C

APPENDIX C

# OVERVIEW OF TERMS AND CONCEPTS USED IN HEALTH RESEARCH AND EPIDEMIOLOGIC STUDIES

Epidemiology is defined as the study of the distribution and determinants of disease frequency in human populations and the application of this study to control health problems (Aschengrau and Seage, 2008). This appendix provides an overview of terms and concepts used in health research that you may encounter as you conduct your community health needs assessments and develop your implementation strategies.

## *Data quality*

Evaluate the reliability and validity of data that will be used in the community health needs assessment.

*Reliability –* Reliability refers to consistency of measurements. A reliable measure will give identical or nearly identical values when measuring the same thing over time.

*Validity –* Validity refers to accuracy of measurements. A valid measure accurately measures what it is intended to measure.

Example: A woman is conducting a study on obesity. She weighs each study participant twice, and notes that the weights are identical. However, the scale is improperly calibrated, and adds five pounds to each person's weight. The scale is a reliable instrument, but it is not a valid instrument. The data is consistent, but is not accurate.

## *Measures of disease: prevalence and incidence*

*The prevalence* of a disease is the number of existing cases at a specific point in time, while *incidence* reflects the number of new cases that arise within a given time period.

Here is an example that illustrates the difference between prevalence and incidence measures:

A team of researchers are interested in studying asthma in a particular community. In 2010, the research team determines how many people living in the community have been diagnosed with asthma - these are the *prevalent* cases of asthma. In other words, this is the prevalence of asthma in 2010 for this particular community.

C

The research team decides to proceed with the asthma study. In 2011, they enroll the members of the community who have *not* been diagnosed with asthma in their study. The research team monitors this group of people over two years. Because the research team is following the population over time, they know when a new case of asthma - an *incident* case - occurs. Therefore, the number of cases that occur during the study is the asthma incidence rate for this particular community between 2010 and 2011. It is important to note that the existing cases of asthma at the beginning of the study (the prevalent cases) are not included in the incidence measure.

### *Types of disease measurement: counts, proportions, ratios and rates*

In common usage, a proportion, ratio, and true rate are all referred to as rates; however, they are different measures. Proportions, ratios, and rates all contain a denominator, and are useful for comparison purposes. *Rates* are different from proportions and ratios because a unit of time (i.e., minutes, years) is included in the denominator. *Counts* simply list the number of events (births, deaths, disease) and are most informative when they are given in the context of the time period in which they occurred and with a geographic context.

A *proportion* is a division of two related numbers (Aschengrau and Seage, 2008). In a proportion, the numerator is included in the denominator. A percentage is a common example of a proportion. For example, the number of babies born at a certain hospital weighing less than 2,500 grams expressed over the total live births that occurred at the hospital would provide the proportion of low birth weight babies in a specified community over a specified time period.

A *ratio* is a division of two unrelated numbers (Aschengrau and Seage, 2008). The numerator is not included in the denominator in a ratio. An example of a ratio would be the number of doctors in a community divided by the number of hospital beds in the same community.

A *rate* is a division of two numbers and, as was previously stated, time is an integral part of the denominator (Aschengrau and Seage, 2008). A commonly encountered rate is miles per hour. In health-related data, an example of a rate would be 700 new cases of asthma per 100,000 population from 2011 to 2013. This is a rate because the time period, "from 2011 to 2013," is specified.

To monitor the health and well-being of a community, it is often desirable to compare a measure of disease from the community to that of another community. Moreover, it may be informative to compare a measure of disease from the community of interest to the number of cases or rate of disease at the national or state levels. Care must be taken when making such comparisons. Crude counts and rates from two different populations can rarely be accurately compared because their underlying population structures (size and age) are rarely the same. Proportions, ratios and rates are measures of disease that are better suited for comparing two populations and often these measures must be adjusted to make completely accurate comparisons.

Consider the two populations described below:

| | COUNTY A | COUNTY B |
|---|---|---|
| Number of new cases of disease X in 2009 | 125 | 300 |
| Population size | 1,500 | 10,000 |

C

First, it is important to note that the number of cases of disease X in 2014 represent the incident cases. County B has more than double the number of incident cases than County A and so it is tempting to conclude that it has a higher incidence of disease X; however, their differing population sizes have not been taken into account. The number of incident cases must be examined in the context of the population size from which they arose to permit an accurate comparison. In order to take the population sizes into account, divide the number of incident cases by the population size of each county and multiply this fraction by the conversion factor, "100,000 population." These numbers are now comparable and it can be concluded that County A had a greater incidence of disease than county B in 2014.

| | COUNTY A | COUNTY B |
|---|---|---|
| Number of new cases of disease X in 2009 | 125 | 300 |
| Population size | 1,500 | 10,000 |
| Incidence of Disease X | 8,333 cases of disease X per 100,000 population | 3,000 cases of disease X per 100,000 population |

C

Please note that it is most common to express these types of incidence measures in terms of "100,000 population" because larger numbers are easier to conceptualize and work with; however, if the original fraction (number of incident cases divided by population size) is quite large, a conversion factor of "1,000 population" can be used. Drawing on the same example, the equivalent incidence measures using the smaller conversion factor are 83.3 cases per 1,000 population in county A and 30 cases per 1,000 in county B.

Population size is not the only thing that varies between communities; the age structures may also be different. As with population size, the age structure must be taken into account in order to accurately compare disease measurements between two populations. Age is very related to disease occurrence; most diseases and health outcomes occur at different rates within different age groups. For example, even without looking at data from a specific population, it is intuitive that older people are more likely to suffer from arthritis than younger people. Therefore, a community with a larger proportion of older adults will have a higher overall incidence of arthritis compared to a community with a higher proportion of young people, such as a university town.

In order to accurately compare populations with different age structures, disease measures must be age-adjusted to a standard population. Any population can be used as a standard population, and the U.S. population is often used.

Age adjustment can be used to answer the following question, "How many cases of disease will occur if the standard population experiences the same incidence (or rate) of disease as the community of interest?"

To do this, the age-structure of the standard population must be known; this is the number of people in designated age groups (e.g., the number of people aged 0 to 5 years, 6 to 10 years, 11 to 15 years, etc.). Also, the incidence or rate of disease for these age groups must be known from the communities of interest; these are called age-specific rates. Typically, such rates are obtained in the same manner as the incidences of counties A and B for disease X in 2014 that were obtained above, but for each age group; therefore, the incident cases and numbers of people in each age group must be known for the populations of interest. These age-specific rates are applied to a standard population distribution, and the number obtained represents the number of cases of disease *if the standard population had the same incidence (or rate) of disease as the communities of interest*. If two populations are age-adjusted to the same standard populations, the numbers can be compared directly. This method is known as direct age standardization – it is a method for controlling the differences in population age structure when comparing two different populations.

### *Confidence intervals*

Confidence intervals indicate the reliability of a measurement. Most often 90 percent or 95 percent confidence intervals are presented around disease measurements. A 95 percent confidence interval is interpreted as follows: If the data collection and analysis could be replicated many times, the confidence interval should include the true value of the measurement 95 percent of the time (Rothman, 2002).

### *Trends*

A change in value for an indicator from one measurement period to another does not always signify a trend. When considering estimates from two time points, examine the confidence intervals. If the confidence intervals overlap, the trend is not likely to be significant. In order to determine if a true increase or decrease has occurred in the measure of disease between two measurement periods, it is necessary to conduct a statistical test for trends.

### *Stability*

If a measurement is based on a small number of disease counts, there is a high level of uncertainty in the measure and the value is considered unstable or unreliable. This is indicated by a broad confidence interval. The stability of the measurement can be improved by combining several years of data or by combining geographies. For example, instead of reporting the incidence of oral cancer in 2012, report the incidence of oral cancer from 2008 to 2012. There will be more cases of oral cancer over five years than in one year and the number will be more stable. Moreover, instead of considering the rate of disease for one ZIP code in a city, consider the rate of disease for a region in the city comprising 20 ZIP codes. Increasing stability in these ways will reduce the uncertainty surrounding the estimate and the confidence interval will become narrower.

### *Error*

Errors can occur in health research and epidemiologic studies that distort the true measurement of disease in a population or that distort the true relationship between an exposure and a disease. Errors come in two types: random error and nonrandom error (also known as bias).

Random error can occur chiefly in two ways, through sampling variability and measurement errors. Nonrandom error can occur when there are systematic differences in the study sample (selection bias), systematic differences in measurement of exposure or outcome (information bias) or confounding (a third variable either makes it appear as if there is a relationship between an exposure and a disease when there is not a true relationship or vice versa). Measurement error can also be nonrandom error (bias) in the case that it causes misclassification of the exposure or disease in a systematic way (i.e., it is not due to chance). For example, a woman is conducting a study on obesity. She weighs each study participant twice, and notes that the weights are identical. However, the scale is improperly calibrated, and adds five pounds to each person's weight each time. These are biased measurements. It is important to consider the possibility of error in your study measurements and data sources.



*For additional information on the topics covered in this section, the following resources are available:*

Aschengrau, Ann and George R. Seage III. *Essentials of Epidemiology in Public Health*. Sudbury: Jones and Bartlett, 2008.

Gordis, Leon. *Epidemiology*. 4th edition. Philadelphia: Saunders, 2009.

Jewell, Nicholas P. *Statistics for Epidemiology*. Boca Raton: Chapman & Hall/CRC, 2004.

Rothman, Kenneth J. *Epidemiology: An Introduction*. New York: Oxford University Press, 2002.