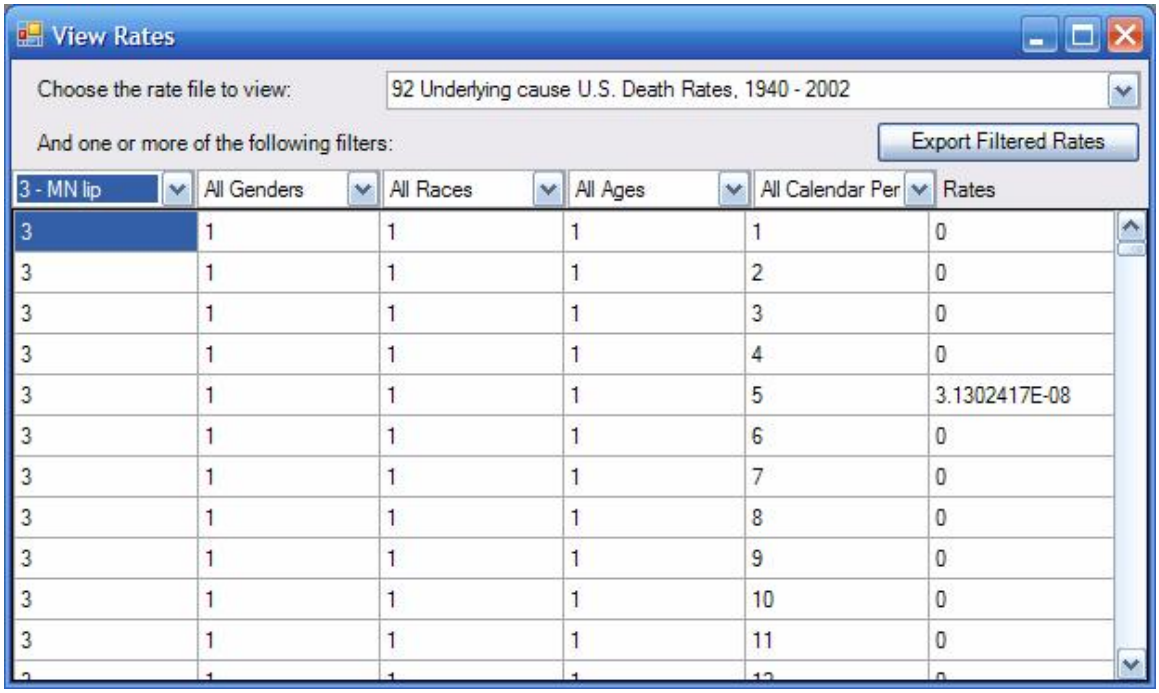
Viewing & Exporting Rate Data

To view or export rate data, access the ‘View Rates’ option under the ‘Manage Rates’ menu and follow these steps to select the subset of rate data that you want to view:

1. Select the rate file you wish to view

2. Select at least one of the filters

Once these steps are complete, the rate data will be displayed as shown below.



3 - MN lip	All Genders	All Races	All Ages	All Calendar Per	Rates
3	1	1	1	1	0
3	1	1	1	2	0
3	1	1	1	3	0
3	1	1	1	4	0
3	1	1	1	5	3.1302417E-08
3	1	1	1	6	0
3	1	1	1	7	0
3	1	1	1	8	0
3	1	1	1	9	0
3	1	1	1	10	0
3	1	1	1	11	0

To export the rate data press the 'Export Filtered Rates' button. This will allow you to save the selected subset of rate data to a comma-delimited file.

Building Rates

Building rates in LTAS.NET involves these two types of files:

- Cause Map – a mapping of disease codes to cause categories.
- Rate Set – a set of rates for each gender/race/age/calendar/cause category combination.

While building rates in PC-LTAS also involved two types of files, the contents and formats differ. Any existing PC-LTAS structure and raw data rate files must be converted to the new format before they can be used by LTAS.NET.

Cause Maps

A Cause Map defines the cause categories and maps the disease codes from your study data to those cause categories. The Cause Maps provided with LTAS.NET map ICD codes to a set of NIOSH cause categories. Because ICD codes were revised over time, mapping them to a smaller set of NIOSH cause categories simplifies the structure of the Rate Set and the process of performing analysis.

Cause categories are defined by a hierarchy of majors and minors. A minor is a single cause category that one or more related disease codes map to. A major is a group of one or more related minors. A special cancers grouping is also used to group cancer-related majors. The grouping of minors and majors determines the options that will be available for performing analysis on your study.

While similar to the PC-LTAS structure file, the Cause Map description file contains only cause information such as the list of majors and minor. The Cause Map does not contain any information specific to rates, such as the gender, race, age and calendar categories. This allows the same Cause Map to be used by multiple Rate Sets.

LTAS.NET currently includes two different Cause Maps representing the NIOSH 92 and NIOSH 119 cause categories.

How to Import a Cause Map

To update an existing Cause Map or add a new Cause Map to LTAS.NET you will need to import a Cause Map Description file using the LTAS.NET application. The steps for performing the import are described below.

1. Start LTAS.NET and access the 'Import Cause Map' option under the 'Manage Rates' menu.

2. From the 'Import Cause Map' dialog, click on the '…' button to select the Cause Map Description file and press the 'OK' button to continue.

3. If the Cause Map you are importing already exists, you will be prompted to confirm that you wish to replace the existing Cause Map with the new data. You must click on the 'Yes' button in order to replace the existing Cause Map. Please note that you should not replace NIOSH cause maps with your own maps. Instead, be sure to use a different name for the map.

4. Once the Cause Map Description file has been selected and confirmed a progress dialog will be displayed as the Cause Map data is imported. It should not take any longer than a few minutes to load the data depending on the speed of your computer.

5. Once the Cause Map has finished importing, an 'Import complete' message will be displayed. If you were adding a new Cause Map, you can now import Rate Sets that use this Cause Map. If you were updating an existing Cause Map, you will need to repeat the Import and Stratify steps on any of your studies that use Rate Sets that are based on the Cause Map in order to reflect any changes to the disease code mappings.

How to Create a Cause Map Description File

The Cause Map Description file is an XML formatted file. XML files use tags to label and delimit data within a file. Below is an example of data formatted as XML:

```
<Name>Niosh92</Name>
```

This example can be broken down into three parts:

1. <Name> is the start tag identifying an element of data called "Name".
2. Niosh92 is the value of the Name element.
3. </Name> is the end tag for the Name element.

XML offers significant advantages over fixed and delimited data formats and has been chosen as the format for the Cause Map Description file because it is well suited for representing structured data.

If you do not have an XML editor, any text editor can be used to create an XML formatted file. However, you will need to be careful to ensure that the XML is well-formed. Well-formed XML requires that:

- Every start tag has an end tag

- Any less than characters (<) that occur within a data value are replaced with <
- Any ampersand characters (&) that occur within a data value are replaced with &

The top-level XML element for the Cause Map Description File is CauseMapSet. This element does not contain any data value, but does contain the following subelements:

- Name – a short name that uniquely identifies the Cause Map up to 50 characters in length
- Description – a description of the Cause Map up to 80 characters in length
- IsIcd – a Boolean value of true or false that indicates if the Cause Map maps ICD codes
- Causes – the major/minor grouping hierarchy
- CauseMapData – the mapping data

The grouping hierarchy defined by the Causes subelement contains a number of different types of elements including:

- Major – defines a grouping of minors
- Cancers – defines a grouping of majors that are Cancer related
- Id – the major or minor code. Codes must be numbers that start at 1
- Description – a description of the grouping or specific minor up to 75 characters in length
- Label – a short description of the grouping or specific minor up to 35 characters in length

The CauseMapData element actually contains multiple lines of data in a comma-delimited format in order to reduce the size and complexity of the XML file. The format of each line of data contains the following three values separated by commas:

1. ICD Revision
2. Disease Code
3. Cause Category

Unlike PC-LTAS, LTAS.NET maps specific disease codes, not a range. This means that you need to have a line of data for every possible Disease Code in every ICD Revision. While quite exhaustive, this mapping approach is better at detecting invalid disease codes.

Guidelines for Cause Labels and Descriptions

Before creating the Label and Description elements for each Cause, it is important to understand how this information is used by LTAS for creating Analysis reports.

Labels are used for column and row headings when Cause is used as either the horizontal or vertical axis. Because the space available for a column or row heading is small, you will want to keep the cause labels as short as possible to allow them to fit properly in the report. Avoiding all caps will also conserve space as lower case letters tend to be narrower than upper case letters.

Descriptions on the other hand are used for page headings when Cause is not used as either the horizontal or vertical axis. Since you have nearly the entire page width available for the page heading, descriptions can be much longer than the labels.

Sample Cause Map Description File

```
<CauseMapSet>
    <Name>Niosh92</Name>
    <Description>92 Cause Standard NIOSH</Description>
    <IsIcd>true</IsIcd>
    <Residual>92</Residual>
    <Causes>
        <Description>ALL CAUSES (92 Base)</Description>
        <Label>All Causes (92 Base)</Label>
        <Major>
            <Id>1</Id>
            <Description>TUBERCULOSIS</Description>
            <Label>Tuberculosis</Label>
            <Minor>
                <Id>1</Id>
                <Description>RESPIRATORY TUBERCULOSIS</Description>
                <Label>Respiratory tuberculosis</Label>
            </Minor>
                :
            Additional Minors
                :
        </Major>
        <Cancers>
            <Description>ALL CANCERS</Description>
            <Label>All Cancers</Label>
            <Major>
                <Id>2</Id>
                <Description>MN OF BUCCAL CAVITY AND PHARYNX</Description>
                <Label>MN buccal &amp;amp; pharynx</Label>
                <Minor>
                    <Id>3</Id>
                    <Description>MN OF LIP</Description>
                    <Label>MN lip</Label>
                </Minor>
                    :
                Additional Minors
                    :
            </Major>
                :
            Additional Majors
                :
        </Cancers>
            :
        Additional Majors
            :
    </Causes>
    <CauseMapData>
05,001,092
05,001.A,092
05,001.B,092
:
Additional comma-delimited lines of cause mapping data
:
    </CauseMapData>
</CauseMapSet>
```

Rate Sets

A Rate Set defines incident rates for each gender, race, age, calendar period and cause category combination. The Rate Set also defines the gender, race, age and calendar period categories that are included in the rates and identifies which Cause Map that the rate set is based on. The Cause Map defines the actual cause categories that are included in the rates. This allows multiple Rate Sets to use the same Cause Map.

While similar to the PC-LTAS rate data file, the Rate Set description file contains all of the information needed to define the rate set, not just the raw rate data. This includes information that used to be part of the PC-LTAS structure file, such as gender, race, age and calendar categories.

LTAS.NET currently includes four different Rate Sets based on the NIOSH 92 and NIOSH 119 Cause Maps.

How to Import a Rate Set

To update an existing Rate Set or add a new Rate Set to LTAS.NET you will need to import a Rate Set Description file using the LTAS.NET application. If your Rate Set is based on a new Cause Map, you must first import the new Cause Map. The steps for performing the Rate Set import are described below.

1. Start LTAS.NET and access the 'Import Rate Set' option under the 'Manage Rates' menu.

2. From the 'Import Rate Set' dialog, click on the '…' button to select the Rate Set Description file and press the 'OK' button to continue.

3. If the Rate Set you are importing already exists, you will be prompted to confirm that you wish to replace the existing Rate Set with the new data. You must click on the 'Yes' button in order to replace the existing Rate Set. Please note that you should not replace NIOSH Rate Sets with your own Rate Sets. Instead, be sure to use a different name for the map.

4. Once the Rate Set Description file has been selected and confirmed a progress dialog will be displayed as the Rate Set data is imported. It should not take any longer than a few minutes to load the data depending on the speed of your computer.

5. Once the Rate Set has finished importing, an 'Import complete' message will be displayed. If you were adding a new Rate Set, this rate set will now be available for selection for any of your studies. If you were updating an existing Rate Set, you may need to repeat the Import and Stratify steps on any of your studies that use this Rate Set in order to reflect any changes to gender, race, age or calendar period categories.

How to Create a Rate Set Description File

The Rate Set Description file is an XML formatted file. XML files use tags to label and delimit data within a file. Below is an example of data formatted as XML:

<Name>Us92ur</Name>

This example can be broken down into three parts:

1. <Name> is the start tag identifying an element of data called "Name".
2. Us92ur is the value of the Name element.
3. </Name> is the end tag for the Name element.

XML offers significant advantages over fixed and delimited data formats and has been chosen for as the format for the Rate Set Description file because it is well suited for representing structured data.

If you do not have an XML editor, any text editor can be used to create an XML formatted file. However, you will need to be careful to ensure that the XML is well-formed. Well-formed XML requires that:

- Every start tag has an end tag

- Any less than characters (<) that occur within a data value are replaced with <
- Any ampersand characters (&) that occur within a data value are replaced with &

The top-level XML element for the Cause Map Description File is RateSet. This element does not contain any data value, but does contain the following subelements:

- Name – a short name that uniquely identifies the Rate Set up to 50 characters in length
- Description – a description of the Rate Set up to 80 characters in length
- Multicause – a Boolean value of true or false that indicates if the rates are multicausal
- Proportions – a Boolean value of true or false that indicates if the rates are proportionate
- Diagnosis – a Boolean value of true or false that indicates if the rates contain diagnosis
- CauseMapSet – name of the Cause Map used by this Rate Set
- Genders – set of Gender categories used by this Rate Set
- Races – set of Race categories used by this Rate Set
- Ages – set of Age categories used by this Rate Set
- CalendarPeriods – set of Calendar Period categories used by this Rate Set
- RateData – the raw rate data

The Genders, Races, Ages and CalendarPeriods elements contain multiple Gender, Race, Age and CalendarPeriod elements respectively for each category being defined. Each of these elements contains the following subelements:

- Id – the category id specified as a number starting at 1
- Description – a short description of the category

The RateData element actually contains multiple lines of data in a comma-delimited format in order to reduce the size and complexity of the XML file. The format of each line of data contains the following values separated by commas:

1. Gender category id
2. Race category id
3. Age category id
4. Calendar period category id
5. Minor id
6. Rate value

The number of lines should equal the number of Gender/Race/Age/Calendar period/Minor combinations.

Sample Rate Set Description File

```
<RateSet>
    <Name>Us92ur</Name>
    <Description> 92 Underlying-cause U.S. Death Rates, 1940 - 1997</Description>
    <Multicause>false</Multicause>
    <Proportions>false</Proportions>
    <Diagnosis>false</Diagnosis>
    <CauseMapSet>Niosh92</CauseMapSet>
    <Genders>
        <Gender>
            <Id>1</Id>
            <Description>Male</Description>
        </Gender>
            :
        Additional gender categories
            :
    </Genders>
    <Races>
        <Race>
            <Id>1</Id>
            <Description>White</Description>
        </Race>
            :
        Additional race categories
            :
    </Races>
    <Ages>
        <Age>
            <Id>1</Id>
            <Description>15 - &amp;amp; 20</Description>
            <EntryPointYears>15</EntryPointYears>
        </Age>
            :
        Additional age categories
            :
    </Ages>
    <CalendarPeriods>
        <CalendarPeriod>
            <Id>1</Id>
            <Description>1940 - 1944</Description>
            <BeginDate>1940-01-01</BeginDate>
            <EndDate>1944-12-31</EndDate>
        </CalendarPeriod>
            :
        Additional calendar period categories
            :
    </CalendarPeriods>
    <RateData>
1,1,01,01,1,0.000085116
1,1,01,01,2,0.000013435
1,1,01,01,3,0.000000000
:
Additional comma-delimited rate data lines for each gender, race, age, calendar
   period and minor combination.
:
    </RateData>
</RateSet>
```

Chapter 10. Glossary

B

BN: Benign neoplasm.

C

Cause Map file: A required rate file that defines the cause categories and maps the disease codes from the user's Outcome file to those cause categories.

CDC: See Centers for Disease Control and Prevention.

Centers for Disease Control and Prevention: A Federal agency organized under the Public Health Service. Parent agency of NIOSH.

COD: Cause of death. Represented in LTAS.NET as either underlying cause (UCOD) or multiple cause(s) (MCOD).

Cohort: A group of people under study for the effects of a common exposure.

Confidence level: Probability that an interval calculated by a statistical procedure will contain the true value of a parameter.

Cumulative level of exposure: A time-dependent measure of accumulated exposure up to a specific point in time, typically in "exposure-time" units such as ppm-years. Used in analyses of exposure-response trends.

Cut Point: The cut points are the values that mark the borders of the strata.

D

Date Last Observed: Most recent date on which subject was known to be alive.

Death rate: The total number of deaths from a certain cause during a certain time period, divided by the population's person-time at risk during that period. Usually standardized by age, race, sex, and calendar time. Used in Standardized Mortality Ratio (SMR) and Standardized Rate Ratio (SRR) runs.

Direct standardization: Standardization of rates for age, race, sex, and calendar time in which the rates of different cumulative level of exposure groups are weighted by the person-time of the entire exposed cohort. Directly standardized rates are used for internal comparisons in which one (usually the low) cumulative level of exposure (or low duration) group is the referent,and SRR (standardized rate ratios) are calculated.

DLO: See Date last observed.

DOB: Date of birth.

DOD: Date of death.

Duration: Duration of exposure, which may be used when no data are available on exposure level or intensity. In SMR runs using person-time and duration, an "exposure level" of 1 is assigned to each day of exposure for each subject. Duration accumulates over time; each person-day for each subject is assigned the cumulative duration, and then that person-day is allocated to the duration category. If employment duration is also desired (and is distinguishable from exposure), then the user should include a variable of value "1" associated with periods of active employment.

E

Exp: Exposure

Export: To create processed data from a software system in a form amenable for storage, display,or processing by other software. LTAS.NET permits the export of the following data: Exceptions and Observed deaths in the Import step; in the Stratify step, the Cohort (person-time and observed deaths by age, race, sex, calendar time and any fixed or time-dependent variables which were stratified upon), Personal Cumulative Dose for each person, and the rejected persons.

Exposure: see cumulative level of exposure, exposure level, or exposure file.

Exposure Begin Date: Earliest date on which a worker was exposed.

Exposure level: Also known as exposure intensity or exposure concentration. For example, parts permillion (ppm) or milligrams per cubic meter (mg/m3). This can change over time. Itis specified in the exposure file for combinations of time period, plant, department,and operation (area exposure) or for combinations of time period and worker (personalexposure). A level of "0" means no exposure.

H

History file: A required user input file listing dates over which each time-dependent exposure variable is unchanging and the daily average value of each time-dependent exposure variable during that period. Linked to the Person and Outcome files via the Person ID.

Hypothesis test: A statistical test for the null hypothesis of no difference between an exposed group and a referent group. Often a test of whether a rate ratio (SMR or SRR) is statistically different from the null value of 1.0 by chance alone, assuming that the true (unknown) rate ratio were 1.0.

I

ICD: International Classification of Diseases; a system of disease classification used by the World Health Organization (WHO). It is a list of codes representing diseases.There are several revisions of this list, the latest of which is the tenth revision, which took effect in 1999.

ICD File: Also called a cause map file. This file is used by LTAS to convert ICD numbers intomajors and minor groupings which are listed in the rate file.This conversion is necessary because the rate data file is listed by minors. See Appendix E.

ICD time period: the time during which a given ICD revision is in effect.

Indirect standardization: Standardization for age, sex, race, and calendar time in which the stratified rates of an external referent population are weighted by the stratified person-time of the exposed group. Used in SMR analyses.

J

Job: Specific work assignment of subject. A subject may have multiple jobs or work assignments or work histories over a working lifetime.

L

Lag: The time elapsed between exposure and the expected effect of that exposure. Used to discount recent exposures, especially in cancer mortality studies in which one assumes thata minimum induction period latency is required prior to observing an effect.

M

Major: short name for Major cause-of-death category.

Major cause-of-death category (major): A related group of one or more minor cause-of-death categories.

MCOD: See Multiple causes of death.

Minor: short name for Minor cause-of-death category.

Minor cause-of-death category (minor): A related group of one or more ICD codes.

MN: Malignant neoplasm.

Mortality Files: The Cause Map file and the Rate Set file for mortality rates. Note that cancer incidence rate files are available as well.

Multiple cause mortality: Considers all causes on the death certificate, not only underlying cause.

Multiple causes of death: All causes of death listed on the death certificate as coded using the NCHS' Record Axis system of coding. This term replaces the use of "contributing causes of death" in previous versions of LTAS. In LTAS.NET, MCOD and UCOD are handled separately. Because of coding rules associated with ICD, the UCOD may or may not be one of the MCOD. Therefore, multiple-cause analysis in LTAS.NET includes only outcomes with "underlying=False".

N

NIOSH: National Institute for Occupational Safety and Health.

NUB: Neoplasm of uncertain or unspecified behavior (i.e., benign vs. malignant).

O

Outcome file: A required user input file that contains information about the outcomes (incident cases or deaths from diseases) for each subject who experienced an event. Linked to the History and Person files via the Person ID. Must also contain vital status, ICD code corresponding to the event, date of occurrence, and T/F indicators of whether the event was terminal (i.e., a mortality event) and underlying (i.e., the underlying cause or a multiple cause).

P

PCMR: See Proportionate Cancer Mortality Ratio.

Person-time at risk: Amount of time at which each subject is at risk of dying or having incident disease occur from the exposure of interest. Often begins at time of first exposure or enrollment into the cohort.; continues until end of follow-up. Summarized across subjects and standardized by potential confounders (age, race, sex, and calendartime). May also be stratified by cumulative level of exposure (or duration), time since first exposure, or any other time-dependent exposure for analyses of trends. Used in SMR and SRR studies. (Also called person years, person years at risk, etc.)

Person file: A required user input file describing the workers in a cohort; includes demographic information and any fixed user-defined covariates. Linked to History and Outcome files via the Person ID.

Person years Begin Date: Date from which a subject's time at risk is initially calculated. In LTAS.NET, the user has great flexibility in choosing from among a list of candidate dates. The latest date selected is always employed. This is often the date of first exposure, or the person years begin date field in the Person file record for the subject. This depends on the person year begin date option (see Chapter 6, Import Process). Sometimes person years begin at the study begin date if the date of first exposure is before the study begin date. For additional information on the person years begin date field, see Appendix C, "Person Import Files Requirements".

Person years End Date: Ending date of the accumulation of person years. This is the date of death for dead people. For survivors, it is usually the study end date. The DLO (Date Last Observed) option ends the accumulation of person years at the date in the DLO field in the Person file record of the subject. For additional information on the date last observed field, see Appendix C, "Person Import Files Requirements".

PMR: See Proportionate Mortality Ratio.

Proportionate Cancer Mortality Ratio (PCMR): ratio of exposed cancer deaths per unexposed cancer deaths. No person time is accumulated.

Proportionate Mortality Ratio (PMR): ratio of exposed deaths per unexposed deaths. No person time is accumulated.

PY: person year.

PY Begin Date: See Person years Begin Date.

PY End: See Person years end date.

R

Rate Set File: A file containing reference population disease incidence or mortality rates or proportions for each gender, race, age, calendar period combination for each of the NIOSH minors. (The rate data file for a PMR run contains proportions instead of rates.) This file is used to calculate expected numbers of deaths.

S

SMR: See Standardized Mortality Ratio.

SRR: See Standardized Rate Ratio.

SSN: Social Security number.

Standardization: Adjustment or weighting of rates after stratification by age, race, sex, and calendar time to avoid confounding by these variables in a comparison of summary rates between an exposed cohort and a referent population. Standardization can be direct or indirect. Note that standardization will occur if and only if the variable is represented in the external rate file. Rate files supplied with LTAS.NET contain information on age, sex, race (white/other) and calendar period.

Standardized Mortality Ratio (SMR): Ratio comparing the summary mortality rate for an exposed cohort to the summary mortality rate for an external population (such as the U.S. population in the U.S. Rates file), after stratification of rates by age, race, sex, and calendar time, and weighting of stratified rates by indirect standardization.

Standardized Rate Ratio (SRR): Ratio comparing a summary mortality rate for a group (usually, more highly exposed or longer duration group) to a summary mortality rate for a different exposed (usually, low exposure or low duration) internal referent group, after direct standardization of rates by age, race, sex,and calendar time.

Stratification: Categorization of rates or proportions by potential confounders (age, race, sex,calendar time, or any number of user-defined fixed or time-dependent covariates). Rates (deaths per person-time) are stratified for SMR and SRR runs. Proportions (deaths from specific cause divided by all deaths) are stratified for PMR runs.

Stratum: A single category of age, race, sex, and calendar time, for example, white males aged 50-54 in 1960-1964, which also may be classified by user-defined fixed or time-dependent covariate strata. Stratification by cumulative duration/ exposure, by time since first exposure (TSFE), time since last exposure (TSLE), or any combination of user-defined covariates, is also provided by LTAS.NET, for example 5-10 years duration in TSFE category 20-25 years, or 100-500 ppm in TSFE category 10-20 years.

Study Begin Date: First date considered for all workers and the starting date for the reports. Any exposure before this date is not considered in the study. Workers having final employment dates occurring before the study begin date are excluded from the analysis.

Study End Date: Last date considered for all workers and the ending date for the reports. Any workexposure after this date is not considered in the study. Workers having beginningemployment dates occurring after the study end date are excluded from the analysis.

Study Files: The person, outcome and history files. Each subject has his or her own record in each file. Unlike PC-LTAS, for LTAS.NET the history file must specify exposures for individual subjects. These are ASCII text files.

T

Time period: in the Cause Map file this is the time during which a mapping from ICD codes to minor cause of death is in effect. This usually corresponds to "ICD time period."

Time Since First Exposure/Employment: Calendar time elapsed between the date on which the subject began exposure and another specified date (such as date of death or end of study). A time-dependent measure that accumulates as the subject moves through time. Each person-day for each subject is assigned the time-since-first-exposure amount, and each person-day is allocated to the user-specified category of time-since-first-exposure. Used in analyses of trends with time-since-first-exposure. Sometimes called "latency" or "potential latency."

Time Since Last Exposure/Employment: Calendar time elapsed between the date on which the subject last experienced exposure and another specified date (such as date of death or end of study). A time-dependent measure that accumulates as the subject moves through time. Each person-day for each subject is assigned the time-since-last-exposure amount, and each person-day is allocated to the user-specified category of time-since-last exposure. Used in analyses of trends with time-since-last-exposure. TSLE is defined as zero before the date of last exposure.

TSFE: See Time Since First Exposure/Employment.

TSLE: See Time Since Last Exposure/Employment

U

Underlying cause of death: The cause coded by a nosologist as the underlying cause.

UNN: Neoplasm of unspecified nature (i.e., primary vs. secondary)

V

Vital Status: A variable indicating whether a subject is alive or dead. Vital status may also be listed as unknown.

VS: See Vital status.

Chapter 11. References

1. Waxweiler R, Beaumont J, Henry J, Brown D, Robinson C, Ness G, Wagoner J, Lemen R. A modified life-table analysis system for cohort studies. **J Occup Med.** [1983]; **25**:115-124.

2. ET Lee. **Statistical Methods for Survival Data Analysis**. [1980], Belmont, CA, Lifetime Learning Publications, pp. 88-96.

3. Steenland K, Beaumont J, Spaeth S, Brown D, Okun A, Jurcenko L, Ryan B, Phillips S, Roscoe R, Stayner L, Morris J. New developments in the Life Table Analysis System of the National Institute for Occupational Safety and Health. **J Occup Med** 32:1091-1098, [1990].

4. Checkoway H, Pearce N, and Crawford-Brown DJ. **Research Methods in Occupational Epidemiology**. [1989], New York, Oxford University Press, pp. 116-136.

5. Rothman KJ and Boice JD Jr. [1979] **Epidemiologic Analysis With a Programmable Calculator.** Washington, DC, US Government Printing Office.

6. Rothman KJ [1986] **Modern Epidemiology**. Boston, Little, Brown, and Company.

7. DG Kleinbaum, LL Kupper, and H Morgenstern. **Epidemiologic Research: Principles and Quantitative Methods**. 1982, Belmont, CA, Lifetime Learning Publications.

8. Miettenin O and Wang J, An alternative to the PMR, **Am J Epidemiol** , 114: 144-148, [1981]

9. Vandenbroucke P, A shortcut method for calculating the 95% confidence interval of the standardized mortality ratio, **Am J Epidemiol 115:** 303-304, [1982].

10. Steenland K, Nowlin S, Ryan B, and Adams S. Use of Multiple-Cause Mortality Data in Epidemiologic Analyses. **Am J Epidemiol** [1992] **136**:855-862.

11. Israel RA, Rosenberg HM, and Curtin LR. Analytical potential for multiple cause-of-death data. **Am J Epidemiol** [1986] **124**:161-179.

12. Checkoway H, Heyer NJ, Seixas NS, Welp EAE, Demers PA, Hughes JM, Weill H. Dose-response associations of silica with nonmalignant respiratory disease and lung cancer mortality in the diatomaceous earth industry. **Am J Epidemiol** 145:680-688, [1997].

13. Checkoway H, Pearce N, Kriebel D. **Research Methods in Occupational Epidemiology, Second Ed.** New York: Oxford University Press. [2004], 392 pp.

14. Preston DL, Lubin JH, Pierce DA, McConney ME. **Epicure Users Guide**. Hirosoft International Corp. Seattle, WA. 330 pp, [1993].

15. Breslow NE and Day NE [1987], **Statistical Methods in Cancer Research: Volume II The Design and Analysis of Cohort Studies.** IARC Scientific Publication No. 82.

16. Flanders WD [1984], **Approximate variance formulas for standardized rate ratios.** J Chron Dis 37(6): 449-453.

17. Kahn HA and Sempos CT [1989], **Statistical Methods in Epidemiology**. New York: Oxford University Press.

18. Neter J, Kutner MH, Nachtsheim CJ and Wasserman W [1986], **Applied Linear Statistical Models**, Irwin Book Team.

19. Rothman KJ and Greenland S [1988] **Modern Epidemiology, Second edition.** Lippincott Raven Publishers.

Chapter 12. Index

U

V