

## MEASURES OF DISEASE IN EPIDEMIOLOGY

The primary purpose of this exercise is to illustrate some methods of measuring the occurrence and outcomes of disease in populations. These methods are commonly used in describing the effects of disease in the population (descriptive epidemiology) and in investigations that test hypotheses about disease occurrence (analytic epidemiology).

The importance of the calculations and appropriate use of measures of disease (proportions and rates) will be demonstrated by a variety of examples chosen from infectious and non-infectious disease studies. The definitions, applications and limitations of the principal rates used in epidemiology are also summarized in this exercise.

### Types and Uses of Measures of Disease

The magnitude of disease can be measured as the frequency of illness within specific populations. Time and place must always be specified.

Proportions appear in the following form:  $\text{Proportion} = \frac{X}{Y} \times k$

where  $x$  = number of times an event has occurred during a specific interval of time.

$y$  = number of persons exposed to the risk of the event during the same time interval.

$k$  = some round number (100; 1,000; 10,000; 100,000; etc) or base, depending on the relative magnitude of  $x$  and  $y$ .

Proportions are usually expressed as either crude proportions or specific proportions. Crude proportions are concerned with the total number of events occurring in a defined population during a specified interval of time. Specific proportions are expressed as the number of events of a disease occurring in a specifically defined population (e.g., age, race, and sex) during a specified interval of time.

### *Examples:*

$$\text{Crude proportion} = \frac{\text{total \# myocardial infarctions (MI)}}{\text{total population of US males}} \times 1,000 \text{ per year}$$

$$\text{Specific proportion} = \frac{\text{\# white male MI's ages 45 - 54}}{\text{\# white males in population ages 45 - 54}} \times 1,000 \text{ per year}$$

Note that both proportions above are expressed per 1000 per year.

### Incidence

Incidence is a measure of the number of NEW cases of a disease in a population during a specified period of time. Incidence measures the risk of developing a disease.

$$\text{Cumulative Incidence} = \frac{\# \text{ NEW cases of disease over specified time period}}{\text{population at risk during specified time period}} \times 100,000$$

Note that in the above example incidence is expressed per 100,000 for a specified time period.

Only new cases (fatal as well as non-fatal) are included in the numerator. The denominator includes an estimate of the population at risk (i.e. persons without the disease) at the beginning of the time period during which incidence is to be measured. It is a fundamental principle of incidence that all persons included in the denominator must be at risk of being included in the numerator (of becoming a "case"), and that all of those included in the numerator must also be included in the denominator.

### **Question 1.**

Among 1,950 persons at risk (all ages) at an initial examination of civil service employees (persons free of coronary heart disease), 20 new cases of coronary heart disease occurred during the subsequent one-year period. What is the incidence of coronary heart disease in this sample?

### **Question 2.**

Among 1,000 males at risk in the 40-59 year age group in the civil service sample: 100 had initial serum cholesterol levels of less than 220 mg/dL, 600 had levels of 220-270 mg/dL and 300 had levels greater than 270 mg/dL. Among these three cholesterol classes, (1) less than 220; (2) 220-270; and (3) greater than 270, the number of new cases of coronary heart disease during the subsequent year were 0, 5, and 5 respectively. What was the incidence of coronary heart disease within each cholesterol class?

### **Question 3.**

In City A, 150 new cases of disease X were diagnosed among women aged 30-39 between January 1999 and December 2001. The census of 2000 for City A reported 10,000 women in that age group. What was the yearly incidence of disease X among women aged 30 - 39? (Assume the size of the population was constant over the three years.)

### Person-Time Incidence

The person-time incidence is the number of new cases that occur per unit of population-time. The denominator reflects the person-time experience of the population at risk. For example, a sample of 5 individuals who were each followed for 2 years would result in 10 person-years at risk. Another sample of 5 individuals, 3 of whom were followed for 1 year each, and 2 of whom were followed for 3 years each, would contribute 9 person-years at risk. Person-time incidence is commonly distinguished from the cumulative incidence which uses a count denominator of individuals rather than a person-time denominator.

#### **Question 4.**

In a health district, a group of 80 school children who lacked antibody to measles were followed for a total of 125 person-months for infection with measles. At the end of follow-up, 10 children had acquired antibody to measles. What was the person-time incidence rate for measles infection in this population of children?

#### **Prevalence**

**Prevalence** is a measure of the total number of people in a defined population who already have a specified disease at a given point in time. Prevalence measures the amount of disease existing in a population; it does not measure the risk of acquiring a disease. Prevalence is a proportion but not a rate, because the denominator does not include a unit of time.

$$\text{Prevalence} = \frac{\text{total \# of cases of a disease at a time}}{\text{total population}}$$

While incidence measures the appearance of new cases of disease, prevalence measures the existence of disease. Prevalence depends on two factors--the incidence and the duration of disease (Prevalence = Incidence \* Duration, or  $P=ID$ ). For example, given the same incidence, the prevalence will be twice as high for disease A as compared with disease B if the duration of A is twice that of B. A change in prevalence may thus reflect a change in disease incidence and/or duration of disease. For example, improvements in therapy can decrease prevalence if cases are considered cured, however, if these improvements increase duration of life with the disease, the result may be an increase in prevalence. Because prevalence is a product of incidence times duration, it may increase or decrease without a change in incidence if the duration of a disease is altered.

Prevalence is used by health planners because it measures the current need for services, i.e., treatment facilities, staffing needs. Prevalence may be determined by a single survey ("cross-sectional"), whereas incidence rates are more difficult to determine. Prevalence may be used for descriptive purposes only if appropriate caution is exercised in interpretation, with consideration given to variations in duration of disease.

The same limitations in the calculation and interpretation of prevalence apply as mentioned above for incidence rate. The numerator must be included in the denominator and all persons in the denominator must be at risk of inclusion in the numerator. The accuracy of numerator and denominator vary with the methods available for ascertaining them,. Differences in prevalence are influenced by exposure and susceptibility, as well as by duration of disease and accuracy of identification.

When the term "prevalence" is used, "point prevalence" is what is usually meant, and is defined as above. "Period prevalence" is used rarely, and is defined as initial point prevalence plus incidence during a subsequent period of time.

### Questions 5 & 6.

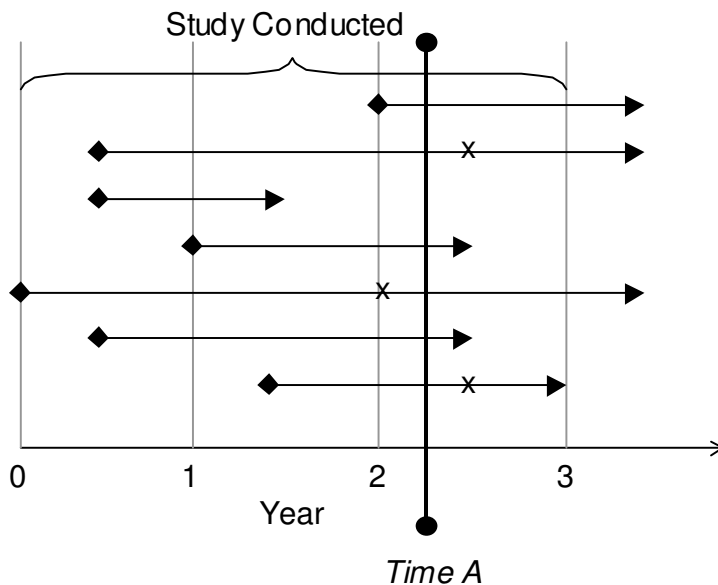
Ten percent (10%) of a sample of 20,000 civil servants underwent a clinical exam. These exams revealed that 50 persons had clinical evidence of coronary heart disease.

5. What is the prevalence of coronary heart disease in the sample population?

6. Would these 50 people be included in a study of the incidence of coronary heart disease in the above civil service sample population? Explain your answer.

### Question 7.

Each horizontal line in the following diagram represents a single person. (The diamond represents when the person entered a study and the arrowhead indicates when the person completed the study.) The "x" represents when that person was diagnosed with Alzheimer's disease. What is the Alzheimer's disease *period* prevalence during the first 3 years of the study? What is the Alzheimer's disease *point* prevalence at time "A"?



### Attack Rates

$$\text{Attack Rate} = \frac{\text{Number of persons affected}}{\text{Population at risk}}$$

An attack rate is usually expressed as a percent and generally reserved for acute outbreaks of

infectious or toxic diseases, as in an epidemic or common source outbreak. The attack rate is actually a proportion, but the term is in common usage.

The numerator (number of persons affected) is ascertained through case reports, enumerations, investigations, etc. The reliability of this figure is influenced by (1) the adequacy of the method of obtaining it, and (2) the nature of the disease. If symptoms of the diseases are obscure and the diagnosis difficult, the numerator will be less reliable.

The denominator is the population at risk and may refer to the whole population of a civil subdivision or of an entire nation, or it may be a restricted or selected group such as a specific age or sex category, the persons served by a certain milk company or the people who attend a certain clinic. The data can be obtained through census, enumeration on a special occasion, by investigations, etc. The reliability of this figure depends on the accuracy of these methods. The size of the denominator is the average population size during the time interval under consideration, often taken at the midpoint.

The attack rate reflects the impact of a disease upon a population at a given time. Any one attack rate may then be compared with similar ones for other groups, or for the same group at a different time.

**Question 8.**

Referring to question 4: At the end of the follow-up period what was the attack rate of measles?

Secondary attack rate is a term reserved for infectious disease studies, and refers to the number of new cases occurring per stated unit of contacts of a 'primary' case, during a time interval that covers the incubation period of the disease population:

$$\text{Secondary attack rate} = \frac{\# \text{ cases among contacts}}{\# \text{ contacts exposed}}$$

Comparisons based on secondary attack rate are likely to be more exact than those based on total attack rate because: (1) there is greater uniformity of exposure among known contacts (2) the numerator is more reliable if, as is usually the case, the group is under closer observation and (3) the denominator is more accurate since the group is usually small, predetermined, and actually counted. For greatest precision, the denominator can be constructed to include only persons known to be susceptible.

### The Incidence Number

For notifiable diseases, such as reportable acute communicable disease, the number of cases during a specified time period is usually recorded by week, month, or year, rather than as a rate per 100,000 population. The rate, however, is implied by recording the population size of the communities in which the reported cases occurred. In comparing the incidence number of cases between communities, taking into consideration population size and other factors, such as age, is important.

| HEALTH ISTRICT     | POP<br>in 1000's | Mumps | Gonorrhea | Hepatitis<br>Infectious | Salmonella | Tuberculosis<br>all forms | Typhoid<br>Fever |
|--------------------|------------------|-------|-----------|-------------------------|------------|---------------------------|------------------|
| Central            | 200              | 3     | 150       | 10                      | 0          | 14                        | 1                |
| East Los Angeles   | 156              | 8     | 32        | 4                       | 1          | 5                         | 0                |
| Hollywood Wilshire | 368              | 1     | 310       | 14                      | 6          | 7                         | 1                |
| Inglewood          | 432              | 7     | 85        | 10                      | 3          | 5                         | 0                |
| Pomona             | 314              | 19    | 56        | 6                       | 0          | 2                         | 0                |
| San Antonio        | 282              | 6     | 68        | 8                       | 2          | 4                         | 0                |
| Southeast          | 102              | 0     | 101       | 4                       | 2          | 9                         | 0                |
| Southwest          | 310              | 4     | 272       | 6                       | 0          | 9                         | 0                |

### **Question 9.**

Which Health District has the highest gonorrhea incidence *number*? Which district has the highest incidence *rate*?

### Mortality Rates

Mortality (death) rates measure the frequency of deaths within specific populations and are calculated for a given time interval and place. A crude death rate expresses the proportion of a population who die of all causes, while a cause-specific death rate is the proportion of a population who die of a certain disease. The numerator is the number of persons dying; the denominator is the total population (such as the midyear population) in which the deaths occurred. The unit of time is usually a calendar year.

$$\text{Crude mortality} = \frac{\text{total \# deaths from all causes in 1 year}}{\text{total population at midyear}} \times 10,000 \text{ per year}$$

$$\text{Cause specific mortality} = \frac{\text{total \# deaths from a specific cause in 1 year}}{\text{total population at midyear}} \times 10,000 \text{ per year}$$

Note that both rates above are expressed per 10,000 person-years.

### Questions 10-12

Between January 1 and December 31, 1995, there were 62,942 deaths (from all causes) in Los Angeles County. Of 9,333 deaths certified as Myocardial Infarction (MI) in Los Angeles County that year, 593 occurred among white males age 45-54 and 35 occurred among black males in the same age group. The Census estimated that the midyear population was 7,000,000. The Census also calculated that there were 366,899 white males age 45-54 and 34,741 black males age 45-54 in this population.

10. Calculate the crude mortality rate.

11. Calculate the cause-specific mortality rate for MI deaths.

12. Calculate the cause-specific mortality rate for MI deaths among black males age 45-54.

### Proportionate Mortality

Proportionate mortality is the proportion of total deaths due to a specific cause. Because proportionate mortality does not express deaths in terms of all people at risk of death, **it is not a measure of risk**. Proportionate mortality is commonly used to illustrate the relative importance of a given cause of death among total deaths. It is expressed as a percentage of the total (per 100).

$$\text{Proportionate mortality} = \frac{\# \text{ deaths from a specific cause}}{\text{total \# deaths (all causes)}} \times 100$$

### Question 13.

Of 62,942 deaths (all causes) in Los Angeles County, 1995, 10,920 were certified as cancer. What is the proportionate cancer mortality?

### Case Fatality Rate

Case fatality rates are commonly used to indicate severity of a given disease. The numerator consists of persons who die from a specific disease and the denominator (population at risk) is composed of a group of persons ill with the disease. The resulting measure is often stated as a biological characteristic of a disease, but it should be defined by a time period and by a specific locality, since it may well vary by time and place. It is expressed as a percentage (per 100) and is usually a proportion rather than a true "rate"

$$\text{Case fatality rate} = \frac{\text{\# deaths resultant from a specified disease in a time interval}}{\text{\# cases of the same disease during same time interval}} \times 100$$

### **Question 14.**

Of the approximately 45,000 Myocardial Infarctions (MI) reported during 1995 in Los Angeles County, there were 9,333 deaths certified as MI on the death certificates. What is the case fatality rate for MI's?

### **Question 15.**

What factors are likely to influence case fatality rates? How?

The case fatality rate (CFR), as determined from hospital records, is often considerably greater than the 'true' rate because only the more severe cases are hospitalized.

Some other mortality rates which are important are:

Infant Mortality Rates = # of deaths in children <1 years old / # of live births

Fetal Death Rates = # of fetal deaths after 20 weeks gestation / # of pregnancies >20 weeks in the population

Perinatal Death Rates = # of fetal deaths after 24 weeks gestation + # of deaths in newborns ≤7 days / number of births

Neonatal Death Rates = # of deaths in newborns ≤28 days / # of live births

Maternal Mortality Rates = # of deaths in women related to pregnancy and delivery / # of pregnancies in the population

### **Illustrative Examples for the Use of Measures of Disease**

Example A. Although patients with mental illness have at times occupied nearly half of the hospital beds in the United States on any given day, and are numbered in the hundreds of thousands, it is

generally recognized that such persons represent only the tip of the iceberg of mentally ill persons. In order to obtain better estimates of the magnitude of the problem, a number of investigators conducted community surveys in various parts of Europe and the U.S. during the 1930's. These were independent studies, employing a variety of procedures for case-finding and diagnosis.

The overall description of mental disorders found in these surveys are as follows:

**Mental disorders found in 4 population surveys**

| Place          | Population Size | Characteristics  | # Cases | # Cases per 1000 |
|----------------|-----------------|------------------|---------|------------------|
| Thuringia, Ger | 37,561          | Agric/indus/Prot | 479     | 12.8             |
| Bavaria, Ger   | 8,628           | Agric/Catholic   | 517     | 59.9             |
| Bornholm, Denm | 45,930          | Agric/fish/tour  | 716     | 15.6             |
| Baltimore, MD  | 55,129          | Urban/poor/black | 1,721   | 31.2             |

**Question 16.**

Do you think the # of cases per 1000 in the last column is incidence or prevalence? Why?

**Question 17.**

How might you account for the differences in results found in these surveys?

While diagnostic criteria have not been standardized among studies of mental illness, individual studies usually have standard criteria for identifying illness within the study. A common approach for detecting mental illness is to identify symptoms rather than diagnostic categories. For example, in a study in Nova Scotia, self reported symptoms were used to determine the need of psychiatric treatment. This need was validated by the judgment of a physician. The authors reported 94% agreement between the criteria for needed treatment and the judgment of a physician, an 'excellent' degree of validity. Then the authors examined a variety of individual characteristics in relation to mental illness in order to obtain etiological leads. One such characteristic examined was social class, and the following data were reported.

**Need for psychiatric treatment by social class,  
Stirling County, Nova Scotia**

| Social Class | Number<br>examined | Number<br>with need | %    |
|--------------|--------------------|---------------------|------|
| I (upper)    | 203                | 18                  | 8.9  |
| II           | 443                | 160                 | 36.1 |
| III          | 153                | 85                  | 55.6 |
| IV (lower)   | 113                | 95                  | 84.1 |

Findings similar to the Nova Scotia study, with regard to social class and mental illness, have been reported from other population surveys. In the Midtown Manhattan Study, income and ethnicity were examined in relation to mental illness with the following results:

**Mental illness by income and ethnicity, Midtown Manhattan (expressed as per 100)**

| Annual Income | Puerto<br>Rican | Black | Jewish | Irish | Other |
|---------------|-----------------|-------|--------|-------|-------|
| <\$3,000      | 36.4            | 26.3  | 10.5   | 26.7  | 32.3  |
| \$3,000-4,000 | 37.3            | 14.5  | 29.2   | 21.2  | 30.3  |
| \$5,000-7,499 | 31.0            | 15.7  | 18.6   | 12.3  | 15.5  |
| \$7,500+      | 41.2            | 21.4  | 15.6   | 11.4  | 22.6  |

**Question 18.**

From the data presented in Stirling County, Nova Scotia study are you willing to accept social class as an important factor in the etiology of mental illness (i.e. cause of mental illness)? Why or why not?

**Question 19.**

How would you interpret the results by income in the Midtown Manhattan data?

**Question 20.**

Do you think income is a good approximation for a social class measure? In what other ways might social class be measured?

Example B. When only some members of a relatively homogeneous population group are affected with the same illness at about the same time in a common source outbreak, and the others are not, the identification of the specific source may be difficult because it is not possible to describe a "population at risk" by meaningful differences in age, sex, residence, etc. This is the usual situation in the investigation of food-borne outbreaks after the probable time and place of exposure have been determined, and the contaminated food items remain to be detected.

Analogous problem situations might be the presence of an allergen in one brand of deodorant among many for sale in community, or the transient presence of a toxic gas in one shop area among many in an industrial establishment. Since all persons do not necessarily have known exposure, the problem is to compute the attack rate of persons presumed to have been exposed to each of the items under suspicion with the attack rate of persons believed not to have been exposed.

An example of this approach is the investigation of an outbreak of acute gastroenteritis following a church supper in Oswego County, New York (unpublished data widely used in practical epidemiology training courses). Of the 80 persons who had been present, 46 suffered an acute illness characterized by nausea, vomiting, diarrhea and abdominal pain, with an incubation period ranging from 3 to 7 hours. Interviews were possible with 75 of the people present at the supper and detailed information was obtained regarding occurrence of symptoms, time of onset, and food consumed at the supper.

From the detailed *line-listing* of cases, the following table could be prepared.

| Food and Beverages Consumed and Illness |                 |         |                 |                         |         |                 |
|-----------------------------------------|-----------------|---------|-----------------|-------------------------|---------|-----------------|
| Food or Beverage                        | Persons who ate |         |                 | Persons who did not eat |         |                 |
|                                         | Total           | No. ill | Attack Rate (%) | Total                   | No. ill | Attack Rate (%) |
| Baked ham                               | 46              | 29      | 63              | 29                      | 17      | 59              |
| Spinach                                 | 43              | 26      | 60              | 32                      | 20      | 63              |
| Mashed potato                           | 37              | 23      | 62              | 38                      | 23      | 52              |
| Cabbage salad                           | 28              | 18      | 64              | 47                      | 28      | 60              |
| Jello                                   | 23              | 16      | 70              | 52                      | 30      | 58              |
| Rolls                                   | 37              | 21      | 57              | 38                      | 25      | 66              |
| Brown Bread                             | 27              | 18      | 67              | 48                      | 28      | 58              |
| Milk                                    | 4               | 2       | 50              | 71                      | 44      | 62              |
| Coffee                                  | 31              | 19      | 61              | 44                      | 27      | 61              |
| Water                                   | 24              | 13      | 54              | 51                      | 33      | 65              |
| Cakes                                   | 40              | 27      | 67              | 35                      | 19      | 54              |
| Van. Ice Cream                          | 54              | 43      | 80              | 21                      | 3       | 14              |
| Choc. Ice Cream                         | 47              | 25      | 53              | 28                      | 21      | 74              |
| Fruit Salad                             | 6               | 4       | 67              | 69                      | 42      | 61              |

**Question 21.**

How would you handle data from people who cannot recall whether or not a particular food was eaten?

**Question 22.**

What would be the ideal finding in the table above to incriminate a particular food or beverage item? Why is the ideal unlikely to be achieved?

**Question 23.**

Which food above is the most likely to have caused the illness?

**Question 24.**

The consumption of some foods in the table appear to have been negatively associated with illness (associated with the absence of disease). How might you explain this?

**Question 25.**

After incriminating a food statistically, as above, what further steps would be necessary to prove that it was the source?