COMMON DATA TERMS

CI:NOW

Table of Contents

CI:NOW

-A-

Administrative data: data generated in the everyday course of business, like sales data in a grocery store, attendance data in a school, or diagnosis data in a doctor's office. Administrative data is a type of **secondary data**.

Age-adjusted rate: rate that has taken into account influences on a crude rate, such as differences in age composition of the population. Age adjustment, using the direct method, is the application of age-specific rates in a population of interest to a standardized age distribution in order to eliminate differences in observed rates that result from age differences in the population composition. This adjustment is usually done when comparing two or more populations (such as race/ethnic groups) at one point in time or one population at two or more points in time. Age-adjusted rates are useful for comparison purposes only, not to measure absolute magnitude. (To compare absolute magnitude, numbers or crude rates are used.)

Age-specific rate: rate obtained for specific age groups (for example, age-specific fertility rate, death rate, marriage rate, illiteracy rate, school enrollment rate, etc.).

Aggregate data: individual data records that have been "rolled up" to a summary level. Data can be aggregated in many different ways, but we often see aggregation by geography, like zip code, or by some characteristic like race/ethnicity or age group.

AISP: Actionable Intelligence for Social Policy is an initiative housed at the University of Pennsylvania that "focuses on the development, use, and innovation of integrated data systems (IDS) for policy analysis and program reform." AISP focuses specifically on IDS, not community data in general. See Integrated Data Systems.

-B-

Big data: a buzzword and a complicated concept, but the term is generally intended to mean datasets that are so large or complex that they can't be handled – managed, analyzed, stored, transferred – using normal data tools. "Big" data typically means petabytes of data (1,024 terabytes, where a terabyte is 1,024 gigabytes [GB]) or exabytes (1,024 petabytes) of data. By definition, community indicator data – for example, percent of households in poverty – and any other data that can be worked with in Excel, Filemaker, Access, or in a similar tool isn't big data.

-C-

Clinical Classification Software: coding system developed by the Agency for Healthcare Research and Quality that classifies ICD-9 codes into disease classifications.

Community Indicators Consortium (CIC): CIC is an organization that offers resources and tools to help communities and practitioners advance the practice and effective use of community indicators to improve quality of life. CIC focuses specifically on community indicators rather than on community data and information systems in general.

Crude rate: rate of any demographic or vital event that is based on an entire population.

-D-

Dashboard: a buzzword that has come to mean some kind of visual display of current (and sometimes historical) values for a set of indicators.

Data: broad concept that generally means a collection of values or pieces of information; sometimes called a dataset. Among other characteristics, data may be quantitative (numerical) or qualitative (non-numerical, like words or images), raw or processed, record-level or aggregated (grouped), and primary (collected/created for the purpose of answering a question) or secondary (created for

some other purpose). "Data" and "indicators" are not the same thing; indicators are calculated from data.

Demography: study of populations including their size, age-sex composition, distribution, density, growth, natality, mortality, marriage/divorce, migration, and any other characteristics which may affect these factors.

Denominator: number below the line in a common fraction.

-E-

Ethnicity: classification of a population that shares common characteristics, such as, religion, traditions, culture, language, and tribal or national origin.

Extant data: data that already exists, also called **secondary data.** Two kinds of extant data are survey data that was collected for some other purpose and **administrative data** generated in the course of an organization's everyday operations.

-F-

Fertility rate: number of live births, regardless of age of mother, per 1,000 women of reproductive age, 15-44 years.

-H-

Health information exchange: form and purpose can differ, but in general, health information exchange (HIE) refers to the electronic transfer of health-related information among organizations, and "HIE" is often understood to refer to a central database of health-related information about a large number of patients, and also to the organization that assembles and manages that data.

-I-

ICD-10: <u>International Classification of Diseases</u>, 10th edition. A system for classifying diseases and injuries developed by the World Health Organization and used worldwide to improve comparability of cause of death statistics reported from different countries.

Indicator: general term for a thing that tells us the state or level of something. "Four-year graduation rate" tells us something about how well kids in a high school do, and "temperature" tells us something about how hot or cold it is. An indicator isn't necessarily a *good* indicator. Often used interchangeably with measure. "Indicator" is not synonymous with "data;" indicators are calculated from data.

Integrated data system: link records across datasets, usually from schools and other human service agencies, to assemble a more complete data "picture" of individual people and/or families. Can vary widely in purpose, topic, size, and functionality. See also **health information exchange** and **administrative data.**

-L-

Life expectancy: average number of years that a person can anticipate living after a given age, usually birth. Most often based upon the current mortality experience of a population.

-M-

Margin of error: when we can't measure all of something, like people in a city, we sample them – measure only some to get an idea of (estimate) what's true of everyone. Sampling introduces error and uncertainty, and the margin of error – for example, "plus or minus three percentage points" – is a measure of how much uncertainty there is. The smaller the sample in relation to the total population, generally, the larger the margin of error.

Mean: arithmetic average of a set of values. It is calculated as the sum of the values divided by the number of values.

Median: value in an ordered set of values above and below which there are an equal number of values; the 50th percentile.

Morbidity: refers to the occurrence of diseases in a population.

Mortality: death as a component of population change.

-N-

Natality: birth as a component of population change.

NNIP: National Neighborhood Indicators Partnership is "a collaborative effort by the Urban Institute and local partners to further the development and use of neighborhood information systems in local policymaking and community building." NNIP defines "information systems" very broadly.

Numerator: number above the line in a common fraction.

-O-

Open data: defined by the Open Knowledge International as data that anyone is "free to use, reuse, and redistribute – subject only, at most, to the requirement to attribute and/or share-alike."

-P-

Percentage point increase/decrease: one way of describing the difference between your current measurement and a past measurement, without relating the change to the past measurement. It's just the difference between the two values, and it's usually phrased as "decrease of X percentage points." If the percent of the population that smokes cigarettes decreased from 19% in 2014 to 17%

in 2015, you'd have a two percentage point decrease, because the difference between 19 and 17 is two. See also **percent increase/decrease**.

Percent increase/decrease: one way of describing the difference between your current measurement and a past measurement, relating it to the past measurement. The percent change is the difference between the two values, divided by the past value, and it's usually phrased like "percent decrease from prior year" or "percent increase over prior year." For example, if the percent of the population that smokes cigarettes decreased from 19% in 2014 to 17% in 2015, you'd have a 10.5% (percent) decrease, because the difference between 19 and 17 is two, and two divided by 19 is 10.5%. See also **percentage point increase/decrease**.

Population: total of all individuals in a given area.

Proportion: portion of a population in relation to another portion of the population or to the population as a whole. Proportions are a special type of ratio in which the denominator always includes the numerator. (See also ratio.)

-R-

Race: geographical population of humankind that possesses inherited distinctive physical characteristics that distinguish it from other populations.

Range: distance between the smallest and largest numbers in a set of numbers.

Rate: amount of one thing in relation to a unit of a related thing in a time period, often the frequency of an event in a specified period of time divided by the population at risk of the event. An example would be "31 violent crimes per 1,000 population in 2015." A percent is a rate where the denominator is always 100.

Ratio: relation of one population subgroup to another subgroup, or to the whole population. The denominator of a ratio may or may not include the numerator. If the denominator includes the numerator, it is a special type of ratio known as a proportion. (See also proportion.)

Residence data: data compiled by the usual place of residence without regard to the geographic location where the event occurred. For births and fetal deaths, the mother's usual residence is used as the place of residence.

-S-

Secondary data: secondary data is existing quantitative data that has already been collected by someone else, likely for some purpose different from yours. Also called **extant data.**

Statistical cut-off: date by which records of vital events for a specific year must be received in order to be included in the statistical analyses for that year.

Statistical significance: used to evaluate the likelihood that chance variability may be considered an explanation for observed results. An appropriate mathematical test of statistical significance is calculated to determine the p value, which is the probability that the observed results may be due to chance alone. If the p value is less than an arbitrarily chosen value, commonly selected as 0.05, the findings are accepted as statistically significant at the 5 percent level. This indicates there is less than 5 percent probability that the observed results are due to chance alone.

-V-

Vital statistics: demographic data on abortions, births, deaths, fetal deaths, marriages and divorces.

-Y-

Years of potential life lost (YPLL75) rate: measure of premature mortality for a specific population, like residents of a county. YPLL is the sum of all the years of life "lost" by individuals in that population

who died before age 75. A person who died at age 60 would contribute 15 years to the population's YPLL, a person who died at age 48 would contribute 27 years, and a person who died at 75 or older would contribute zero. The YPLL rate is the total YPLL divided by the number of people in the population, usually multiplied by 1,000 to make the number easier to read.

A Note on Sources

These definitions were developed locally to serve as a non-technical resource for people interested in data terms. Two sources used to develop an earlier version of this document are:

1. Shryock, H.S., and Siegel, J.S. *The Methods and Materials of Demography*. San Diego, CA: Academic Press, 1976.
2. Arthur Haupt and Thomas T. Kane, *Population Handbook* (Washington, D.C.: Population Reference Bureau, Inc., 1978), p. 54.

Additional Information

Below are a few of the many free resources available online for those who would like to learn more about data, from the basics to advanced concepts and skills.

1. School of Data. https://schoolofdata.org/handbook/courses/what-is-data/
2. Data-Pop Alliance. http://datapopalliance.org/item/what-is-data-literacy/
3. Oceans of Data Institute. http://oceansofdata.org/our-work/big-data-big-promise

If you'd like to suggest additions or changes to this list, please contact CI:Now via our website at http://cinow.info/contact/.

Last rev. 3/2017