Measures used in Epidemiology

Sinan Iyisoy

Measures in Epidemiology

Epidemiologists ask “Who gets the disease? In what frequency? Is the frequency changing over time?

How does the frequency in one population compare to the frequency in another population?”

Answering such questions first requires a case definition. Second, we then need measures to estimate disease frequency. Third, we may want to compare measures in different populations (association).

Measures in Epidemiology

Suppose you are asked to estimate the population prevalence of attention deficit hypersensitivity disorder (ADHD) among U.S. school-age children. How will you identify the children who should be counted as cases of ADHD? What defines a ‘case’?

The definition of a ‘case’ is critical in planning an epidemiological investigation. The case definition must be carefully formulated to meet objectives of the investigation, while also permitting valid comparisons with results from other studies.

In this example, it may be of interest to consider whether the proportion of school age children with ADHD has changed over a period of time. If the case definition changes significantly from one time period to the next, comparisons with previous years are problematic.

Outcomes of Epidemiology

- **Death**
- **Disease/Illness - Physical signs, laboratory abnormalities**
- **Discomfort - Symptoms (e.g., pain, nausea, dyspnea, itching, tinnitis)**
- **Disability - Impaired ability to do usual activities**
- **Dissatisfaction - Emotional reaction (e.g., sadness, anger)**
- **Destitution - Poverty, unemployment**

Death and disease are the most frequently used outcomes.

Measures in Epidemiology

We consider two groups of measures that are used in Epidemiology

- **Measures of Disease Frequency**
- **Measures of Association (Effect)**

Measures of Disease Frequency

Count (Sayi)

the number of individuals who meet the case definition

5621 cases of invasive colorectal cancer in Konya in 2016.

Calculating the magnitude of disease occurrence with a count is simple and useful for certain purposes, such as allocating health resources. For other purposes, it is more helpful to have a denominator under the count that indicates the size of the study populaton. The following four measures address this.

Measures of Disease Frequency

Proportion (Yuzde)

A/(A+B): a fraction in which the numerator (A) includes only individuals who meet the case definition and the denominator totals the numbers of individuals who meet the case definition plus those in the study population who do not meet the case definition and are at risk.

$$\frac{Numerator}{Denominator}$$

A proportion is *not dependent* upon time. It may be expressed as a fraction or a percentage. A proportion indicates the fraction of the population that is affected by the disease or condition.

Measures of Disease Frequency

Proportion

30% of persons over 50 years of age have been screened for colon cancer.

Calculating the proportion of women with cervical cancer requires a special consideration. Cervical cancer only occurs in women with a cervix. A woman who has had a complete hysterectomy is no longer at risk for developing cervical cancer. This is a large segment of the population of older women.

Population at Risk

The National Women's Health Information Center of the US Department of Health and Human Services reports that 1 in 3 women have had a hysterectomy by age 60. Thus, women with hysterectomies are not included in the denominator when calculating the proportion of women with cervical cancer as part of the *population at risk*.

Measures of Disease Frequency

Ratio (Oran)

A/B: a special fraction in which the numerator includes only individuals who meet one criterion (e.g. the case definition) and the denominator includes only individuals in the study population who meet another criterion (e.g. do not meet the case definition but are *at risk*).

A ratio is *not dependent* upon time. A ratio as a measure of disease frequency is used infrequently, in special situations. (not to be confused with an odds-ratio or risk-ratio)

Measures of Disease Frequency

Ratio

Ex: 1 case of colon cancer for every 1 case of breast cancer.

Ex: 2 female cases of major depression to 1 male case of major depression.

A ratio is not bounded from above, it can be any non-negative value. But a proportion is bounded between 0 and 1.

Measures of Disease Frequency

Rate

a fraction in which the numerator includes only individuals who meet the case definition and the denominator includes individuals in the study population who do or do not meet the case definition but could meet the case definition (at-risk).

A rate is *dependent* upon time. In other words, *a proportion over a particular period of time*. An epidemiological rate will contain the following: disease frequency (numerator), unit of population size, and the time period during which the event occurred

Ex: 44 cases of colon cancer per 100,000 population in Konya during 2000.

Measures of Disease Frequency

Risk

the probability of an individual meeting the case definition in a period of time.

Risk *is dependent* upon time.

Ex: The 10-year risk that a 45-year-old male will develop prostate cancer is 5% (typically derived from a cohort study in which each at-risk person is followed over time until he/she is no longer at-risk)

Measures of Disease Frequency

The smoking *proportion (~~rate~~)* among adults is the number of adults in a population who smoke divided by the total number of adults in the population.

A rate can never be expressed as a percentage, while a proportion should always be able to be expressed as a percentage.

Rates are usually expressed per 100, 1,000, or 100,000 persons. In a strict application, "rate" should only be used when the denominator is an estimate of the total person-time at risk.

Measures of Disease Frequency

For measuring disease frequency, proportions and rates are very helpful when comparing groups, because they relate the number of people with disease to the size of the population in which they occur.

Prevalence and incidence are the two fundamental measures of disease frequency.

Measures of Disease Occurence

Point Prevalence

the proportion of the population that has disease at a particular time (period).

	Existing Cases	Population Size	Prevalence (as a decimal fraction)	Prevalence (per 10000 population)
City A	75	30000	0.0025	25 per 10000
City B	35	7000	0.0050	50 per 10000

Period Prevalence

Measures of Disease Occurence

Incidence

a measure of the occurrence of new cases of disease during a span of time.

Incidence quantifies development of disease. Incidence can be estimated using data from a disease registry data or a cohort trial. There is an implicit assumption of a period of time, such as new cases within a month.

Measures of Disease Occurence

Incidence Proportion (Cumulative incidence, risk)

the probability of developing disease over a stated period of time; as such, it is an estimate of risk.

If a population initially contains 1,000 non-diseased persons and 28 develop a condition over two years of observation, the incidence proportion is 28 cases among 1,000 persons per two years, i.e. 2.8% per two years.

Measures of Disease Occurence

A newspaper article states that women who are 60 years of age have a 2% risk of dying from cardiovascular disease.

Incidence proportion must specify a certain period of time. A 2% risk has a different meaning if it is over 12 months or 10 years.

The concept of risk is fairly intuitive - if a group of disease-free people were followed for a period of time, one could determine the *proportion* of people who developed the disease *at some point* during the observation period in order to arrive at an estimate of the probability of developing that disease, i.e. the risk.

Measures of Disease Occurence

The incidence proportion doesn't distinguish when a disease occurs, as long as it is within the follow-up period.

For example, if a population is followed for 20 years, it would make a difference to the person and to the epidemiologist if the cancer occurred after two years or after 20 years, but both of these outcomes would count the same with the incidence proportion.

Measures of Disease Occurence

For this reason, the incidence proportion is generally used in situations where the follow-up time is relatively short and there is relatively little loss to follow-up.

Otherwise, epidemiologists generally use the

Incidence Rate (Incidence density (rate))

Incidence rate is a measure of incidence that incorporates time directly into the denominator.

An incidence rate is generally calculated from a long-term cohort follow-up study,

Measures of Disease Occurence

wherein enrollees are followed over time and the occurrence of new cases of disease is documented.

Each person is observed from an established starting time until one of four "end points" is reached: onset of disease, death, migration out of the study ("lost to follow-up"), or the end of the study.

Similar to the incidence proportion, the numerator of the incidence rate is the number of new cases identified during the period of observation. However, the denominator differs.

Measures of Disease Occurence

The denominator is the sum of the time each person was observed, totaled for all persons. This denominator represents the total time the population was at risk of and being watched for disease.

Thus, the incidence rate is the ratio of the number of cases to the total time the population is at risk of disease.

After two years of observation 28 cases occurred per 2500 person-year. So incidence rate is 0.0112 cases per person-year, or 11.2 cases per 1000 person-years.

Measures of Disease Occurence

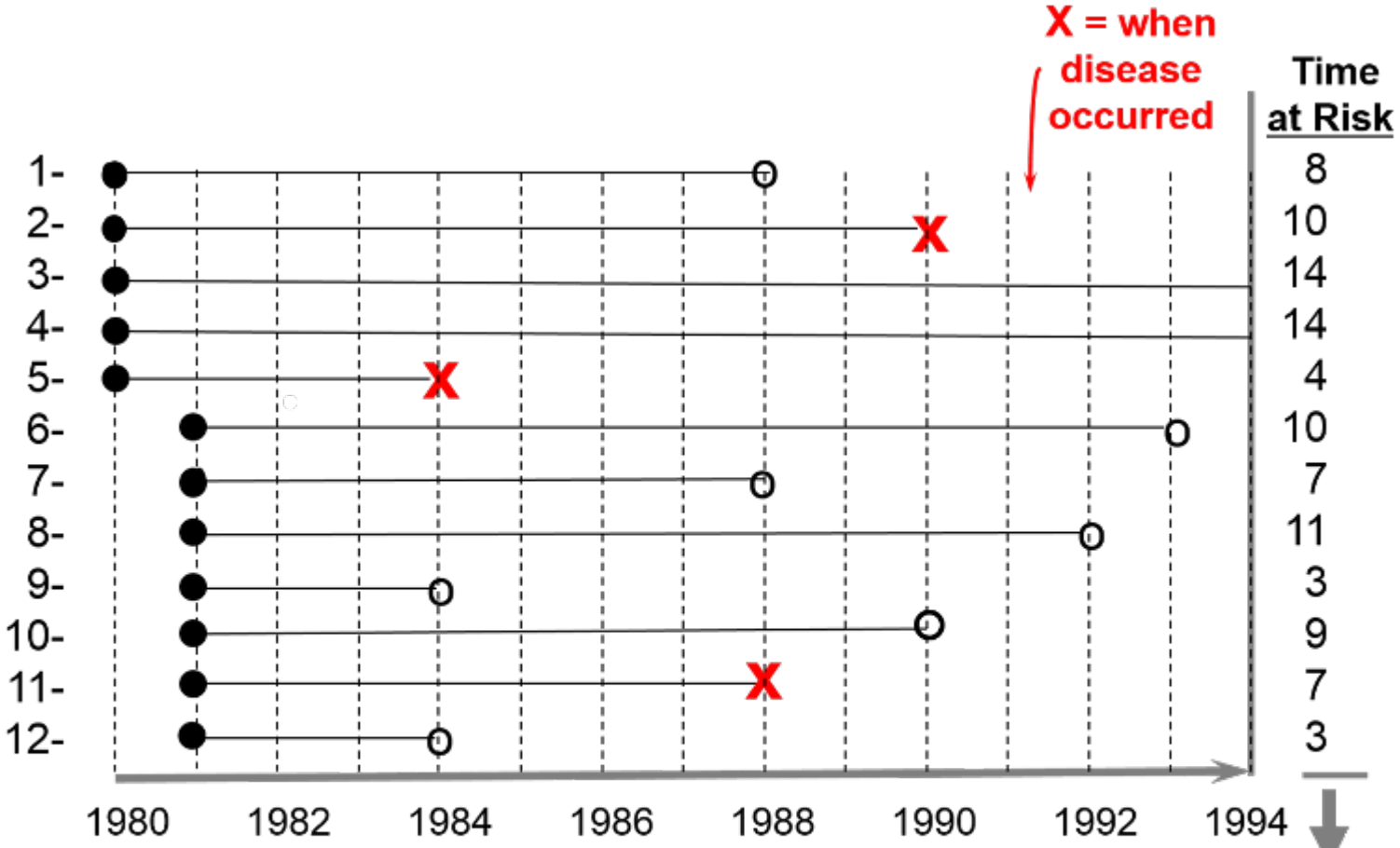
Measure	Numerator	Denominator
Incidence proportion (or attack rate or risk)	Number of new cases of disease during specified time interval	Population at start of time interval
Incidence rate (or person-time rate)	Number of new cases of disease during specified time interval	Summed person-years of observation or average population during time interval
Point prevalence	Number of current cases (new and preexisting) at a specified point in time	Population at the same specified point in time
Period prevalence	Number of current cases (new and preexisting) over a specified period of time	Average or mid-interval population

Measures of Disease Occurence

Consider a hypothetical clinical trial that was conducted to determine whether taking low-dose aspirin reduced the frequency of heart attacks in middle-aged and elderly men.

The time line below summarizes events 12 subjects labeled 1-12, all of whom were allocated to the placebo-treated group.

Measures of Disease Occurence



$$IR = \frac{3}{100 \text{ p-yrs.}}$$

Measures of Disease Occurence

Incidence proportion= 3/12 = 25% over 14 years.

Incidence rate = 3/100 = 0.03 cases per person-years or 3 cases per 100 person-years.

Measures of Association

The key to epidemiological analysis is comparison. Occasionally you might observe an incidence rate among a population that seems high and wonder whether it is actually higher than what should be expected based on, say, the incidence rates in other communities.

Or, you might observe that, among a group of case-patients in an outbreak, several report having eaten at a particular restaurant. Is the restaurant just a popular one, or have more case-patients eaten there than would be expected? The way to address that concern is by comparing the observed group with another group that represents the expected level.

Measures of Association

A measure of association quantifies the relationship between exposure and disease among the two groups.

Exposure is used loosely to mean not only exposure to foods, mosquitoes, a partner with a sexually transmissible disease, or a toxic waste dump, but also inherent characteristics of persons (for example, age, race, sex), biologic characteristics (immune status), acquired characteristics (marital status), activities (occupation, leisure activities), or conditions under which they live (socioeconomic status or access to medical care).

Measures of Association

Consider the following example regarding the management of Hodgkin lymphoma, a cancer of the lymphatic system. Years ago when a patient was diagnosed with Hodgkin Disease, they would frequently undergo a surgical procedure called a "staging laparotomy".

At times, the surgeons performing this procedure would also remove the patient's appendix, not because it was inflamed; it was done "incidentally" in order to ensure that the patient never had to worry about getting appendicitis.

Measures of Association

Some surgeons felt that doing this "incidental appendectomy" did the patient a favor by ensuring that they would never get appendicitis, but others felt that it meant unnecessarily increasing the patient's risk of getting a post-operative wound infection by spreading around the bacteria that was once inside the appendix.

To address this, the surgeons at a large hospital performed a retrospective cohort study.

Measures of Association

Relative Risk (Risk Ratio)

For the study examining wound infections after incidental appendectomy, the risk of wound infection in each exposure group is estimated from the cumulative incidence. The relative risk (or risk ratio) is an intuitive way to compare the risks for the two groups. Simply divide the cumulative incidence in exposed group by the cumulative incidence in the unexposed group:

Had Incidental Appendectomy?	Wound Infection	No Wound Infection	Total
Yes	7	124	131
No	1	78	79

$$Relative\ Risk = \frac{CI_{exposed}}{CI_{unexposed}} = \frac{7/131}{1/79} = 4.2$$

Measures of Association

Interpretation

In this study, patients who underwent incidental appendectomy had 4.2 times the risk of post-operative wound infection compared to patients who did not undergo incidental appendectomy.

Relative Risk<1

It is also possible for the relative risk to be less than 1; this would suggest that the exposure being considered is associated with a reduction in risk.

Measures of Association

Treatment	Myocardial Infarction	No Infarction	Total	Cumulative Incidence
Aspirin	139	10898	11037	139/11037 = 0.0126
Placebo	239	10795	11034	239/11034 = 0.0217

$$Relative\ Risk = \frac{0.0126}{0.0217} = 0.58$$

Interpretation

Those who take low dose aspirin regularly have 0.58 times the risk of myocardial infarction compared to those who do not take aspirin.

Measures of Association

Note also that the unexposed (comparison, reference) group must be specified. For example, if we simply said, "Those who take low dose aspirin regularly have 0.58 times the risk of myocardial infarction", the question is "compared to what?".

Is it those who didn't take any aspirin, those who took low-dose aspirin but used it irregularly, those who took high dose aspirin, those who took acetaminophen...?

Measures of Association

Pitfalls

Note that in the interpretation of RR both the appendectomy study (in which the RR > 1), and the aspirin trial (in which RR < 1) used the expression "times the risk." To be precise, it is not correct to say that those who had an incidental appendectomy had 4.2 times *more* risk (wrong) or 4.2 times *greater* risk (wrong).

In fact, those with the incidental appendectomy had a 320% increase in risk. Conversely, in the aspirin study it is not correct to say that those on aspirin had 0.57 times *less* risk (wrong). In fact, they had 43% less risk.

Measures of Association

Rate Ratio

Rate ratios are closely related to risk ratios (Relative Risk), but they are computed as the ratio of the incidence rate in an exposed group divided by the incidence rate in an unexposed (or less exposed) comparison group.

$$Rate\ Ratio = \frac{IR_{exposed}}{IR_{unexposed}}$$

Measures of Association

Consider an example from The Nurses' Health Study. This prospective cohort study was used to investigate the effects of hormone replacement therapy (HRT) on coronary artery disease in post-menopausal women.

The investigators calculated the incidence rate of coronary artery disease in post-menopausal women who had been taking HRT and compared it to the incidence rate in post-menopausal women who had not taken HRT. The findings are summarized in this table:

Measures of Association

Post-menopausal Hormone Use	# with Coronary Artery Disease	Person-Years of Disease-free Follow-up
Yes	30	54308.7
No	60	51477.5

the rate in those using hormones was 30 / 54308.7 = 55.2 per 100000 person-years

the rate in those NOT using hormones was 60 / 51477.5 = 116.6 per 100000 person-years.

So, the rate ratio was 0.47.

Interpretation: Women who used postmenopausal hormones had 0.47 times the rate of coronary artery disease compared to women who did not use postmenopausal hormones.

Measures of Association

Multiple Groups

Some cohort studies and clinical trials compare the risk of disease or other outcomes among three or more exposure groups.

In this situation, results can be summarized in a table with multiple rows to accommodate the multiple exposure groups. This is a logical extension of the basic "2 x 2" table and is sometimes referred to as an "r x c" table (row and columns).

Measures of Association

The table below summarizes a study examining the association between exposure to magnetic fields, e.g., from high tension wires, and the risk of leukemia. In this study there are no unexposed subjects, but we can classify them as having low, medium, and high exposure.

To compute the relative risks , it is logical to use the least exposed group as a "reference group" against which we can compare the other two exposure (or "index") groups. Note that in this example the investigators calculated and compared cumulative incidence.

Measures of Association

Magnetic Field Exposure	Leukemia	No Leukemia	Total	Cumulative Incidence
High	30	644	674	30 / 674 = 0.0445
Medium	61	1408	1469	61 / 1469 = 0.0415
Low (Reference)	2264	65160	67424	2264 / 67424 = 0.0336

The risk ratio for medium exposure compared to low exposure (the reference group) is 0.0415/0.0336 = 1.23.

The risk ratio for high exposure compared to low exposure (the reference) is 0.0445/0.0336 = 1.33.

Measures of Association

Interpretation

Compared to children exposed to low magnetic field levels,

those exposed to medium levels have 1.23 times the risk of leukemia (a 23% increase in risk),

and those exposed to high levels have 1.33 times the risk (a 33% increase in risk).

Measures of Association

Risk Difference

Instead of comparing two measures of disease frequency by calculating their ratio, one can compare them in terms of their absolute difference.

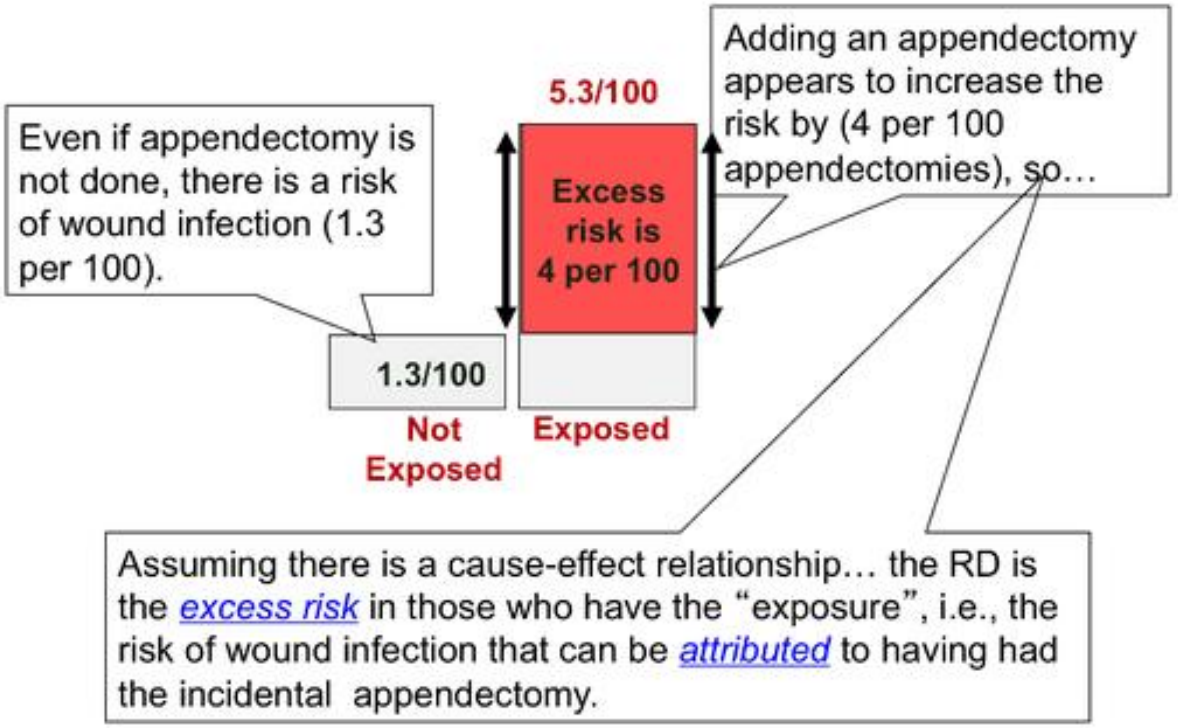
The risk difference is calculated by subtracting the cumulative incidence in the unexposed group (or least exposed group) from the cumulative incidence in the group with the exposure.

$$Risk\ Difference = CI_{exposed} - CI_{unexposed}$$

Measures of Association

The risk difference focuses on absolute effect of the risk factor, or the excess risk of disease in those who have the factor compared with those who don't. Recall that in the wound infection study, the cumulative incidence of infection was 5.3% in the incidental appendectomy group, and only 1.3% in the group without appendectomies. The risk ratio was 4.2, but we can also compute the absolute difference, which is 5.3/100 - 1.3/100 = 4 per 100 excess wound infections among those who had the incidental appendectomy. $CI_{unexposed}$ provides an estimate of the baseline risk (i.e., in the absence of the exposure), and the exposure factor imposes an additional (excess) risk on top of that.

Measures of Association



An older term for the risk difference is "attributable risk," that is the excess risk than can be attributed to having had the exposure.
However, many discourage the use of this terminology because it presumes a causal relationship between the exposure and the outcome.

Measures of Association

Rate Difference

Analogous to the risk difference, the rate difference is calculated by subtracting the incidence rate in the unexposed group (or least exposed group) from the incidence rate in the group with the exposure.

$$Rate\ Difference = IR_{exposed} - IR_{unexposed}$$

Measures of Association

In the Nurses' Health Study, the difference between highest and lowest weight categories was about 62.3 cases per 100,000 person-years. If you followed another 100,000 women with BMI's >29 for one year, you would expect about 85 of them to have a non-fatal MI, and we could attribute about 62 of these to their obesity. Conversely, if you got these 100,000 women to lose enough weight to get them down into the leanest category, you might expect to prevent about 62 non-fatal MIs.

Measures of Effect

RR

OR

HR

AdjustedRR

Attributable Risk

NNT NNH