

# About Age-Adjusted Rates, 95 Percent Confidence Intervals and Unstable Rates

## What is age-adjustment?

Age-adjustment is a statistical process applied to rates of disease, death, injuries or other health outcomes that allows communities with different age structures to be compared.

## Why do we do age-adjustment?

Almost all diseases or health outcomes occur at different times in different age groups. Most chronic disease, including most cancers, occurs more often among older people. Other outcomes, such as many types of injuries, occur more often among younger people. The age distribution determines what the most common health problems in a community will be. One way of examining the pattern of health outcomes in communities of different sizes is to calculate an incidence or mortality rate, which is the number of new cases or deaths divided by the size of the population. In chronic diseases and injuries, rates are usually expressed in terms of the number of cases/deaths per 100,000 people per year.

A community made up of more families with young children will have a higher rate of bicycle injuries than a community with few young children. A community with a larger number of older individuals will have higher rates of cancer than one with younger individuals. This is true even if the individuals in the two communities have the same risk of developing cancer or being injured. Epidemiologists refer to this as confounding. Confounding happens when the measurement of the association between the exposure and the disease is mixed up with the effects of some extraneous factor (a confounding variable). Age-adjustment is a statistical way to remove confounding caused by age.

## How is age-adjustment done?

Age confounding occurs when the two populations being compared have different age distributions and the risk of the disease or outcome varies across the age groups. The process of age-adjustment by the direct method changes the amount that each age group contributed to the overall rate in each community, so that the overall rates are based on the same age structure. Rates that are based on the same age distribution can be compared to each other without the presence of confounding by age. Adjustment is accomplished by first multiplying the age-specific rates of disease by age-specific weights. The weights used in the age-adjustment of cancer data are the proportion of the standard U.S. population within each age group. The weighted rates are then summed across the age groups to give the age-adjusted rate.

## What is a confidence interval?

A confidence interval is a range around a measurement that conveys how precise the measurement is. Ninety five percent confidence intervals are provided for the age-adjusted rates in data released by the North Dakota Statewide Cancer Registry. These are analogous to the margins of error that are provided for news polls.

In the simplest terms, an age-adjusted breast cancer incidence rate of 132.6 cases per 100,000 women +/- 1.0 case per 100,000 means that there is a 95 percent chance that the rate was between 131.6 and 133.6 cases per 100,000 women. Conversely, there is a 6 percent chance that the rate was lower than 131.6 or higher than 133.6.

The precise statistical definition of the 95 percent confidence interval is that if the measurement were

conducted 100 times, 95 times the true value would be within the calculated confidence interval and five times the true value would be either higher or lower than the range of the confidence interval.

## What are confidence intervals used for?

Confidence intervals often are used in research that requires that a measurement be taken on a sample of the population, such as a survey. In that case, the confidence intervals are a way to measure sampling error and are related to the size of the sample taken or the number of people surveyed. The cancer incidence and mortality rates derived from the Cancer Registry and the Division of Vital Records are not samples and are, therefore, not subject to sampling error. The rates are, however, subject to what is termed random error, which arises from random fluctuations in the number of cases over time or between different communities. The 95 percent confidence intervals are an easily understood way to convey the stability of the rates. A stable rate is one that would be close to the same value if the measurement were repeated; e.g., if the rate did not vary greatly from one year to the next. An unstable rate is one that would vary from one year to the next due to change alone. Wider confidence intervals in relation to the rate itself indicate instability. For example, if a rate is five cases per 100,000 people, but the 95 percent confidence interval is plus or minus 2.5 cases per 100,000 people, then the rate is relatively unstable. In one year, you might have a rate of three cases per 100,000 and in the next year have six cases per 100,000. This would be a 100 percent increase in the rate but would still be within the range of the confidence intervals.

On the other hand, narrow confidence intervals in relation to the rate tell you that the rate is relatively stable and you would not expect to see large fluctuations from year to year. If differences are observed between stable rates (those with narrow confidence intervals), then it is likely that the differences represent true variations, rather than random fluctuations in the number of cases.

## How are confidence intervals calculated?

Confidence intervals are calculated based on the standard error of the rate. The standard error, in turn, is based on the rate and the number of cases or deaths. For crude rates, the formula for the 95 percent confidence intervals is as follows:

$$\begin{aligned} 95\% \text{ CI} &= \pm 1.96 \times \text{se} \\ &= \pm 1.96 \times \text{rate} \div \sqrt{\text{cases}} \end{aligned}$$

## Why are rates based on fewer than 20 cases marked as being unstable?

Annual rates based on fewer than 20 cases or five-year average rates based on fewer than four cases per year are considered unstable because they have a large relative standard error (RSE). The RSE is the standard error as a percentage of the measure itself. For age-adjusted incidence and mortality rates, the RSE is equal to  $1 \div \sqrt{\text{cases}}$ . A RSE of 50 percent indicates that the standard error is half the size of the rate.

The RSE of an incidence or mortality rate is based on the number of cases or deaths, unlike the standard error and confidence intervals, which are based on both the number of cases and the size of the population.

Example:

If there are 20 testicular cancer deaths among males every year, the rate of testicular cancer deaths is 0.4 per 100,000 males, and the standard error is 0.09. The relative standard error is 22 percent.

If there are about 20 prostate cancer deaths in every year, the rate is 21.9 deaths per 100,000 males and the standard error is 4.8. Again, the relative standard error is 22 percent.

So, even though the standard error for the testicular cancer rate is much smaller than the standard error for the prostate cancer rate, both have the same magnitude relative to the rate itself. This means that when

there are only a few cases or deaths, small changes in the number of cases or deaths have a much bigger effect than small changes in a large number of cases or deaths.