TECHNICAL NOTES

A. SOURCES OF DATA

1. Mortality:

Pennsylvania's Certificate of Death is the source for the injury mortality data contained in this report. A copy of this certificate is included as Appendix I. Death certificates are usually completed by hospital personnel, physicians, and funeral directors.

There is a cut-off date of March 1 of the following year for compiling annual mortality files, e.g., the cut-off date for 1995 death certificates was March 1, 1996. Any data pertaining to a 1995 death for which a certificate was filed after the March 1, 1996 deadline are not included in this report. Files are updated on an annual basis after the cut-off date. Data from certificates received after a cut-off date or corrections to previously filed records are made at these times. Therefore, it is possible that injury mortality data previously obtained from or published by the Division of Health Statistics may differ slightly from the data that appear in this report.

2. Population:

Population estimates for the years 1985-1989 and 1991-1995 used to compute rates were produced jointly by the Census Bureau and the State Data Center of the Pennsylvania State University at Harrisburg under the Federal-State Cooperative Program for Local Population Estimates. The estimates are published by the Bureau of the Census in *Current Population Reports, Series P-26*. They are generated by a combination of a component procedure which uses administrative records and a regression method. The 1990 state and county populations by age, sex, and race used in this report are based on the U.S. Bureau of the Census enumerated population figures as of April 1, 1990. The 1985-1989 black population figures used to compute rates were derived from 1980 and 1990 Census data. They were estimated by advancing 1980 Census data incrementally to finally match 1990 Census figures.

B. DATA QUALITY

The Division of Health Statistics maintains query and field programs for improving the quality of data collected on death certificates. A query program is a system used to follow back with hospital personnel, funeral directors, and physicians concerning incomplete or conflicting information. The follow-back contact is usually done by telephone and is based both on a manual and a computer editing procedure.

The National Center for Health Statistics (NCHS) monitors Pennsylvania's coding of statistical data on death certificates. A 0.5 percent sample of death records coded and submitted monthly by the State are used as a quality control mechanism by NCHS. NCHS codes these sample records independently and then conducts an item-by-item computer match of codes entered by the state and NCHS. NCHS has established a two percent upper limit for coding differences involving any one data item of these sample records, with the exception of cause of death. A five percent limit is established for that item due to the complexity of that coding.

Deaths to Pennsylvania residents that occurred in other states are included in this report. Inclusion of these data is possible due to an agreement among all registration areas in the United States for resident exchange of copies of death certificates.

C. DATA CONFIDENTIALITY

The confidentiality of death certificates is strictly maintained by the Division of Health Statistics. All published reports of vital statistics contain only aggregate data. Individually identifiable information is only released, per Act 66, Vital Statistics Law of 1953, for health research purposes and to government agencies in pursuance of their official duties upon completion, review, and approval of an application for access to protected data. Certified copies are also released to family members or legal representatives for determination of personal or property rights.

D. DEFINITIONS

AGE-SPECIFIC RATE – Number of events for a specified age group per 100,000 population in the same age group.

AGE-ADJUSTED RATE (Direct Method) – Age-specific rates for a selected population are applied to a standard population (in this report the 1940 United States population was used) to calculate what rate **would be expected** if the selected population had the same age distribution as the standard population. The total of these **expected** deaths divided by the total of the standard population and multiplied by 100,000 yields the age-adjusted rate per 100,000. (It is important to use the same standard population in the computation of each age-adjusted rate to allow comparability. Age-adjusted rates are artificial measurements and should never be compared with any other type of rate or be used to calculate the actual number of events.) The 1940 United States standard population distribution used in calculating age-adjusted death rates for this report is shown below:

AGE	POPULATION
All Ages	1,000,000
Under 1	15,343
1-4	64,718
5-14	170,355
15-24	181,677
25-34	162,066
35-44	139,237
45-54	117,811
55-64	80,294
65-74	48,426
75-84	17,303
85+	2,770

YEARS OF POTENTIAL LIFE LOST or YPLL – The YPLL represents the number of years of life lost for a specified cause of death before age 65. It is calculated by subtracting the age at death from 65 for each decedent being studied, and then adding all these differences for a total YPLL. This statistic emphasizes mortality due to causes of death that tend to be more predominant among younger persons.

YEARS OF POTENTIAL LIFE LOST PER DEATH or YPLL RATE – This rate is computed by dividing the total years of potential life lost for a specified cause of death by the total number of deaths resulting from that same cause.

E. RELIABILITY OF RATES

Rates, even when they are based on full population counts (as in this report), should be considered estimates and subject to error. The observed death rate is an estimate of the true or underlying death rate. Rates are subject to chance variation and users of this report should be aware of this problem. The variation of the rate is directly related to the number of events used to calculate the rate. The smaller the number of events used in the calculation of a rate, the higher the variability of the rate will be. Rates based on unusually small numbers of events over a specified period of time or for a sparsely populated geographic area should be of particular concern and be used cautiously.

When few events or small populations are evident in calculating/studying rates, multiple-year summary rates, referred to as average annual rates, will sometimes provide a much better perspective or measurement of an outcome. This was the method used in this report. Expanding the period of time studied enlarges the absolute numbers and adds more credence to a statement regarding a rate.

It is also common practice among data users familiar with health statistics to calculate what is called a standard error (SE) of a rate. This statistic defines an observed rate's variability and can be used to calculate a confidence interval (CI) to determine the range of probable values for the true or underlying rate. The following section describes the formulas and methods used in this report to compute standard errors and confidence intervals.

F. COMPARING AGE-ADJUSTED RATES

As mentioned above, a first step in comparing rates is the computation of a standard error, defining the rate's variability. The usual formula given for computing the standard error of an age-adjusted rate (Chiang, 1961) is very complex and not often understood or used by the average health data user. However, the standard error can be approximated with the following less complex formula (Keyfitz, 1966):

$$SE = R / \sqrt{N}$$

where:

R = age-adjusted rate
N = number of deaths

This estimate assumes the rate to be a binomial proportion. As an example, let's use the state's 1991-1995 age-adjusted death rate for homicide of 7.0 to calculate an estimated SE. The rate was based on 3,870 homicides. The square root of 3,870 is 62.2. By dividing the rate of 7.0 by 62.2, one obtains the estimated SE of 0.11.

This estimated SE can then be used to compute a 95% confidence interval (CI) for the rate. The standard formula for determining the 95% CI of a rate is:

$$R \pm (1.96 \times SE)$$

Following this formula, for the rate we are using, produces an equation of 7.0 ± (1.96 × 0.11). The result is 7.0 ± 0.22. Then, by subtracting and adding 0.22 against the original rate of 7.0, a range can be calculated. One could then state, with 95% certainty, that the true age-adjusted homicide rate for the state during 1991-1995 was between 6.78 and 7.22.

To compare a county's age-adjusted homicide rate for 1991-1995 with the state's corresponding rate, one must go through the same steps shown above to obtain the 95% CI for that county's rate. If the CIs for the state and county rates do not overlap, then the two rates are considered to be significantly different at the 95% confidence level. For example, the table below shows homicide rates and confidence intervals for Pennsylvania, Fayette County, and Lehigh County for 1991-1995.

	Deaths	Rate	CI
Pennsylvania	3,870	7.0	6.78 - 7.22
Fayette County	36	5.2	3.50 - 6.90
Lehigh County	69	4.9	3.74 - 6.06

Both counties had homicide rates below the state rate. However, upon closer examination, we see that Fayette County's CI overlaps the state's CI. Therefore, we can say (with 95% confidence) that there was no significant difference between Fayette County's homicide rate and the state's homicide rate for the years 1991-1995. On the other hand, Lehigh County's CI does not overlap and is below the state's CI. Thus, we can say that Lehigh County's homicide rate was significantly lower than the state's homicide rate for the years 1991-1995.

This example points to an important result that the user should note. The smaller the number of events upon which the rate is based, the larger the SE and range of the CI will be. This clearly demonstrates the wider variability (and less reliability) of rates based on smaller numbers. As a general rule, **age-adjusted rates based on less than twenty events should be considered unstable and are not recommended for comparative use or in determining significance**. For this reason, the CIs were not computed, compared and shown for any age-adjusted death rate in this report based on less than twenty events.

Another important note concerns the significance tests in Tables 1-9 (D) of this report. The confidence limits shown are rounded to the nearest tenth. However, significance tests use the numbers before they are rounded, to determine significance. Therefore, a county's upper or lower confidence limit (rounded) may appear to equal the state's upper or lower confidence limit, but could be labeled as significantly different than the state rate.

G. CAUSES OF DEATH

Resident deaths with an underlying cause assigned to one of the International Classification of Diseases, Ninth Revision (ICD-9) codes listed below are included in this report:

CAUSE OF DEATH	ICD-9 CODES
Unintentional Injuries	E800-E949
Motor Vehicle Crashes	E810-E825
Drug Poisoning	E850-E858
Falls and Fall-Related Injuries	E880-E888
Fire and Flames	E890-E899
Drowning and Submersion	E830,E832,E910
Suicide	E950-E959
Homicide	E960-E969
Firearm-Related Injuries	E922.0-E922.3, E922.8-E922.9, E955.0-E955.4, E965.0-E965.4, E970, E985.0-E985.4