Data Source
The Pennsylvania Department of Health’s vital statistics registration system was the source for the birth and death statistics that appear in this report except for work-related injury deaths which were from the Census of Fatal Occupational Injuries as conducted by the U.S. Department of Labor. The National Center for Health Statistics was the source for the U.S. birth and death statistics that appear in this report. The U.S. birth statistics are final 1998 data. However, the 1998 U.S. death statistics are preliminary, and only 1997 U.S. data are available for breast cancer.

The Department's Communicable Disease Surveillance, Sexually Transmitted Disease, and Tuberculosis Control Programs were the sources for the number of measles, syphilis, and tuberculosis cases reported. For the number of AIDS cases reported, data from the Department's AIDS Reporting System were used.

The U.S. Bureau of the Census 1995 income data were used for the estimated number and percentage of related children ages 5-17 and all children under age 18 living below the poverty level by county. Access their website at www.census.gov to review complete data tables, including confidence intervals and data limitations.

The 1990 population used in this report is U.S. Bureau of the Census enumerated population figures as of April 1, 1990. Population estimates for the years 1992 through 1998 used to compute rates were produced jointly by the U.S. Bureau of the Census and the State Data Center of the Pennsylvania State University at Harrisburg under the Federal-State Cooperative Program for Local Population Estimates. The estimated county population figures used to compute the rates that appear in this report are available from the Bureau of Health Statistics upon request. The 1940 United States standard million population used in calculating age-adjusted death rates follows:

<table>
<thead>
<tr>
<th>Age</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Ages</td>
<td>1,000,000</td>
</tr>
<tr>
<td>Under 1</td>
<td>15,343</td>
</tr>
<tr>
<td>1-4</td>
<td>64,718</td>
</tr>
<tr>
<td>5-14</td>
<td>170,355</td>
</tr>
<tr>
<td>15-24</td>
<td>181,677</td>
</tr>
<tr>
<td>25-34</td>
<td>162,066</td>
</tr>
<tr>
<td>35-44</td>
<td>139,237</td>
</tr>
<tr>
<td>45-54</td>
<td>117,811</td>
</tr>
<tr>
<td>55-64</td>
<td>80,294</td>
</tr>
<tr>
<td>65-74</td>
<td>48,426</td>
</tr>
<tr>
<td>75-84</td>
<td>17,303</td>
</tr>
<tr>
<td>85+</td>
<td>2,770</td>
</tr>
</tbody>
</table>

DEFINITIONS of TERMS
Death rates by cause (and for all causes) are per 100,000 population (except the rate for breast cancer which is per 100,000 females) and are age-adjusted to the 1940 standard million U.S. population (see above) except the rate for work related injury which is a crude rate per 100,000 population. Infant death rates are per 1,000 live births for the specified years.

Incidence rates are per 100,000 population for the specified years.

The International Classification of Diseases (ICD-9) codes for the selected causes of death shown in this report is as follows:

- Motor Vehicle Crash (E810-E825)
- Suicide (E950-E959)
- Lung Cancer (162)
- Female Breast Cancer (174)
- Cardiovascular Disease (390-448)
- Heart Disease (390-398, 402, 404-429)
- Stroke (430-438)
- Homicide (E960-E978)

Low Birth Weight is less than 2,500 grams or 5 pounds and 9 ounces.

Hispanics can be of any race.

All calculations exclude any unknowns.

Age-Adjusted Rates
There are many characteristics of a population that can render a crude rate of little use, especially when comparing different populations. (A crude rate is usually defined as: total number of events divided by total population at risk, then, multiplied by 1,000 or 100,000.) Any unique demographic factors such as those related to age, sex or race are used or compared. The median age of Pennsylvania’s population has been for many years one of the highest among all the states. Therefore, age-adjusted rates offer a more refined measurement to compare experiences over geographic areas or periods of time. However, there are limitations to their use and one should be familiar with these types of rates when using them.

It is important to use the same standard population in the computation of each age-adjusted rate to allow comparability. Also, note that 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 age-adjusted death rates that appear in this report were calculated using the direct method and the 1940 United States standard million population distribution (shown in the column on the left).

To calculate an age-adjusted rate using the direct method, the age-specific rates must first be calculated for each of the age groups (as shown in the 1940 standard population distribution) using the enumerated or estimated population figures for the time period and population under study. Each age-specific rate is then multiplied by the population figure of the corresponding age group in the standard population breakdown. The resultant figures are the number of deaths to be expected.
if the population under study had the same age distribution as the standard population. The total of these expected events is then divided by the total of the standard population (in the report 1,000,000). This dividend is then multiplied by 100,000 to yield the age-adjusted rate per 100,000 population.

**Reliability of Rates**

All rates are subject to variation. This variation 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 will be the variability of the rate. Rates (or percentages) 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 used cautiously. When few events or small populations are evident in calculating/studying rates, multiple-year summary rates usually referred to as average annual rates, will sometimes provide a much better perspective or measurement of an outcome. Expanding the period of time studied enlarges the absolute numbers and adds more credence to a statement regarding a rate. Another approach is to expand the geographical area of study, thereby, enlarging the number of events. Adjoining counties can be grouped into regions according to any demographic features they may share, i.e., rural counties with mostly white, older populations.

It is also common practice among data users familiar with health statistics to calculate what is called a standard error (SE) of a rate when comparing rates. This statistic defines a rate's variability and can be used to calculate a confidence interval (CI) to determine the range of possible values for the true rate. If a set standard, goal or target value is included in a rate's confidence interval, there is no significant difference between the two. However, there are various statistical formulas for comparing rates depending on the types of rates or populations being studied and the number of events involved. The following section discusses various statistical formulas that were used to compare the rates that appear in this report.

**NOTE:** Before we proceed with presenting formulas for comparing rates and ratios/percentages, the user should understand that these statistical tools for analyzing/comparing rates are crude and rather conservative approaches, especially the formulas presented for comparing age-adjusted rates. A user may wish to utilize more precise and sophisticated calculations performed by computer software such as SPSS or SAS. Consultation with a statistician or other professional familiar with analyzing health statistics may also be a consideration before pursuing any further study.

**Comparison of Age-Adjusted Rates**

As mentioned above, a first step in comparing rates is the computation of a standard error (SE), 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 average user can approximate the standard error of an age-adjusted rate with the following less complex formula (Keyfitz, 1966):

\[
SE = \frac{R}{\sqrt{N}}
\]

where:

- \( R \) = (age-adjusted) rate
- \( N \) = number of events (deaths)

This estimate assumes the rate to be a binomial proportion. As an example, let's use the state's average annual (1996-1998) age-adjusted death rate for suicide of 10.7 to calculate an estimated SE. The rate was based on 4,171 suicides. The square root of 4,171 is 64.58. By dividing the rate of 10.7 by 64.58, one obtains the estimated SE of 0.1657. The 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 10.7 \( \pm (1.96 \times 0.1657) \) and the result is 10.7 \( \pm 0.32 \). Then, by subtracting and adding 0.32 against the original rate of 10.7, a range can be calculated and considered the estimated 95% confidence interval for the state, i.e., 10.38 - 11.02. One could then state, with 95% certainty that the actual age-adjusted suicide rate for the state during 1996-1998 was between 10.38 and 11.02.

To compare a particular county’s age-adjusted suicide rate for 1996-1998 with the state’s corresponding rate, one must go through the same steps shown directly above to obtain the 95% CI for that county’s rate. If the rate for the state is not included in the CI, then the county rate is considered to be significantly different, at the 95% confidence level. For example, at first glance, Bradford County’s age-adjusted suicide rate for 1996-1998 of 14.4 (based on 31 deaths) seems much higher than the corresponding state rate of 10.7. However, calculation of a 95% CI for Bradford County’s rate would produce a rather wide range of 9.33-19.47. Since this range for Bradford County also includes 10.7 or the state rate, we can say that the county rate is not significantly different than the state rate, at the 95% confidence level. If we were comparing two counties, any significant difference would be determined by whether their confidence intervals overlapped or not. However, please note that the formula for computation of the SE that we are using is not as precise as others and the application of a more precise methodology may produce somewhat different results. Another important result the user of this formula should note is that, the smaller the number of events, upon which the rate is based, the larger the SE and 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 mortality rate in this report based on less than twenty events.

Comparison of Crude Rates/Ratios
A crude rate is easily computed and usually based on the number of vital events and the total population for a specific area or group, i.e., number of births or deaths among a specific population per 1,000 (or 100,000) of that specific population. A ratio is simply a proportion or percentage, usually a rate per 100. Any of the indicators that are not presented in this report as age-adjusted rates can be considered crude rates or ratios. Before comparison of these figures can be done, they should first be identified as dependent or independent and then defined as being based on a small or large number of events.

DEPENDENT vs. INDEPENDENT CRUDE RATES:
Two crude rates or ratios are considered dependent when the same events are included in their numerators. Examples of this include a state rate and a county or city rate or rates that share or overlap the same time periods, i.e., two multiple-year summary rates for the state – one for 1980-1985, the other for 1980-1989. Two rates are considered to be independent when they do not include any of the same data or events in their numerators, such as rates for two different counties.

NUMBER of EVENTS: When comparing two dependent or independent rates, determining whether a significant difference exists between the two rates or whether the difference is caused solely by chance requires a rather complex statistical computation. The number of events upon which the two observed rates are based is of primary importance. The statistical formula for determining significance is different for a rate based on a small number of events as compared to the formula for a rate based on a large number of events. Exactly what is considered a small number of events is arbitrary but, as a general rule, one can define “small number” as less than 100 events. Crude rates or ratios based on less than ten events should be considered unstable and are not recommended for comparative use or in determining significance. The formulas are also different depending on whether the rates being compared are dependent or independent.

Four formulas for comparing crude rates and ratios are presented next: one recommended for use in comparing dependent rates based on a small number of events; another, for comparing dependent rates based on a large number of events; a third, for independent rates based on less than 100 events; and, a fourth, for comparing independent rates based on 100 or more events. A sample step-by-step calculation is shown for the first formula to demonstrate its use.

COMPARISON of DEPENDENT CRUDE RATES BASED on SMALL NUMBER OF EVENTS: When the (county or local) crude rate or ratio to be compared to a standard (state or national) rate or ratio is based on 10-99 events, actual and estimated numbers of events are used to determine statistical significance. The formula for this situation is shown below:

\[
\mu = \frac{(o-e)}{\sqrt{e}}
\]

where:
- \(o\) = the number of events for the county or local area to be compared
- \(e\) = the expected number of events for the county or local area (based on the state or national crude rate)

If \(\mu\) has a value greater than +1.96, the county rate is considered to differ significantly at a 95% confidence level from the state rate to which it is being compared. The value for \(o\) is a readily available figure; however, \(e\) must be specially computed. To compute the expected number of events for the county based on a state or national crude rate, first change the state rate to a percentage or rate per person. For example, if the state rate was 14.5 per 1,000 population, simply divide 14.5 by 1,000; the result is .0145. Then, multiply the value of the denominator in the county rate (the population used to compute the rate) by this figure to obtain the value for \(e\) or the expected number of events for the county.

As an example for computation of this formula, use a county infant death rate of 13.8 per 1,000 resident live births. This rate was based on 58 resident infant deaths occurring among 4,205 resident live births for the county. The comparable state rate that year was 9.5. Step-by-step computation would yield the following results:

1. \(\sqrt{e} = \sqrt{39.9} = 6.3\)
2. \((o - e) = 58 - 39.9 = 18.1\)
3. \((o - e) / \sqrt{e} = 18.1 / 6.3 = 2.9 \text{ or } \mu\)

Since the value of \(\mu\) in the previous computation exceeds the value of 1.96, it can be stated that the difference between the county’s infant death rate and the state’s rate that year was significant at the 95% confidence level. In other words, the user can be up to 95% confident that the county’s true infant death rate that year was significantly higher than the infant death rate for the state. A negative value of more than -1.96 would mean a significantly lower rate.

COMPARISON of DEPENDENT CRUDE RATES BASED on LARGE NUMBER OF EVENTS: The following formula for determining the significance between two observed, dependent crude rates with 100
or more events in the numerator of the county or local rate is more complex than the previous formula for dependent rates.

\[ \mu = (r - s) \sqrt{\frac{n}{(s - s^2)}} \]

where:

- \( r \) = the county or local rate to be compared, expressed as a rate per person
- \( s \) = the state (or national, regional, etc.) rate expressed as a rate per person
- \( n \) = the population figure used for computing the county or local rate

To compute a rate per person, divide the rate by the population number used to express the rate. For example, the rate per person for a death rate of 23.5 per 100,000 would be calculated by dividing 23.5 by 100,000. The result is 0.000235.

Determining significance according to the \( \mu \) value follows the same rules as listed in the previous section for comparing dependent rates based on a small number of events.

**COMPARISON of INDEPENDENT CRUDE RATES BASED on SMALL NUMBER OF EVENTS:** The following formulas can be used to compute a 95% confidence interval to determine the statistical significance of the difference between two independent crude rates when both rates are based on 10-99 events. The first step is to calculate the difference (D) between the two rates with the following formula:

\[ D = r_1 - r_2 \]

where:

- \( r_1 \) = rate for County 1
- \( r_2 \) = rate for County 2

The 95% confidence interval (CI) is then computed using the following formula:

\[ CI = D \pm 1.96 \sqrt{\left(\frac{1}{d_1} + \frac{1}{d_2}\right)} \]

where:

- \( d_1 \) = number of events for County 1
- \( d_2 \) = number of events for County 2

If the range of numbers derived from the confidence interval (CI) contains the value of 1, then a significant difference does not exist, at 95% confidence. If the range of numbers does not contain the value of 1, then it can be stated that the ratio between the two county rates is significantly different, with 95% confidence.

**COMPARISON of INDEPENDENT CRUDE RATES BASED on LARGE NUMBER OF EVENTS:** If two independent crude rates or ratios are being compared and both or one of the figures is based on 100 or more events, a two-step calculation is performed to construct a 95% confidence interval for the ratio between the two rates. Please note, however, that whenever only one of the two rates is based on 100 or more events, then that rate must be used as \( r_2 \) in the following formula.

The formula for calculating the ratio (R) between the two rates is:

\[ R = \frac{r_1}{r_2} \]

where:

- \( r_1 \) = rate for County 1
- \( r_2 \) = rate for County 2

The formula for the 95% confidence interval (CI) for the ratio between the two independent rates is:

\[ CI = R \pm 1.96 (R) \sqrt{\left(\frac{1}{d_1} + \frac{1}{d_2}\right)} \]

where:

- \( d_1 \) = number of events for County 1
- \( d_2 \) = number of events for County 2

If the range of numbers derived from the confidence interval (CI) for the ratio contains the value of 1, then a significant difference does not exist, at 95% confidence. If the range of numbers does not contain the value of 1, then it can be stated that the ratio between the two county rates is significantly different, with 95% confidence.