

## Glossary – data online tool

**Age-adjusted (age-standardized) rate:** A rate statistically modified to eliminate the effect of different age distributions in different populations. It allows one to compare rates between regions that have different age distributions (a region with a larger senior population will have higher disease rates simply because they have more people who are likely to have a disease).

**Age-specific rate:** A rate limited to a particular age group. The numerator is the number of deaths or people with condition/disease in that age group; the denominator is the number of persons in that age group in the population. For example, the rate of diabetes for people aged 55-65 years.

**BBTD:** Baby Bottle Tooth Decay

**BMI (see Body Mass Index)**

**Birth weight (high/low):** A birth weight is considered “high” if it is equal to or greater than 4.0 kg (eight pounds 12 ounces) or “low” if less than or equal to 2.5 kg (five pounds seven ounces).

**Body Mass Index (BMI):** a measure of body weight and height used to classify people as: underweight (BMI below 18.5), normal weight (BMI 18.5-24.9), overweight (BMI 25-29.9), obese (BMI 30 and above).

One big criticism of using BMI is that it does not distinguish between body fat and muscle mass (which weighs more than fat).

**Chi-square test:** Chi-square is a statistical test commonly used to compare observed data with data we would expect to obtain according to a specific hypothesis. For example, if, according to Mendel's laws, you expected 10 of 20 offspring from a cross to be male and the actual observed number was 8 males, then you might want to know about the "goodness to fit" between the observed and expected. Were the deviations (differences between observed and expected) the result of chance, or were they due to other factors. How much deviation can occur before you, the investigator, must conclude that something other than chance is at work, causing the observed to differ from the expected. The chi-square test is always testing what scientists call the null hypothesis, which states that there is no significant difference between the expected and observed result.

**Coefficient of variation (CV):** Used to assess variation in an estimate, the CV is the ratio of the standard deviation (SD) to the estimate (e.g mean or proportion). Estimates

## Glossary – data online tool

with CV between 16.5% and 33.3%, reflect moderate to high sampling variability and are supplemented with an 'E' to indicate cautious interpretation. Estimates with a CV greater than 33.3%, reflect extreme sampling variability and are suppressed (denoted by an 'F').

**Confidence interval (CI):** Sometimes referred to as the 'plus or minus' around a statistic. The CI represents a range of possible values for the estimate, within a certain probability. The specified probability is called the confidence level (usually 95%), and the end points of the confidence interval are called the confidence limits.

**Confidence limit:** The minimum or maximum value of a confidence interval. Using the above example of  $120 \pm 39.2$ , the confidence limits are 80.8 lb ( $120-39.2$ ) and 159.2 lb ( $120+39.2$ ). Therefore 95% of the data fall between the ranges of 80.8 lb and 159.2lb, with a mean of 120lb.

**Crude rate:** The number of events in a given time period divided by the population at risk produces crude rates. The result is multiplied by a constant (typically 1,000 or 100,000) for ease of presentation. Common crude rates include birth, death, marriage, and divorce.

**Denominator:** The lower portion of a fraction used to calculate a rate, proportion or ratio. In a proportion, the denominator is usually the population. In a rate, the denominator is expressed as population experience (e.g. person-years) at risk. For example:  $\frac{1}{2}$  - "2" is the denominator

**Distribution:** In epidemiology, the frequency and pattern of characteristics and events in a population. In statistics, the observed or theoretical frequency of values of a variable.

**E:** Estimates with CV between 16.5% and 33.3%, reflect moderate to high sampling variability and are supplemented with an 'E' to indicate cautious interpretation.

**Frequency:** The number of times something occurs over a particular point or period in time. For example, the number of car accidents in the past year: 'the frequency of car accidents in Sackville was 64 in the past year'.

**F:** Estimates are suppressed due to low cell size ( $n < 5$ ) or very high sampling variability ( $CV > 33.3\%$ ).

## Glossary – data online tool

**Mean:** The measure of central location commonly called “the average”. It is calculated by adding together all the individual values in a group of measurements and dividing by the number of values in the group.

**Median:** The measure of central location which divides a set of data into two equal parts. When the data is sorted lowest to highest, it is the central number; half of the data is higher and half of the data is lower.

‘n’: generally refers to the sample size (i.e. how many people answered a question).

‘N’: (capital n) generally refers to the total population.

**Percentage:** A fraction or ratio with 100 understood as the denominator; for example, 0.98 equals a percentage of 98 or 98 percent (%).

**Physical activity:** Based on total energy expenditure (EE), which is calculated using the following formula:

$$EE = \sum(N_i * D_i * MET_i / 365 \text{ days})$$

$N_i$  = number of occasions of activity  $i$  in a year,

$D_i$  = average duration in hours of activity  $i$ , and

$MET_i$  = a constant value for the metabolic energy cost of activity  $i$ .

Frequency and duration of physical activities were reported for the 12 months prior to the survey, and the metabolic equivalent value (MET value) of each activity was independently established (Ainsworth et al., 2000). Children with energy expenditures of less than 1.5 kcal/kg/day were considered to be inactive; those with energy expenditures between 1.5 kcal/kg/day and 2.9 kcal/kg/day were considered to be moderately active; and those with energy expenditures of 3 kcal/kg/day or greater were considered to be active.

**Prevalence:** The proportion of a population found to have a condition (typically a disease or a risk factor such as smoking or seat-belt use). It is arrived at by comparing the number of people found to have the condition with the total number of people studied, and is usually expressed as a fraction, as a percentage or as the number of cases per 10,000 or 100,000 people.

## Glossary – data online tool

**Proportion:** A type of ratio in which the numerator is included in the denominator. The ratio of a part to the whole, expressed as a "decimal fraction" (e.g., 0.2), as a fraction (1/5), or, loosely, as a percentage (20%).

**Sex-specific rate:** A specific rate among either males or females.

**Standard deviation (SD):** A measure of variation away from the average. It is the most widely used measure of dispersion of a frequency distribution, equal to the positive square root of the variance. A large standard deviation means that the data points are spread out, whereas a small standard deviation means the data points are clustered around the mean (i.e. a tighter estimate). One standard deviation away from the mean captures 68% of the data, and two standard deviations away from the mean represents 95% of the data.

**Standard error:** The standard deviation of a theoretical distribution of sample estimates about the true population estimate.

**Statistical significance (or difference):** The difference between groups or categories is considered statistically significant if the 95% confidence interval for each estimate does not overlap, after Bonferonni adjustment. Confidence intervals were reported using either a range (e.g., 95% CI: 87.5, 91.5) or a plus/minus (e.g., 95% CI: +/- 2.0).

**Variance:** A measure of the dispersion shown by a set of observations, defined by the sum of the squares of deviations from the mean, divided by the number of degrees of freedom in the set of observations.