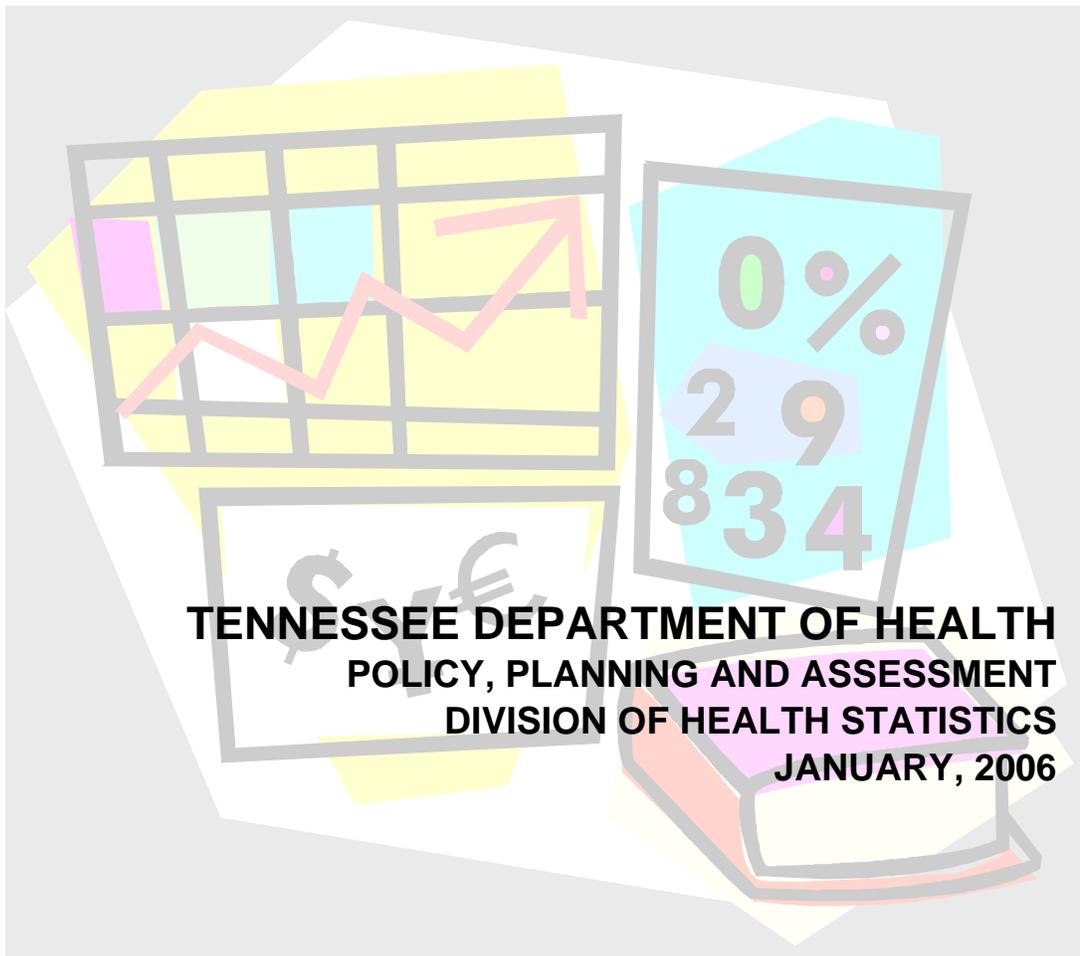
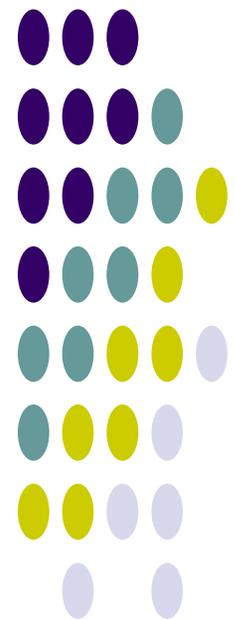


GUIDELINES FOR PREPARING STATISTICAL TABLES AND REPORTS



**TENNESSEE DEPARTMENT OF HEALTH
POLICY, PLANNING AND ASSESSMENT
DIVISION OF HEALTH STATISTICS
JANUARY, 2006**



GUIDELINES FOR PREPARING STATISTICAL TABLES AND REPORTS

Tennessee Department of Health
Authorization No. 343783, No. of copies 100.
This public document was promulgated at a cost
of \$0.85 per copy. 03-06



Table Of Contents

	Page
Introduction	
Analysis and Interpretation of Vital Statistics Data.....	1
Preparing Statistical Reports and Publications	2
Definitions of Terms, Rates, and Ratios	4
Guidelines to be Followed in Statistical Work	6
Guidelines to Observe When Writing the Narrative of a Statistical Report.....	7
Guidelines for Preparing Graphs	8
Guidelines for Preparing Tables	8
Guidelines for Preparing Maps	8
Principles for Checking and Finalizing a Statistical Report.....	9
Sources of Inconsistencies in Health Statistics Work.....	10
Policy for Release of Aggregate Data.....	12
How to be a Vital Statistician	13
Department of Health Guidelines for Age-Adjustment	18
Appendix A	
Comparable Code Numbers for the Ninth and Tenth Revisions	
International Classification of Diseases.....	21

INTRODUCTION

Analysis and Interpretation of Vital Statistics Data

The analysis and interpretation of data from vital records for the State of Tennessee vital statistics data has been going on since at least 1917. It has been, and still is, predicated upon standards and methodology formulated at the national level by the agency that is now called the National Center for Health Statistics (NCHS), of the Centers for Disease Control and Prevention (CDC). The methods and procedures used at the both the national and state levels have evolved over a long time period with considerable thought and attention given to the various data elements selected, the manner in which they are collected, the vital statistics measurement so derived, and the proper analysis, interpretation, and presentation. There is a wealth of history, tradition and culture associated with these processes. This is often misunderstood, especially by researchers using these data for the first time. The purpose of this document is to provide some guidance to those working with health statistics data to both understand and apply common standards and usages.

It should be appropriate to point out that health-related data is formulated from various sources and settings. There are biostatistical data created in laboratory and scientific environments. There are epidemiological data evolving from the surveillance of disease and adverse health conditions. There are also health statistics data that are gathered as a byproduct of the reporting of vital events. While there is some overlap in these types of health data and often these terms are used interchangeably, still they should generally be recognized as separate fields or subfields of a more global concept of health related statistical data. Each category is a unique area of study in its own right. And this is most certainly true in vital statistics. However, this situation is not well recognized outside of the vital statistics environment.

Vital statistics, through guidance from NCHS and all of the other states' vital statistics systems, are very well defined as to how to collect, assemble, aggregate, analyze and present data, and there are 'right' and 'wrong' ways of doing this. There are certain well-defined measures used; defined methodologies to use; and structured ways of presenting these data. While NCHS provides a wide variety of documentation and resource material on the subject and while many senior staff in the Division of Health Statistics have been formally trained in this area and are familiar with these standards and procedures, these resources are often overlooked by persons attempting to analyze vital statistics data.

It would take a long time to fully detail all of the knowledge base for analyzing and presenting vital statistics data. It cannot be done within the confines of this document. The attempt here will be simple and threefold: first, to illustrate some common sources of inconsistent measures and analysis of supposedly similar data; secondly, to give a broad general overview of how vital statistics data are analyzed; and thirdly, to provide the reader with some basic samples of resource material available from NCHS.

Preparing Statistical Reports and Publications

Statistical reports and publications are prepared from many different data sources. Joint annual reports of health care facilities, hospital discharge data, communicable disease reports, and vital records data supply some of the information used in the reports prepared by the Division of Health Statistics. The information obtained annually from these data sources is compiled and checked for accuracy and consistency. The compiled data is transferred into a tabular format which is used for analysis and reference. It is then possible to obtain specific and select data sets such as, births by age of mother, cause of death, or county of residence of hospital patients. Other data sources include the United States Bureau of the Census, the National Center for Health Statistics and the Centers for Disease Control.

Data obtained from certificates of “vital events” are generally referenced as vital statistics data and/or statistics. Copies of certificates and reports for births, deaths, induced abortions and fetal deaths occurring to Tennessee residents in other states are forwarded to Tennessee in the interstate transfer of these certificates and reports. Thus, it is possible to compile birth, death, induced abortion and fetal death statistics for Tennessee residents. Resident data is data compiled according to the usual place of residence without regard to the geographic place where the event occurred. Since at the present time there is no interstate transfer of marriage and divorce certificates, resident data for these events is not available.

Recorded data are obtained by classifying certificates and reports according to the place of occurrence of birth, death, fetal death, induced abortion, marriage or divorce without regard to place of residence. For certain purposes, such as knowledge of the numbers of births or deaths occurring in hospitals, recorded data are desirable and are presented in some of the summary tables. Many residents of the state and of nearby states use health facilities in the large cities of Tennessee; consequently, the numbers of recorded births, deaths and induced abortions are large in these cities.

Resident data are compiled according to the usual place of residence of the person to whom the event occurred without regard to the geographic place where the event occurred. Resident data are preferable to recorded data for comparing birth, death and induced abortion rates and ratios for the residents of the counties of Tennessee. Most of the statistics prepared by the Division of Health Statistics are based on resident data.

On May 2, 1989, the marriage law (T.C.A. 36-3-103) was amended to include marriages for which the couple obtained a Tennessee marriage license but had a ceremony performed out-of-state. These marriages are included in statistical analyses just as those with in-state ceremonies.

On January 1, 1989, the Office of Vital Records for the State of Tennessee began using revised certificates of birth and death and report of fetal death which are based upon, and very similar to, the U.S. Standards in content and format. The modifications and additions made to the Tennessee birth certificate and fetal death report were extensive. The previous revision contained a minimum of medical information whereas the 1989 revision includes check box sections entitled Medical Risk Factors for This Pregnancy, Other Risk Factors for This Pregnancy (tobacco and alcohol use, weight gain), Obstetric Procedures, Complications of Labor and/or Delivery, Methods of Delivery, Abnormal Conditions of the Newborn, and Congenital Anomalies of Child. Additional non-medical items call for information on parents' occupation and business/industry and Hispanic origin. Further changes to the birth certificate report starting January 1, 2004, have changed the content and format of the data items on the database even more significantly, especially in the areas of race and ethnicity, medical health information, and smoking. Race of mother and father has been expanded to include 15 racial categories. Maternal smoking has been expanded to include data on smoking before and during pregnancy. Care should be taken in comparing data across these major data reporting system changes. Changes to the Tennessee death certificate were less extensive but still significant. The cause of death section was

expanded by the addition of a fourth line to provide more room for describing the conditions. New items on the 1989 death certificate include education, Hispanic origin, and a check box for autopsy findings available prior to completion of cause of death. The rationale for changes to the two certificates and one report was to provide better and more comprehensive statistical information on public health needs and to measure the effectiveness of initiatives designed to address these problems.

Beginning with the 1991 reported data, natality is presented by race of mother. Before 1991, this data was presented by race of child. The Tennessee Department of Health implemented this change to be consistent with the National Center for Health Statistics (NCHS), which implemented the change in 1989. Race of child had been computed by an algorithm based on race of mother and race of father. Persons of mixed parentage were classified according to the race of the nonwhite parent. Several related events influenced this decision, most importantly the regular revision of the certificate of live birth. These certificates include many new items which include considerable health and demographic information related to the mother. Over the years, the percent of births where parents were not of the same race has been increasing. Also, with the increasing proportion of out-of-wedlock births in which father's race is missing, it was decided by NCHS to implement this change to race of mother which will provide for a more uniform approach.

Changing the basis for tabulating birth data from race of child to race of mother resulted in more white births and fewer births of other races. Therefore, infant mortality rates under the new classification tend to be lower for white infants and higher for infants of other races than they are when computed based on live births tabulated by race of child.

Many factors must be considered for the correct interpretation of vital statistics. Completeness and timeliness of registration is one of these factors. Because of concentrated efforts of the Tennessee Department of Health to record vital events, as well as the public's need for vital records, it is believed that virtually all events in Tennessee are registered. Characteristics peculiar to a certain area, such as proximity to a military base, affect the vital statistics of that area. In addition, knowledge of the racial and age composition of the population is important for proper interpretation.

The population of Tennessee is based on estimates prepared from the 2000 census and revised by the Division of Health Statistics. These revised population figures were based on updated county total estimates released by the Bureau of the Census. The elements of this estimated population are used for calculation of rates per population for the state and for the counties by race, sex and/or age groups. Estimates for total city populations and city populations by race were also obtained from the 2000 census estimates. The population for Nashville was adjusted to reflect Metro-Nashville including all of Davidson County.

It should be noted that the format of the population data changed from that used prior to 1994. The data is now being presented using the racial categories of white and black rather than white and all other races. Time series data continues to use the broader racial categories of white and all other races. This is due to the fact that a detailed breakout of estimated population data for white, black and other races does not exist for the years prior to 1990.

Beginning in 1999, causes-of-death data, both nationally and at the state level, were coded, tabulated and presented according to the *Tenth Revision of the International Classification of Diseases (ICD-10)* classification system. This system was developed collaboratively with the World Health Organization (WHO) and is the required standard to be used within the United States. Such revisions in cause-of-death classifications are periodically made on an international basis to reflect new discoveries and advances in our knowledge of the etiology of diseases; to improve our ability to describe and measure emerging health problems; and for adaptation to changing definitions and nomenclature in the medical field.

Tennessee had previously classified deaths according to the *Ninth Revision of the International Classification of Diseases (ICD-9)* in its *Vital Statistics Bulletins* for the time period of 1979-1998, but in the 1999 *Vital Statistics Bulletin*, Tennessee began implementing *ICD-10*. Unlike previous revisions in the International Classification of Diseases in which the changes were relatively minor, *ICD-10* is radically different from its predecessor, *ICD-9*. Diseases are now coded in an alphanumeric, rather than numeric format. There are new cause-of-death definitions, coding structure, codes and titles, and new ways of categorizing them. Some causes-of-death have been moved from one section to another. Where *ICD-9* classified causes-of-death into about 4,000 categories, *ICD-10* expands them into about 8,000 categories. Where *ICD-9* organized causes-of-death into 72 major groups, *ICD-10* organized them into 113 major groups. Comparisons of data for years prior to 1999 are somewhat difficult to make for many cause-of-death categories and continuity is not always assured.

In addition to the *ICD-10* cause-of-death coding changes being implemented with 1999 data, the National Center for Health Statistics has also changed its age-adjusted rates age distribution standard. The new population standard for the age standardization of death rates is based upon the estimated 2000 United States population age distribution and replaces the existing standard based upon the estimated 1940 population age distribution. The results of this change in population distribution are that the age-adjusted rates presented for 1999 and later have differing scales of values than those age-adjusted rates calculated for previous years. In general, the rates determined by the new age-adjustment methodology standard will be higher than those determined by the replaced standard. The rates are thus not comparable to any rates prior to 1999 and no comparison should be made. The calculation of age-adjusted rates will be discussed in more detail later.

Definitions of Terms, Rates, and Ratios

Age-Adjusted Death Rate - Number of deaths per 100,000 age-adjusted population. (Population adjusted to the age distribution of the estimated United States 2000 standard population.)

Birth Rate - Number of live births per 1,000 population.

Cause-Specific Death Rate - Number of deaths from a specific cause per 100,000 population with the exception of the following: abnormal conditions of the newborn which is per 1,000 live births; maternal death rate which is per 10,000 live births; and congenital malformations, deformations, and chromosomal abnormalities which is per 100,000 live births.

Crude Death Rate - Number of deaths per 1,000 population.

Divorce Rate - Number of persons divorced per 1,000 population.

Fertility Rate - Number of resident live births per 1,000 females 15-44 years of age.

Fetal Death (reportable) - A fetal death of 500 grams or more, or, in the absence of weight, of 22 completed weeks of gestation or more. (Note: Induced abortions are excluded from the fetal death count.)

Fetal Death Ratio - Number of fetal deaths per 1,000 live births.

Hispanic Origin - Hispanic origin refers to persons whose ancestry, national group, lineage, heritage, or country of birth originated from a Spanish speaking country or culture.

Induced Abortion - The intentional termination of pregnancy for a purpose other than to produce a live-born infant, or to remove a dead fetus.

Induced Abortion Ratio - The number of induced abortions per 1,000 live births by race of mother.

Infant Death - A death of a live-born infant under one year of age.

Infant Mortality Rate - Number of infant deaths per 1,000 live births.

Legal Intervention Death – Death inflicted by the police or other law-enforcing agents, including military on duty, in the course of arresting or attempting to arrest lawbreakers, suppressing disturbances, maintaining order, and other legal action. Legal execution is also included.

Live Birth - A birth in which the child shows evidence of life (includes heart action, breathing, or a coordinated movement of a voluntary muscle) after complete birth.

Low Birthweight - A live birth weighing less than 2,500 grams (5 pounds, 8 ounces).

Low Birthweight Rate - Number of live births weighing less than 2,500 grams per 100 live births.

Marriage Rate - Number of persons married per 1,000 population.

Maternal Death - A death due to complications of pregnancy, childbirth, and the puerperium (ICD-10 codes A34, O00-O95, O98-O99).

Maternal Death Rate - Number of maternal deaths per 10,000 live births. Rates by race for 1970-1974 are by race of child per 10,000 live births (race of mother is not available). Rates by race since 1975 are by race of mother per 10,000 live births.

Neonatal Death - A death of a live-born infant under 28 days of age.

Neonatal Mortality Rate - Number of neonatal deaths per 1,000 live births.

Other Termination - Total number of fetal deaths and induced abortions.

Other Termination Ratio - Number of other terminations per 1,000 live births by race of mother.

Postneonatal Death - A death of an infant 28 days or more but under one year of age.

Postneonatal Mortality Rate - Number of postneonatal deaths per 1,000 live births.

Pregnancies - The sum of live births, reportable fetal deaths, and reported induced abortions for women 10-49 years of age.

Pregnancy Rate - Number of pregnancies per 1,000 female population aged 10-49.

Prenatal Care - The amount of prenatal care obtained by the mother is determined by information on the certificate of live birth concerning the month that prenatal care began and the number of prenatal visits.

Race - The terms "white", "black", and "all other races" are used to denote the racial groups. Definitions used for classifying race are those specified by the National Center for Health Statistics which is consistent with the Office of Management and Budget standards.

Recorded Data - Data compiled according to geographic place where the event occurred without regard to residence. (Recorded data for Tennessee include events which took place within the state, irrespective of the residences of the persons to whom the events occurred. Beginning in 1989, recorded data for marriages includes those obtaining a Tennessee license with the ceremony performed out of state.)

Resident Data - Data compiled according to the usual place of residence of the person to whom the event occurred without regard to the geographic place where the event occurred. For births, infant deaths, and fetal deaths, the residence is of the mother. (Resident data for Tennessee include events which occurred to residents of the state irrespective of where the events took place.)

Unmarried - For live births and fetal deaths, mother married to father item on certificate or report checked "no".

Unmarried Birth Rate - Number of births to unmarried mothers per 100 live births.

Very Low Birthweight (VLBW) - A live birth weighing less than 1,500 grams (3 pounds, 4 ounces).

Notes:

Death rates for specific causes are not calculated for white or all other races populations of less than 5,000.

In cases where a cause of death is sex-specific, the associated mortality rate is calculated per 100,000 population of the sex at risk.

Percents greater than or equal to 0.0 but less than 0.05 are indicated by 0.0.

Rates based on births are not calculated when the number of births is less than 100.

A dash (-) used in place of a number of events means there were no events, and percents greater than or equal to 0.0 but less than 0.5 are indicated by 0.0.

Guidelines to be Followed in Statistical Work

1. Completed Units – Usually in statistical work, completed units are used. In completed months, all periods of time less than 1 month are used as 0, periods from exactly 1 month up to 2 months are used as 1. A period of 1 month 29 days is considered as 1 month.
2. Rounding of a number – Use computer rounding.
3. Rates – In calculation of rates, care should be taken to be sure that the numerator and denominator are specific for the same population group. (See Definition of Terms.)
4. Never release data for printing or public website usage that has not been checked.
5. When using “Provisional” data, always label as such in title. Provisional vital statistics for a data year should include delayed certificates.

Guidelines to Observe When Writing the Narrative of a Statistical Report

1. Two numbers should not appear adjacent in a statement except where one number is in parenthesis. To avoid this, revise statement so that the above does not occur.
Example: In 2003, 524 deaths were attributed to home accidents.
Revised: In 2003, home accidents accounted for 524 deaths.

2. In a statistical report, statements which describe the occurrence of certain events should be documented as to time of occurrence. When a comparison is made of date for one time period with that for another, be sure to clarify specific time with data, and be certain that the meaning/comparison is easily understood on first reading.
Example: In 2003, deaths from diabetes mellitus numbered 623, compared with 605 deaths in 2002. This number was 3 percent larger than for the preceding year.
Revised: In 2003, deaths from diabetes mellitus numbered 623, compared with 605 deaths in 2002. The number of deaths in 2003 was 3 percent larger than that for the preceding year.

3. Generally, one writes of a “decrease from” or “increase over.” And “increase from” is not ordinarily used.

4. The term “preceding” is used rather than “previous” to describe that which is immediately before.

5. Numbers are described as “larger/largest” or “smaller/smallest,” whereas, percents and rates are termed “higher/highest” or “lower/lowest.”

6. “Percent” should be spelled out. Do not use symbol (%).

7. Be consistent in use of figures and the number spelled out according to specifications.
 - a. Use words for numbers less than or equal to ten; exceptions are written below.
 - b. Use words when numbers are fractions or have fractions in them. (Example: Two and one-half)
 - c. Use figures when writing rates and percents.
 - d. Use figures for numbers containing decimals.
 - e. Use figures for dates and ages.

Guidelines for Preparing Graphs

1. The simplest graphs are the most effective.
2. Every graph should be self-explanatory.
3. Every graph should have a title and the title should be placed at the top of the graph.
4. Label the scales on both the horizontal and vertical axis. Also, a subject for these scales is needed to identify the unit of value and/or periods of time.
5. Frequency is usually represented on the vertical scale and method of collection on the horizontal scale.
6. When more than one variable is shown on a graph, each should be clearly differentiated by means of legends or keys.
7. No more coordinate lines should be shown than are necessary to guide the eye and each line should be labeled.
8. When shading is used, a key or legend should be placed on the chart.
9. Scale divisions should be clearly indicated as well as the units into which the scale is divided.
10. Indicate the source of the data, usually at the bottom of the chart.

Guidelines for Preparing Tables

1. Every table should have a title.
2. Each table column should be labeled.
3. All tables should be added vertically and horizontally to assure accuracy. Totals and subtotals should be footnoted if they do not add.
4. Indicate the source of the data, usually at the bottom of the table.

Guidelines for Preparing Maps

1. Every map should have a title.
2. A map legend should be placed on the right bottom corner of each map.
3. Map legends should be clearly labeled and shaded to represent data range presented.
4. Indicate the source of the data, usually at the bottom left of the map.

Principles for Checking and Finalizing a Statistical Report

1. If you are not familiar with the type of data contained in the report, review a copy of the preceding year's report (or some similar material) and keep it handy for reference in order to determine if certain figures/statements are reasonable.
2. Before starting to check narrative of report, be sure that all tables/graphs/maps that should accompany narrative have been checked and that they have a title and a number/letter, etc., for identification/reference. The title should clearly and concisely describe the contents of the table/graph/map telling what, where, and when (what data/events/conditions are illustrated, the geographic area where these events occur, and when the events occurred, etc.) in that order. All titles should include the name of the state, i.e., "Tennessee."
3. As you read the narrative, stop and check every numerical value given in the report for accuracy. As you continue to read and check the report, be certain that each statement made is correct and consistent with the remainder of the report (narrative/tables/graphs/maps).
4. If an error is found in the narrative/tables/graphs/maps, draw a line through the error with correction written above (preferable in a different color pencil/pen). Be sure your correction is checked by the person who prepared the report. When the correction has been verified, it should be dotted (or noted in some manner).
5. If a correction is made in narrative/tables/graphs/maps, check the remainder of the report relative to that correction/change.
6. If a statement is made concerning contents of a table/graph/map, review that table/graph/map referenced to make certain that it does contain/show information as stated in the text. Be sure that tables/graphs/maps are correctly numbered and referenced by that number.
7. The person checking/reading a statistical report should be mindful at all times as to whether a value, expression or statement is reasonable (within the realm of possibility).
8. Spell check all work for errors.
9. The person editing a report should identify incorrect sentence structure, punctuation and spelling.
10. When some of the same data/figures appear in two or more tables or places in the report, be sure that they are the same in all instances.
11. All tables should be added vertically and horizontally to assure accuracy.
12. A third reader should review the complete report for logic, comprehension, consistency, etc., but not necessarily check calculations.
13. Any reports sent to an outside agency, such as graphic arts, are proofed again for accuracy.
14. All titles for graphs/tables/maps are compared to the Table of Contents.

Sources of Inconsistencies in Health Statistics Work

1. Different Sources of the Same Data

The same data coming from different sources may provide for slight disagreements in the value of number of events indicated. A good example of this is the comparison of Tennessee vital statistics data with data on Tennessee from a national source such as NCHS. The values may be close, but not exact. This is due, no doubt, to the fact that there may be resident events to Tennessee that did not get counted by Tennessee before the particular data file was officially closed. While Tennessee is a part of the interstate exchange agreement process, this process is not one hundred percent complete. Tennessee receives notification of most, but certainly not all such events. NCHS on the other hand receives data from all states and territories and is in a much better position to discern same. For example, NCHS typically counts four to five more infant deaths than does Tennessee. Academic researchers have to have approval to use vital statistics data from each state where the resident event occurred; not all states are willing to give this approval so that vital statistics analysis from academic sources may count some values less than Tennessee published values.

2. Provisional or Final Data

Vital statistics data are often released on a provisional or incomplete basis before final or complete data are published. Final data are edited for correctness and are obviously more complete than provisional data. Final data usually report higher values for the different statistical measures being reported. Such measures can be more greatly distorted when they are derived from two or more other measures which may have different degrees of completeness.

3. Comparing Resident to Recorded Data

Perhaps the simplest and most common error committed in comparing vital statistics data is comparing resident to recorded data. Resident data is based upon the residence of the person for whom the event is being reported irrespective of where the event occurred. Recorded data is based upon the geographic location of where the event occurred. Typically in Tennessee recorded data for the state as a whole are greater than resident data. Substate data, usually county, varies greatly due to economic and other local and health care referral patterns. Comparing resident and recorded data is grossly inappropriate but is often encountered. When in doubt about which data to use, resident figures are usually the better choice.

4. Using Different Sources of Population Data

Many rates and measures are population-based and incorporate the use of population estimates and projections. The Tennessee Department of Health has a standard set of population estimates that is used for all such calculations. The same rates and measures from other sources such as national data for states from NCHS, generally obtain their population data from the U.S. Census Bureau. Thus there are usually differences in their published rates and Tennessee published rates; these differences, however, are usually minor.

5. Using Different ICD-10 Coding Definitions

Often, researchers on a special project may redefine mortality data that they have obtained in a manner different than the International Classification of Diseases (ICD) coding structure that is used in classifying Tennessee data. The nomenclature and data values used in such studies are technically different than that employed by Health Statistics analysis and published work and is often greatly different. Unless such work is properly documented as to the unique and different procedures used, the source of such difference is not readily apparent.

6. Comparing Data from ICD-9 Coding Schemes to ICD-10 Coding Schemes

Related to the above, especially with regard to cause of death coding, many researchers do not pay careful attention to the details of the International Classification of Diseases (ICD) coding schemes and the methodology associated with it. Currently Tennessee has implemented ICD-10 for its mortality data. Hospital discharge data are still being coded in the ICD-9-CM format. There are major differences in ICD-10 and ICD-9 coding schemes. Comparison across these schemes is still awkward and difficult requiring the use of comparability ratios and other statistical techniques. Furthermore, different researchers do not pay attention to the precision of definition of the codes and hence simply are not comparing the same conditions from one source to another. NCHS has provided data users with detailed tabulated lists for such purposes, but these tabulated lists are often not employed as guidance for research endeavors. Invalid comparisons are quite often presented. (See Appendix A for a brief synopsis of ICD coding changes and comparability ratios.)



STATE OF TENNESSEE
Department of Health
Office of Policy, Planning and Assessment
Cordell Hull Building, 4th Floor
425 Fifth Avenue North
Nashville, TN 37247

**HEALTH DATA RELEASE POLICY
RELEASE OF AGGREGATED DATA**

The Tennessee Department of Health has the responsibility to protect the confidentiality and privacy of the citizens while also adequately presenting information and data concerning conditions that affect public health. With regard to these two responsibilities, the following guidelines should be followed when presenting data for public consumption including publications or public web sites:

Rates

- (1) Any rate may be presented if the population at risk is at least 100 persons. If the population is less than 100, the rate may be presented in certain circumstances as noted in the Exceptions section below. If all Exception criteria are not met, the rate will be suppressed (replaced with an asterisk or similar symbol) and a footnote added to the page indicating “rate not calculated due to population less than 100”.

Numbers

1. Counts for the state, region and county totals can be shown irrespective of the number of events or population size.
2. Counts for state, region and county totals, by sex, can be shown irrespective of the number of events or population size.
3. Counts for state, region and county totals, by race, can be presented if the number of persons in each race group being presented is at least 50.
4. Counts for state, region and county totals, by race and sex, can be presented if the number of persons in each race-sex category presented is at least 50.
5. Counts for the state, region and county totals, by age group can be presented if the number of persons in each age group presented is at least 50.

Exceptions

1. If the Rates criteria are not satisfied, rates may be presented only if both of the following conditions are met:
 - a) Appropriate confidence intervals are presented with all rates.
 - b) Table cell subtracted from the number of total events in the same data file for the same characteristic is at least 10 (“Numerator and Denominator Rule”).
2. If the Numbers criteria are not satisfied, counts may be presented only if the following condition is met:
 - a) Table cell subtracted from the number of total events in the same data file for the same characteristic is at least 10 (“Numerator and Denominator Rule”).

It is understood that a program within the Department of Health may adopt more strict (conservative) guidelines for data presentation. In addition, specific situations related to a public health threat could dictate more detailed data presentation, at the direction of the Commissioner of the Department of Health.

Final October 13, 2003

How To Be A Vital Statistician *

The possible uses of vital statistics are endless. Data users are as diverse as a public health official evaluating a program and using death data, a demographer projecting school enrollments with birth data, and a business person deciding to open a formal wear shop based on marriage data. Many of these users have a thorough knowledge of statistics. Others find the subject confusing and intimidating. For either group, a misunderstanding of what vital statistics mean or an incorrectly used comparison can lead to wrong conclusions. Therefore, this section is included to provide an overview of how to use vital statistics. It is addressed to the person looking at vital events for the first time, but the experienced user may also find a review helpful.

Step 1: Finding The Correct Number

The first step is to determine how many of a particular vital event took place during the year. This involves asking two questions:

Which event or events are appropriate?

DEATHS
INFANT DEATHS
NEONATAL DEATHS
POSTNEONATAL DEATHS
FETAL DEATHS
LOW BIRTHWEIGHT
INFANTS
PREGNANCIES
INDUCED ABORTIONS
MARRIAGES
DIVORCES

This may not be as simple as it sounds. For one thing, examining more than one type of event may be required. For example, a researcher who is concerned with adolescent pregnancies will have to consider abortions and fetal deaths, not simply the number of births.

Deciding which events to use is important since sometimes the choice of one event over another can lead to vastly different conclusions. Consider the researcher who asks, "Have we reduced the likelihood of death for babies?" Selecting neonatal deaths (babies who die before they are 28 days old) for 1992 and 1993 suggests the answer is yes. But selecting post-neonatal deaths (those between 28 days and one year) for the same time period suggests no.

What is the correct population?

Vital events do not usually occur in individuals' homes, which leads to an interesting problem in geography. Vital events are usually reported two ways:

RECORDED

Means that the event (death, birth, marriage) actually took place in the geographic region indicated (either Tennessee or a particular county). The person participating in the event may have lived in another state.

RESIDENT

Means that the person involved in the event lived in the geographic region mentioned, but the event itself may have taken place anywhere in the United States or Canada. In other words, a resident of Davidson County who died in an accident while on vacation in Michigan is included in the Davidson County resident death figure.

* This section is adapted with permission of the Oregon Center for Health Statistics from their *Oregon Vital Statistics, 1983* publication. Examples have been changed to Tennessee examples.

Choosing the appropriate population is important. Recorded birth data is more relevant to a hospital planner who is deciding whether to expand or contract delivery services. Resident birth data should be the choice for projecting school enrollment. When in doubt about which population to use, resident figures are usually the better choice. Most birth and death data are published by residence, which means that comparisons with other states or the United States as a whole are easier. Table headings should indicate the population included in the tables.

Step 2: Making The Number Meaningful With Rates And Ratios

In many instances simply knowing the number of events which occurred is not sufficient. A. Bradford Hill expressed this important statistical concept:

"It is well recognized that white sheep eat more than black sheep--because there are more of them."

In other words, it is the probability of the event which is important. We know more people died in Shelby County than in Lake County, but is the likelihood of dying really different?

In order to answer this question, statisticians calculate rates. This means that the number of events which occurred are compared to the population for which that event could have occurred, and the figure is then standardized to some number (such as 1,000 or 100,000) for convenience.

Here is an example:

$$\text{CRUDE DEATH RATE} = \frac{\text{DEATHS}}{\text{POPULATION}} \times 1,000$$

the number of
a number picked by vital
people who
statisticians to eliminate small
could have died
fractional values

The more specifically a statistician can define the "population at risk" (the bottom part of the rate), the more meaningful the rate is. For example, the crude birth rate, which compares the number of births to the population, is not nearly as informative as the fertility rate. This second rate uses only the number of women of childbearing age (15-44) for comparative purposes and is therefore less distorted by quirks in the population distribution. (Note: The turn of the 20th century notion that only married women between the age of 15 and 44 would be considered at risk of pregnancy has been abandoned for obvious reasons.) The importance of being specific about the population at risk is explained in Step 3.

Unfortunately we do not always have the correct number for the bottom part of the equation. In these situations a substitute is used. For example, how many people are at risk of getting divorced? The number of married people is only available for census years. As a substitute, the crude divorce rate is calculated using the total population regardless of marital status. In other situations, the event is simply compared to another related number to create a ratio. For instance, the abortions to the number of births. This is easier and more accurate than trying to determine the true denominator, which is the total number of pregnant women.

When calculating rates and ratios, great care must be taken to make certain that the same time periods, the same geographical boundaries, and the same population are used in the top and bottom parts of the equation.

Step 3: Comparing Places, Time Periods, And Categories

In many situations, the vital statistician's goal is to compare two or more numbers for different geographical areas, different time periods or different categories such as men versus women. But comparing rates and ratios introduces some new problems which must not be overlooked.

Changes in measurement

The first problem is that the measures being compared are not always the same. Definitions, population estimates, certificates, and coding procedures change from time to time as the need arises. This can create "artificial" differences and can disguise "real" differences. The cause-of-death item provides an excellent example of changes in comparability:

In 1998, 2,237 people died from pneumonia and influenza according to ICD-9.

RATE = 41.2 PER 100,000 POPULATION

In 1999, 1,583 people died from influenza and pneumonia according to ICD-10.

RATE = 28.9 PER 100,000 POPULATION

A new scheme for coding causes of death (ICD-10) was introduced in 1999. This coding change is responsible for fewer deaths being coded to influenza and pneumonia than would otherwise be coded to pneumonia and influenza. Caution should be used when comparing (ICD-10) data with the data from (ICD-9) or earlier causes of death codes.

Taking age and gender into account

In 1923, Mr. G. C. Wipple noted that, "We might find that the death rate of bank presidents was higher than that of newsboys; but this would not be because of different occupations, but because of different ages." We expect older people to die at a higher rate than younger people. We also expect people in their twenties to have more babies than the very young or the very old. Gender and race, as well as age, affect rates drastically.

When comparing two places or two points in time, it is necessary to take these influencing characteristics into account. Here is an example:

The death rate increased between 1950 and 1960 from 8.9 to 9.2 deaths per thousand. But an examination of the death rates for each age group indicates that all decreased. This contradiction is explained by the fact that in 1960 a larger proportion of the population was concentrated in the older age groups where death rates are higher.

	<u>1950</u>	<u>1960</u>
CRUDE DEATH RATE	8.9	9.2
AGE SPECIFIC DEATH RATES		
0 - 4	9.0	7.1
5 - 14	0.6	0.5
15 - 24	1.5	1.2
25 - 44	3.1	2.5
45 - 64	12.6	11.6
65+	62.0	60.3

Before comparing two places or two time periods, always compare the population characteristics first. If discrepancies are noted in any relevant variables, then the rates should be adjusted or standardized in order to make the comparisons free of differences in the structure of the populations.

Small numbers

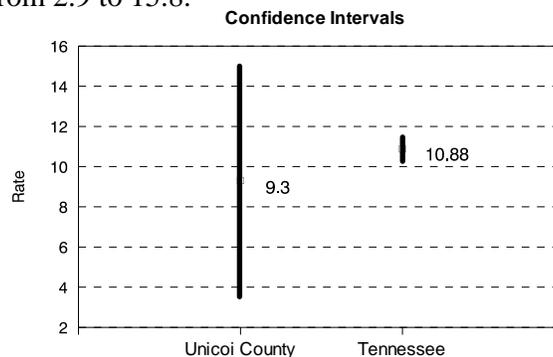
Perhaps the most common error made in vital statistics concerns the use of rates based on very small numbers. When the numbers being used to calculate rates are extremely small, large swings often occur in the rates which do not reflect real changes. Consider Unicoi County's infant mortality rates for a 5-year period:

<u>Year</u>	<u>Births</u>	<u>Infant Deaths</u>	<u>Infant Death Rates</u>
1986	151	2	13.2
1987	175	0	---
1988	199	0	---
1989	165	2	12.1
1990	166	4	24.1
1986-1990	856	8	9.3

The overall rate of 9.3 is quite similar to the state rate of 10.9 for the same time period. Yet the rate varies widely with the addition or subtraction of only two deaths. Public health officials would waste a good deal of energy if they reacted to these annual rates.

Statisticians expect a certain amount of chance variation and have methods to take this into account. The confidence interval uses the number of cases and their distributions to determine what the rate "really is". For example a statistician will say, "We are 95 percent sure that the true infant death rate for Tennessee between 1986 and 1990 was 10.88 plus or minus 0.33, that is, it lies somewhere between 10.55 and 11.21." If two rates have overlapping confidence intervals, then the difference between them may be due to this chance variation. In other words, the difference between them is not statistically significant.

Unicoi County's infant death rate for the same period was 9.35 plus or minus 6.45. Even though 5 years of data were combined in the determination of the confidence interval for Unicoi County, the 95 percent confidence interval ranges from 2.9 to 15.8.



The difference between the state rate and Unicoi County's rate is not statistically significant.

If 20 is too few, how many cases are sufficient to say that a true difference exists? Unfortunately, we have no easy rules for this. To be safe, the vital events statistician should always try to combine several years of data or consolidate geographical areas. Anyone preparing to make important decisions based on small rates should make an effort to determine if rates are statistically significant.

Step 4: Analyzing The Data

The first three steps have been fairly mechanical:

- (1) Choose the correct events and the correct population to determine the number of events which took place for the geographical areas and time periods.
- (2) Calculate the rates.
- (3) Compare these rates to determine if the differences are statistically significant.

Now the vital event statistician must begin to ask the difficult questions: Just because two rates are different, can we assume that we have a case of cause and effect? Is there another factor which is really causing the changes in the rates? If the differences which we expected did not prove to be significant, is there another item which perhaps is masking the difference? Frequently, the statistician has to refine the research question and begin all over again.



State of Tennessee
Department of Health
Cordell Hull Building
425 Fifth Avenue North
Nashville, TN 37247

Guidelines for Age-Adjustment

Overview:

Age standardization, often called “age adjustment,” is one of the key tools used to control for the changing age distribution of the populations, and thereby to make meaningful comparison of vital rates over time and between groups. Age-adjusted rate statistically is a weighted average of the age-specific rates, where the weights represent the fixed population proportion by age.

Although a crude rate, such as a crude death rate, is a very useful indicator in measuring the actual disease burden, it is often inadequate to compare crude rates among different populations (states, counties, and race or gender groups) or across time periods because many outcomes vary significantly according to age. If the age composition for different populations or for the same population across time periods is different and if the outcome being measured is influenced by age, making comparisons using crude rates may lead to incorrect conclusions. Should one need to look at the data from a more scientific and detailed point of view, consideration should be given to using the more refined measures such as age/gender/race-specific rates and age-adjusted rates as would be appropriate to the interpretations in question.

It is also important to be aware of its limitations when using age-adjusted rates. First, an age-adjusted rate is not a direct or actual measure of mortality/morbidity risk. Second, age adjustment may mask important information when the age-specific rates in the populations being compared do not have a consistent relationship. In this case, age-adjusted rates should be supplemented with age-specific rates.

For more details regarding age adjusting rate, please refer to National Center for Health Statistics (NCHS) publication “Age Standardization of Death Rates: Implementation of the Year 2000 Standard” (http://www.cdc.gov/nchs/data/nvsr/nvsr47/nvs47_03.pdf).

Guidelines:

Age-adjusted rates should be used when:

- comparing two or more populations (e.g., states, counties, race/gender groups, etc.).
- in a time series analysis for a single area that spans more than three consecutive years.

Age-adjusted rates should be computed using:

- Tennessee Population Estimates and Projections, 2003 Series from the Division of Health Statistics, Tennessee Department of Health.

- Year 2000 United States Standard Population from NCHS. We recommend the standard 11 age intervals (see Table 1).
- direct method of adjustment.

Comments:

- In some cases, crude rates or age-specific rates can be used. Since age-specific rates already take age into account, they do not need to be adjusted.
 - For certain research questions or direct measure of mortality/morbidity risk, crude rates may be appropriate.
 - Infant mortality rate is an age-specific rate (rate of deaths among children less than one year of age).
- An adjusted rate is an artificially created figure that should only be used in comparison with other adjusted rates that were computed using the same “standard population.” Besides, one can not multiply the age-adjusted rate by the actual population to determine the number of deaths (as one could with the crude rate).
- Always document (e.g., footnote) the source of population data (e.g., Tennessee Population Estimates and Projections, 2003 Series), standard population (e.g., Year 2000 U.S.), and method of adjustment (e.g., Direct or Indirect).

Age-Adjusted Mortality Rate Example:

There are three major components needed to compute age-adjusted mortality rates:

- the number of deaths.
- the population at risk.
- a “standard” population.

The following example uses 2002 provisional death data to compute age-adjusted stroke death rates for Tennessee black and white populations (see Table 1). This example is for illustrative purposes only.

Step 1: Group the study population into 11 age groups.

Step 2: Tabulate the number of deaths within each age group.

Step 3: Compute age-specific rates within each age group

$$(Number\ of\ deaths / Actual\ population * 100,000)$$

Step 4: Compute the number of expected deaths within each age group by applying age-specific rates to the standard population

$$(Age-specific\ rate\ per\ 100,000 * Standard\ population) / 100,000$$

Step 5: Compute the age-adjusted mortality rate.

$$[(sum\ of\ the\ expected\ deaths) / Standard\ million\ population * 100,000].$$

As shown in Table 1, the crude stroke mortality rate in Whites (70.3/100,000) is higher than that in Blacks (66.0/100,000). However, the age-adjusted rates indicate the opposite (67.1/100,000 vs. 98.8/100,000). We can see from the age-specific rates that the rate of stroke mortality is higher for blacks than whites in every age group except for those aged 15-24 years, where the number of deaths for blacks is zero.

Applicability:

This guideline applies to all statistical work prepared by staff in the Tennessee Department of Health.

Table 1. Crude and age-adjusted stroke mortality rates by races, Tennessee 2002

White					Black				
Age Group	Number of Deaths	2002 Population	Age Specific Rate per 100,000	Expected Deaths per 1,000,000	Number of Deaths	2002 Population	Age Specific Rate per 100,000	Expected Deaths per 1,000,000	2000 U.S. Standard Million Population
0	1	57,631	1.74	0	1	16,060	6.23	1	13,818
1-4	1	234,150	0.43	0	1	67,500	1.48	1	55,317
5-14	1	608,686	0.16	0	1	176,660	0.57	1	145,565
15-24	3	619,137	0.48	1	0	160,425	0.00	0	138,646
25-34	6	652,495	0.92	1	3	143,073	2.10	3	135,573
35-44	41	730,156	5.62	9	23	145,773	15.78	26	162,613
45-54	99	683,308	14.49	20	72	120,480	59.76	81	134,834
55-64	205	503,630	40.70	36	70	64,438	108.63	95	87,247
65-74	462	344,338	134.17	89	134	41,798	320.59	212	66,037
75-84	1170	217,886	536.98	241	145	24,142	600.61	269	44,842
85+	1336	75,482	1,769.96	274	190	9,812	1,936.40	300	15,508
Total	3,325	4,726,899	70.3	671	640	970,161	66.0	988	1,000,000

Crude death rate for white = $(3,325 / 4,726,899) \times 100,000 = 70.3/100,000$

Crude death rate for black = $(640 / 970,161) \times 100,000 = 66.0/100,000$

Age-adjusted death rate for white = $(671 / 1,000,000) \times 100,000 = 67.1/100,000$

Age-adjusted death rate for black = $(988 / 1,000,000) \times 100,000 = 98.8/100,000$

Appendix A
Comparable Code Numbers For The Ninth And Tenth Revisions
International Classification Of Diseases

Category Title According To Tenth Revision	International Classification Tenth Revision	Code Number Ninth Revision	Comparability Ratio
Shigellosis and Amebiasis	A03, A06	004, 006	**
Tuberculosis	A16-A19	010-018	0.85
Respiratory Tuberculosis	A16	010-012	0.91
Whooping Cough	A37	033	**
Scarlet Fever and Erysipelas	A38, A46	034.1-035	**
Meningococcal Infection	A39	036	1.00
Septicemia	A40-A41	038	1.19
Syphilis	A50-A53	090-097	0.64
Arthropod-borne Viral Encephalitis	A83-A84, A85.2	062-064	**
Measles	B05	055	**
Viral Hepatitis	B15-B19	070	0.83
Human Immunodeficiency Virus (HIV) Disease	B20-B24	*042-*044	1.06
Malignant Neoplasms, All Sites	C00-C97	140-208	1.01
Colon, Rectum and Anus	C18-C21	153-154	1.00
Trachea, Bronchus and Lung	C33-C34	162	0.98
Breast	C50	174-175	1.01
Prostate	C61	185	1.01
Diabetes Mellitus	E10-E14	250	1.01
Parkinson's Disease	G20-G21	332	1.00
Alzheimer's Disease	G30	331.0	1.55
Diseases of Heart	I00-I09, I11, I13, I20-I51	390-398, 402, 404, 410-429	0.99
Ischemic Heart Disease	I20-I25	410-414, 429.2	0.99
Cerebrovascular Disease	I60-I69	430-434, 436-438	1.06
Atherosclerosis	I70	440	0.96
Influenza	J10-J11	487	1.01
Pneumonia	J12-J18	480-486	0.70
Chronic Lower Respiratory Diseases	J40-J47	490-494, 496	1.05
Hernia	K40-K46	550-553	1.04
Chronic Liver Disease and Cirrhosis	K70, K73-K74	571	1.04
Cholelithiasis and Other Disorders of Gallbladder	K80-K82	574-575	0.96
Nephritis, Nephrotic Syndrome, and Nephrosis	N00-N07, N17-N19, N25-N27	580-589	1.23
Pregnancy, Childbirth, and the Puerperium	A34, O00-O95, O98-O99	630-676	**
Certain Conditions Originating in the Perinatal Period	P00-P96	760-771.2, 771.4-779	1.07
Congenital Malformations, Deformations and Chromosomal Abnormalities	Q00-Q99	740-759	0.85
Accidents	V01-X59, Y85-Y86	E800-E869, E880-E929	1.03
Motor Vehicle Accidents	V02-V04, V09.0, V09.2, V12-V14, V19.0-V19.2, V19.4-V19.6, V20-V79, V80.3-V80.5, V81.0-V81.1, V82.0-V82.1, V83-V86, V87.0-V87.8, V88.0-V88.8, V89.0, V89.2	E810-E825	0.85

Appendix A
Comparable Code Numbers For The Ninth And Tenth Revisions
International Classification Of Diseases (continued)

Category Title According To Tenth Revision	International Classification Tenth Revision	Code Number Ninth Revision	Comparability Ratio
Intentional Self-Harm (Suicide)	X60-X84, Y87.0	E950-E959	1.00
Assault (Homicide)	X85-Y09, Y87.1	E960-E969	1.00
Legal Intervention	Y35, Y89.0	E970-E978	**
Events of Undetermined Intent	Y10-Y34, Y87.2, Y89.9	E980-E989	**

Beginning with data for 1987, NCHS introduced categories *042-*044 for classifying and coding Human Immunodeficiency Virus (HIV) infection. The asterisks before the category numbers indicate that these codes are not part of the Ninth Revision International Classification of Diseases.

**Figure does not meet standards of reliability or precision.