

Technical Notes

Data Sources

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 latest available U.S. birth statistics are final 2006 data. **Please note that the 2006 U.S. prenatal care data that appear in this report are limited and are based on only the eighteen states [Delaware, Florida, Idaho, Kansas, Kentucky, Nebraska, New Hampshire, New York (excluding New York City), North Dakota, Ohio, Pennsylvania, South Carolina, South Dakota, Tennessee, Texas, Vermont, Washington, and Wyoming] that have implemented the 2003 Revision of the U.S. Certificate of Live Birth. These data are not comparable to those states that have not implemented the 2003 Revision.** The latest available U.S. death statistics are preliminary 2007 data (female breast cancer are final 2006 data).

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.

Population estimates, for the years 2005 through 2007, used to compute rates were produced jointly by the U.S. 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 estimated county population figures used to compute the rates that appear in this report are available from the Bureau of Health Statistics and Research upon request. The 2000 United States standard million population used in calculating age-adjusted death rates follows:

<u>Age</u>	<u>Population</u>
All Ages	1,000,000
Under 1	13,818
1-4	55,317
5-14	145,565
15-24	138,646
25-34	135,573
35-44	162,613
45-54	134,834
55-64	87,247
65-74	66,037
75-84	44,842
85+	15,508

The U.S. Census Bureau 2007 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.

Definitions of Terms

Death rates by cause (and for all causes) are per 100,000 population (except the rate for female breast cancer which is per 100,000 females) and are age-adjusted to the 2000 standard million U.S. population, except the rate for work related injury, which is a crude rate per 100,000 population.

Infant mortality 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-10) codes** for the selected causes of death shown in this report are as follows:

	<u>ICD-10</u>
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
Intentional Self-harm (Suicide)	U03, X60-X84, Y87.0
Lung Cancer	C33-C34
Female Breast Cancer	C50 (sex = female)
Cardiovascular Disease	I00-I78
Diseases of Heart	I00-I09, I11, I13, I20-I51
Stroke	I60-I69
Assault (Homicide)	U01-U02, X85-Y09, Y87.1

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 and race may not be accounted for in crude rates. 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.

The age-adjusted death rates that appear in this report were calculated using the direct method and the 2000 United States standard million population distribution

(shown previously). It is important to use the same standard population in the computation of each age-adjusted rate to allow comparability. **Please note that reports for Health Status Indicators prior to the 2001 issue used the 1940 U.S. standard million population to calculate age-adjusted rates. Therefore, the age-adjusted rates that appear in this report should not be compared to the age-adjusted rates that appeared in reports prior to the 2001 issue.** This change in the use of a standard population is in response to national/federal guidelines. 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.

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 2000 standard population distribution on the previous page) 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):

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

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 (2005-2007) age-adjusted death rate for suicide (10.9) to calculate an estimated SE. The rate was based on 4,192 suicides. The square root of 4,192 is 64.75. By dividing the rate of 10.9 by 64.75, one obtains the estimated SE of 0.1684. The estimated SE can then be used to compute a 95 percent confidence interval (CI) for the rate. The standard formula for determining the 95 percent CI of a rate is:

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

Following this formula, for the rate we are using, produces an equation of $10.9 \pm (1.96 \times 0.1684)$ and the result is 10.9 ± 0.33 . Then, by subtracting and adding 0.33 against the original rate of 10.9, a range can be calculated and considered the estimated 95 percent confidence interval for the state, i.e., 10.57 - 11.23. One could then state, with 95 percent certainty that the actual age-adjusted suicide rate for the state during 2005-2007 was between 10.57 and 11.23.

To compare a particular county's age-adjusted suicide rate for 2005-2007 with the state's corresponding rate, one must go through the same steps shown directly above to obtain the 95 percent 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 percent confidence level. For example, at first glance, Perry County's age-adjusted suicide rate for

2005-2007 of 15.0 (based on 21 deaths) seems much higher than the corresponding state rate of 10.9. However, calculation of a 95 percent CI for Perry County's rate would produce a rather wide range of 8.58-21.42. Since this range for Perry County also includes 10.9 or the state rate, we can say that the county rate is not significantly different from the state rate, at the 95 percent 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 of using this formula 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, age-adjusted rates and CIs were not computed, compared and shown in this report when the number of events was less than twenty.

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 2000-2005, the other for 2000-2008. Two rates are considered to be independent when they do not include any of the same data or events in their numerator, 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. Please note that when the number of

events is less than 10, then crude rates are not displayed and significance testing is not performed due to the unreliability of rates based on small numbers. The formulas for determining significance 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:

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

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 μ has a value greater than + 1.96, the county rate is considered to differ significantly at a 95 percent 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. (Note: If comparing percentages, divide by 100.) 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 mortality 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:

$$o = 58$$

$$e = (9.5/1,000) 4,205 = 39.9$$

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

Since the value of μ in the previous computation exceeds the value of 1.96, it can be stated that the difference between the county's infant mortality rate and the state's rate that year was significant at the 95 percent confidence level. In other words, the user can be up to 95 percent confident that the county's true infant mortality rate that year was significantly higher than the infant mortality 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.

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

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 μ 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 percent 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 percent confidence interval (CI) is then computed using the following formula:

$$CI = D \pm [\sqrt{(CL_1^2 + CL_2^2)}]$$

where:

- CL_1 = confidence limit for County 1 rate
- CL_2 = confidence limit for County 2 rate

This computation becomes a three-step process due to the need to construct a confidence limit or CL (the

numerical value that determines the range of the confidence interval) for each rate before the above formula can be calculated for CI. The formula for 95 percent confidence is as follows:

$$CL = [1.96 (r / \sqrt{d})]$$

where: d = number of events

At the end of this three-step process, if the confidence interval or the range of the numbers calculated for the difference between the two rates includes the value of 0, then it can be stated that the two rates are not significantly different, with 95 percent confidence. Of course, if the range of numbers does not contain 0, then the difference between the rates is considered significant, with 95 percent confidence. For example, a computed confidence interval (CI) of 4.38 for a rate difference (D) of 6.8 would result in a range of 2.42 to 11.18. Since that range does not include the value of 0, the difference between the two rates being compared can be considered significant, with 95 percent confidence.

COMPARISON of INDEPENDENT CRUDE RATES BASED on LARGE NUMBER OF EVENTS: If two independent crude rates or ratios are being compared and one or both of the figures is based on 100 or more events, a two-step calculation is performed to construct a 95 percent 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:

$$R = (r_1 / r_2)$$

where:

- r_1 = rate for County 1
- r_2 = rate for County 2

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

$$CI = R \pm [1.96 (R) \sqrt{((1 / d_1) + (1 / d_2))}]$$

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 percent 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 percent confidence.

References

1. **Centers for Disease Control and Prevention**, *Summary of Notifiable Diseases, United States 2007*. Morbidity and Mortality Weekly Report: Vol. 56 No. 53. Atlanta, Georgia: July 09, 2009.
2. **Chaing CL**. *Standard Error of the Age-Adjusted Rate*. Vital Statistics Special Reports 1961;47(9).
3. **Curtin LP** and **Klein RJ**. *Direct Standardization (Age-Adjusted Death Rates)*. Statistical Notes; Number 6. Hyattsville, Maryland: National Center for Health Statistics. March 1995.
4. **Dever, Alan GE**. *Epidemiology in Health Services Management*. Rockville, Maryland: Aspen Systems Corporation. 1984.
5. **Freedman MA**. *Health Status Indicators for the Year 2000*. Statistical Notes; Vol. 1 No. 1. Hyattsville, Maryland: National Center for Health Statistics. 1991.
6. **Keyfitz N**. *Sampling Variance of Standardized Mortality Rates*. Human Biology, 38:309-317, 1966.
7. **Klein RJ** and **Hawk SA**. *Health Status Indicators: Definitions and National Data*. Statistical Notes; Vol. 1 No. 3. Hyattsville, Maryland: National Center for Health Statistics. 1992.
8. **National Center for Health Statistics**, *Births: Final Data for 2006*, National Vital Statistics Report Vol. 57 No. 7; Hyattsville, Maryland. DHHS Publication No. (PHS) 2009-1120 (1/2009).
9. **Centers for Disease Control and Prevention. National Center for Health Statistics**. *Vital Stats*. <http://www.cdc.gov/nchs/vitalstats.htm>. May 12, 2009.
10. **National Center for Health Statistics**, *Deaths: Preliminary Data for 2007*, National Vital Statistics Report Vol. 58 No. 1; Hyattsville, Maryland. DHHS Publication No. (PHS) 2009-1120 (8/2009).
11. **National Center for Health Statistics**, *Deaths: Final Data for 2006*, National Vital Statistics Report Vol. 57 No. 14; Hyattsville, Maryland. DHHS Publication No. (PHS) 2009-1120 (4/2009).
12. **Pennsylvania Department of Health**. *Comparing Rates – Part 1: Dependant Rates*. Statistical News from the State Health Data Center; Vol. 12 No. 3, pp. 4-5. Harrisburg, Pennsylvania: State Health Data Center. 1989.
13. **Pennsylvania Department of Health**. *Comparing Rates – Part 2: Independent Rates*. Statistical News from the State Health Data Center; Vol. 12 No. 4, pp. 4-5, 8. Harrisburg, Pennsylvania: State Health Data Center. 1989.
14. **U.S. Department of Health and Human Services**. *Healthy People 2000: National Health Promotion and Disease Prevention Objectives for the Nation*. Washington DC: Public Health Service. 1991.
15. **U.S. Department of Health and Human Services**. *Healthy People 2010*. 2nd ed. 2 vols. Washington DC; US Government Printing Office, November 2000.
16. **U.S. Department of Health and Human Services**. *International Statistical Classification of Diseases and Related Health Problems*. Tenth Revision. Prepared by the World Health Organization, Geneva 1992.