

Guidelines for Using and Developing Rates for Public Health Assessment

Revision Date: May 31, 2012

Primary Contact: Juliet VanEenwyk, Ph.D., State Epidemiologist for Non-Infectious Conditions

[Rate Basics](#)

[Why are rates used in public health assessment?](#)

[What is the difference between crude, age-adjusted and age-specific rates?](#)

[Should I calculate a rate when the number of events is small?](#)

[Will the rate be compared to other rates?](#)

[How do I know whether two rates are different?](#)

[Guidelines for using and developing rates](#)

[Number of events \(numerator\)](#)

[Population at risk \(denominator\)](#)

[Crude rates](#)

[Age-adjusted rates](#)

[Age-specific rates](#)

[Comparing rates](#)

[Unstable rates due to small numbers](#)

[Methods for age adjustment](#)

[Direct adjustment](#)

[U.S. Standard Populations for direct adjustment](#)

[Indirect adjustment](#)

[Glossary](#)

[References and Resources](#)

Rate Basics

A rate consists of a numerator and a denominator.

The numerator is the number of health events. This is often the same as the number of people who experience an event, but for some health conditions, one person may experience the event more than once. For example, one individual may have multiple hospitalizations for the same condition in a given year.

To measure incidence or prevalence of the condition, you usually want to count people. To measure the public health burden, you may want to count events. Actions based on the data may be different depending on whether the rate represents many individuals with only one event or a smaller number of individuals who have had many events.

It is customary to count only events that occur among the population at risk.

[Guideline: Number of events \(numerator\)](#)

The denominator is also known as the population at risk. Everyone in the population at risk must be eligible to be counted in the numerator if they have the event of interest. For example, in looking at female breast cancer, we cannot include men in the population at risk, because men with breast cancer would not be included in the numerator.

[Guideline: Population at risk \(denominator\)](#)

Once the numerator and denominator are established, how do we decide which rate is the most appropriate to use. The following questions are useful.

Why are rates used in public health assessment?

Much of public health assessment involves describing the health status of a defined community by looking at changes in the community over time or by comparing health events in that community to events occurring in other communities or the state as a whole. In making these comparisons, we need to account for the fact that the number of health events depends in part on the number of people in the community. To account for growth in a community or to compare communities of different sizes, we usually develop rates to provide the number of events per population unit.

Also, the frequency with which health events occur is almost always related to age. For example, acute respiratory infections are more common in children of school age because of their immunologic susceptibility and exposure to other children in schools. Chronic conditions, such as arthritis and atherosclerosis, occur more frequently in older adults because of a variety of physiologic consequences of aging. Mortality tends to increase rapidly after the age of 40. In fact, the relationship of age to risk often dwarfs other important risk factors. Because the relationship of age to risk is often resistant or impervious to interventions, analysts often remove the effects of differences in age structure when comparing rates across populations by calculating age-adjusted and age-specific rates.

What is the difference between crude, age-adjusted and age-specific rates?

Crude rates

Crude rates are recommended when a summary measure is needed and it is not necessary or desirable to adjust for other factors. For example, rates of infectious diseases, such as tuberculosis and hepatitis, are usually not age adjusted, because public health officials are interested in the overall burden of disease in the total population irrespective of age.

A crude rate is calculated by dividing the total number of events in a specified time period by the total number of individuals in the population who are at risk for these events and multiplying by a constant, such as 1,000 or 100,000 [e.g., (numerator/denominator) x constant]. For example, number of deaths in King County for 2010 (numerator) divided by the population of King County in 2010 (denominator) times 100,000 (constant) gives the 2010 crude death rate per 100,000 population for King County.

[Guideline: Crude Rates](#)

Age-adjusted rates

Adjusted rates are used when comparing rates of health events affected by confounding factors. They are used when comparing different populations or for comparing trends in a given population over time. Because the occurrence of many health conditions is related to age, the most common adjustment for public health data is age adjustment.

The process of age adjustment removes differences in the age composition of two or more populations to allow comparisons between these populations independent of their age structures. For example, a county's age-adjusted death rate is the weighted average of the age-specific death rates observed in

that county, with the weights derived from the age distribution in an external population standard, such as the U.S. population. Different standard populations have different age distributions and the choice will affect the resulting age-adjusted rate. If the age-adjusted rates for different counties are calculated with the same weights (i.e., using the same age groups and population standard), the effect of any differences in the counties' age distributions is removed.

Age-adjusted rates should be presented when summary measures are needed to compare across groups, such as counties or years. Data analysts should also inspect age-specific rates (Choi, 1999), because

- Age-adjusted rates can mask important trends. For example Anderson and Rosenberg (1998) show decreasing cancer mortality rates for people under 24 and increasing rates for people over 65, while the age-adjusted rates change very little.
- Age-adjusted rates can over- or under-estimate differences when age-specific rates of the populations being compared do not show a consistent relationship (i.e., the trend is not in the same direction for all age-specific rates or the ratio of age-specific rates is different for different age groups). Comparing age-specific rates across groups can help the data analyst understand the potential for over- or under-estimation.

[Guideline: U.S. Standard Populations for Direct Adjustment](#)
[Guideline: Age-adjusted Rates, Direct Method](#)

When the number of events is relatively small, the age-specific rates needed to calculate an age-adjusted rate by the [direct method](#) are unstable. This may result in unstable age-adjusted rates when using the direct method of age adjustment. Additionally, since the age-adjusted rate calculated by the direct method provides a somewhat arbitrary summary statistic that depends on the choice of a standard, it may not provide the best summary measure in explaining health status to communities. An alternative approach is to use [indirect adjustment](#). Indirect adjustment compares the number of events that occur in a given population to the number expected if the rate in that population is the same as in a standard population. This method is for comparing observed to expected numbers within a population; it cannot be used to compare across multiple populations.

[Guideline: Indirect Adjustment](#)

Age-specific rates

Because age-adjusted rates can mask important trends or over- or under-estimate differences, age-specific rates are used for comparing age-defined subgroups when rates are strongly age-dependent. Age-specific rates are also used when specific causal or protective factors or the prevalence of risk exposures are different at different ages. For example, at highest risk for head injury are males 15-24 years of age (related to motor vehicle occupant injuries) and those 75 or older (mainly due to falls). Restricting the age range in the development of a rate is sometimes called an age-limited rate.

[Guideline: Age-Specific Rates](#)

Should I calculate a rate when the number of events is small?

Rates based on small numbers of events can fluctuate widely from year to year for reasons other than a true change in the underlying frequency of occurrence of the event. [Guidelines for Working with Small Numbers](#) provides approaches for assessing and addressing statistical stability.

[Guideline: Small Numbers](#)

Will the rate be compared to other rates?

When calculating rates, the numerator and denominator (i.e., events and population) must be defined consistently over time and place. Areas where public health professionals are most likely to find inconsistencies include:

- *The definition of the health event:* Because analysts may classify diseases differently, it is useful to compare technical definitions of the number of events before comparing rates. For example, *Healthy People 2020* defines coronary heart disease as International Classification of Diseases, 10th Revision (ICD-10) codes I11 and I20–I25; the National Center for Health Statistics' definition of coronary heart disease does not include I11. Likewise, the National Center for Health Statistics and the National Cancer Institute include use different code groups for several types of cancer.
- *The coding scheme:* Coding schemes that may appear to collect the same information can contain subtle differences. For example, starting with 1999 deaths, state and federal agencies changed from coding causes of death following the International Classification of Diseases (ICD) 9th Revision (ICD-9) to the ICD 10th Revision (ICD-10). Differences in coding practices between ICD-9 and ICD-10 affect mortality rates and trends for some disease categories, but not others.
- *Data collection and definitions:* Data collection processes or definitions for the number of events or the population at risk may change over time.
- *Geographies:* Data collection may vary from one geographic area to another. This issue is generally more relevant to the number of events than the population at risk and more relevant to some types of events than others.

In addition to the previous issues, when comparing age-adjusted rates, the standard population must be the same for all rates to be compared. Although most national, state and local agencies in the United States use the [U.S. 2000 Standard Population](#) for age adjustment, international agencies and older data developed in the United States may use different standard populations.

When comparing age-specific rates, if the age categories are relatively large, there can still be confounding by age. For example, if the proportions of very old individuals in the group "65 years and over" are different in two populations being compared, differences in rates may be a reflection of differences in the age distributions of the populations.

[Guideline: Comparing Rates](#)

How do I know whether two rates are different?

See [Guidelines for Using Confidence Intervals for Public Health Assessment](#).

Guidelines for Using and Developing Rates

Number of events (numerator)

- In developing a rate, only count events that occur among the population at risk.
- For conditions where one person may be counted more than once, note whether you are counting events or people. The count of events may be a better estimate of the public health burden experienced by the total population.
- When reporting rates, provide precise definitions of the number of events (e.g., specify ICD codes) so that readers can calculate comparable rates.

- When reporting rates, specify the data source for the number of events.

Population at risk (denominator)

- Everyone in the population at risk must be eligible to be counted in the numerator if they have the event of interest. For example, in looking at female breast cancer, we cannot include men in the population at risk, because men with breast cancer would not be included in the number of events.
- See [population denominators](#) for guidelines related to choice of populations at risk.
- When reporting rates, specify the data source for the population at risk.

Crude rates

- Crude rates are recommended when a summary measure is needed and it is not necessary or desirable to adjust for age.
- Choose a constant (e.g., rate per 1,000 or rate per 100,000) that is compatible with commonly published rates for the topic (e.g., birth information is generally expressed per 1,000 live births; death information is commonly expressed per 100,000 population.)

Age-adjusted rates

- Age-adjusted rates are recommended when comparing rates of age-related health events across different populations or within a given population over time, especially when comparing populations with **different** age distributions. These comparisons are only valid if the [direct method](#) for age adjustment has been used.
- Only use age-adjusted rates for the purpose of comparison. Because an age-adjusted rate is based on an external standard population, it does not reflect the absolute frequency of the event in a population.
- Present age-adjusted rates when a single, summary measure is needed. Choi 1999 recommends developing age-specific rates before developing age-adjusted rates as explained [above](#). However, this approach is not always practical in public health practice. Whether the data analyst follows this recommendation might depend on the purpose of the publication. For example, the Washington State Cancer Registry (WSCR) produces summary statistics for about 25 types of cancer annually. The purpose of the report is to provide an overview of cancer in Washington. One portion of the report presents age-adjusted incidence and death rates for the state and its 39 counties. Examining age-specific rates for this portion of the report alone could require hundreds of separate analyses. Age-specific analyses are more practical and important when investigating, for example, what a specific county looks like in relation to the state for a limited number of health conditions. If patterns by age in two or more jurisdictions or over time are not similar, presenting age-specific rates, rather than age-adjusted rates, is recommended.
- Use the U.S. 2000 Standard Population in current analyses, unless there is a need to compare to data that have been adjusted to another standard. In the latter case, you must use the standard population used in the comparison data. See [U.S. Standard Populations](#).

Age-specific rates

- Use age-specific rates to compare age-defined subgroups between or within populations when rates are strongly age-dependent.

- Use age-specific rates when causal or protective factors or risk exposures are different at different ages.
- Use age-specific rates to explore summary measures such as age-adjusted rates.
- In defining sub-groups for age-specific rates, select age ranges appropriate to the condition of interest.

Comparing rates

- Only compare rates when the numerator and denominator (i.e., events and population) are defined consistently over time and place. Look for
 - Consistency in definition of event
 - Consistency in coding scheme
 - Consistency over time
 - Consistency among geographies
- If comparing age-adjusted rates, compare rates that have been adjusted to the same standard population and age groupings.
- When comparing age-specific rates, if the age categories are relatively large, consider developing age-specific rates using smaller age groups or age adjust within the broader age group.

Unstable rates due to small numbers

Rates based on small numbers of events can fluctuate widely from year to year for reasons other than a true change in the underlying frequency of occurrence of the event. (See [Guidelines for Working with Small Numbers.](#))

- Calculation of rates is not recommended when there are fewer than five events in the numerator or a relative standard error of greater than 30%, because the calculated rate can be unstable and exhibit wide confidence intervals.
 - Annotating rates with relative standard errors between 30% and 50% as unreliable might offer an alternative to suppression.
 - Additional analysis looking at annual fluctuation can help the data analyst understand the stability of rates based on small numbers.
- Small counts should be included, where possible, even if the rates are not reported, so that the counts can be combined into larger totals (for example, three or five year averages) which would be more stable.

Methods for Age Adjustment

Direct adjustment

Multiply the age-specific rates in the target population by the proportion of the standard population in the corresponding age group.

$$\hat{R} = \sum_{i=1}^m s_i (d_i / P_i) = \sum_{i=1}^m w_i d_i$$

Where m is the number of age groups, d_i is the number of cases (events or people) in age group i , P_i is the population in age group i , and s_i is the proportion of the standard population in age group i . This is a weighted sum of Poisson random variables, with the weights being (s_i / P_i) .

U.S. Standard Populations for Direct Adjustment

In 1998, the federal Department of Health and Human Services (DHHS) directed all DHHS agencies to use the 2000 projected U.S. population as the standard for age adjustment. This standard is generally called the U.S. 2000 Standard Population. This directive improved consistency in age adjustment across DHHS agencies. Prior to the directive, some agencies used the U.S. 1940 and others used the U.S. 1970 Standard Populations. However, not all DHHS agencies use the U.S. 2000 Standard Population in the same manner. The inconsistency most likely to affect rates is the use of different age groupings. Less substantive are small differences in how proportions are calculated. The following table shows the age groupings and proportions used by the National Center for Health Statistics and the National Cancer Institute. These distributions are frequently used for public health statistics. Groupings and proportions for other age ranges and methods for creating proportions for unique groupings are available in [National Center for Health Statistics Statistical Notes 20 January 2001](#).

U.S. 2000 Standard Population: Proportions and Age Groups Used by the National Center for Health Statistics (NCHS) and the National Cancer Institute (NCI)

NCHS		NCI	
11, "10-yr" age groups		19, "5-yr" age groups	
age group	proportion	age group	proportion
<1	0.013818	<1	0.013818
1 – 4	0.055317	1 – 4	0.055316
5 – 14	0.145565	5 – 9	0.072532
15 – 24	0.138646	10 – 14	0.073031
25 – 34	0.135573	15 – 19	0.072167
35 – 44	0.162613	20 – 24	0.066478
45 – 54	0.134834	25 – 29	0.064530
55 – 64	0.087247	30 – 34	0.071045
65 – 74	0.066037	35 – 39	0.080762
75 – 84	0.044842	40 – 44	0.081852
85+	0.015508	45 – 49	0.072117
		50 – 54	0.062718
		55 – 59	0.048454
		60 – 64	0.038794
		65 – 69	0.034264
		70 – 74	0.031772
		75 – 79	0.026998
		80 – 84	0.017843
		85+	0.015509

Indirect Adjustment

When the number of events in a community is small, or when developing statistics for use in communities concerned about the number of events, compare the observed number of events to the expected number, using indirect age adjustment or age and sex adjustment.

- To develop a count of expected events using indirect adjustment, apply the age-specific rates in a larger population (e.g., Washington State) to the number of people in the age-specific group in the population of interest, and total the results for all age groups. Ideally, the larger population should be large enough that the rates in that population are stable (i.e., exhibit little random variation).
- Compare the observed number of events (usually abbreviated "O") and the expected number obtained using indirect adjustment (usually abbreviated "E").
 - Generally, if the confidence interval (CI) around O does not include E, the observed number of events is statistically significantly different from the expected. (Breslow and Day, 1987) The Poisson confidence limits around O can be obtained from standard tables or can be calculated using several software packages. This method assumes that E is developed from stable rates (i.e., the standard population is large enough that there is little random variation in the rates for that population).
 - A more conservative estimate of statistical significance (and the estimate that should be used when the rates in the larger population are not as stable as one would like) is to develop CIs around both E and O. If the CIs do not overlap, the difference between O and E is statistically significant. If they do overlap and either CI contains the point estimate of the other rate, O and E are not statistically significantly different. If the CIs overlap, but neither contains the point estimate of the other rate, a statistical test is needed to determine statistical significance.
- A more precise method of determining whether the observed number of events is different from the expected is to develop a ratio of O to E (O/E) and conduct a statistical test to determine whether the ratio is statistically significantly different from 1. Most biostatistics textbooks provide methods for conducting the statistical test. Fisher and van Belle (1993) provide an approximation for large samples only. Rosner (1990) provides methods for small and large samples.

Glossary

Rate: A rate is a measure of the frequency of an event per population unit. The use of rates, rather than raw numbers, is important for comparison among populations, since the number of events depends, in part, on the size of the population.

Numerator: In calculating rates, the numerator is the number of events in a specified population.

Denominator: In calculating rates, the denominator is the number of people in a specified population. Everyone in the denominator must be eligible to be counted in the numerator. The denominator is often called the "population at risk."

Crude rate: A crude rate is calculated by dividing the total number of events in a specified time period by the total number of individuals in the population who are at risk for these events and multiplying by a constant, such as 1,000 or 100,000 [e.g., (numerator/denominator) x constant].

Age Adjustment: Age adjustment is the process by which differences in the age composition of two or more populations are removed, to allow comparisons between these populations in the frequency with which an age-related health event occurs.

Age-adjusted rate (direct adjustment): An age-adjusted rate adjusted by the direct method is "...the rate that would occur if the observed age-specific...rates were present in a population with an age distribution equal to that of a standard population." (Anderson, 1998)

Age-specific or age-limited rate: An age-specific rate is a rate in which the number of events and population at risk are restricted to an age group (e.g., the birth rate for women age 15 to 19; death rate for people age 45 to 64).

Standard population: The standard population refers to the choice of populations used in developing age-adjusted rates.

References and Resources

[Policy Statement on Changing the Population Standard Used for Age Adjusting Death Rates in DHHS Publications](#)

Anderson RN, Rosenberg HM. Age standardization of death rates: Implementation of the Year 2000 Standard. *Natl Vital Stat Rep.* 1998;47(3). http://www.cdc.gov/nchs/data/nvsr/nvsr47/nvs47_03.pdf. Accessed May 1, 2012.

Breslow NE, Day NE. *Statistical Methods in Cancer Research: Volume II – The Design and Analysis of Cohort Studies*. New York: Oxford University Press, 1987.

Choi BCK, de Guia NA, Walsh P. Look before you leap: Stratify before you standardize. *Am J Epidemiol.* 1999;149:1087-1096.

Fay MP, Feuer EJ. Confidence intervals for directly standardized rates: A method based on the gamma distribution. *Stat Med.* 1997;16:791-801.

Fischer LD, van Belle G. *Biostatistics: A Methodology for the Health Sciences*. New York: John Wiley and Sons, Inc., 1993.

Klein RJ, Schoenborn CA. Age adjustment using the 2000 projected U.S. population. *Healthy People 2010 Stat Notes 20*, January 2001. National Center for Health Statistics, 2001.

Kuller LH. Age-adjusted death rates: A hazard to epidemiology? [editorial] *Ann Epidemiol.* 1999;9(2):91-92.

Last JM [ed]. *Dictionary of Epidemiology*. 3rd ed. New York: Oxford; 1995.

Rosner, BA. *Fundamentals of Biostatistics*. 3rd ed. Boston: PWS-Kent Publishing; 1990.

Selvin S. *Statistical Analysis of Epidemiologic Data*. 2nd ed. New York: Oxford University Press; 1996.

Sorlie PD, Thom TJ, Manolo T, Rosenberg HM, Anderson RN and Burke GL. Age-adjusted death rates: Consequences of the year 2000 standard. *Ann Epidemiol.* 1999;9:93-100.