Epidemiology (4) Measuring disease occurrence

Minato NAKAZAWA, Ph.D.
<minato-nakazawa@people.kobe-u.ac.jp>

MEASURES OF DISEASE OCCURRENCE

- Measurement is a central feature of epidemiology
  - As the study of the occurrence of illness
- Broad interpretation of illness
  - Injuries
  - Birth defects
  - Health outcomes
  - Other health related events and conditions
- As the measures of disease occurrence, 3 fundamental ones are introduced in this chapter
  - Risk
  - Incidence rate
  - Prevalence

Risk and Incidence Proportion

- Risk (Incidence proportion)
  - Probability
  - In a small group, **the probability** that **a person** will develop a given disease
  - In larger groups, let A the number of people who develop the disease among the number of people in the beginning of the observation N, risk can be defined as **A/N** (A < N, thus A/N < 1)
    - It assumes that all of the N people are followed for the entire period
    - The average risk for a group is also referred as **incidence proportion**
- Advantage of using risk
  - Readily understood by many people
- To note, What does "60 years old women have 2% risk of dying of cardiovascular disease" in newspaper mean?
  - Within next 24 hours, it's not true
  - Even for 1 year, it's too high
  - The length of time over which the risk applies is essentially important. Without the observation period, risk has no meaning

Table 4-1. Comparison of incidence proportion (risk) and incidence rate

Property	**Incidence Proportion**	**Incidence Rate**
Smallest value	0	0
Greatest value	1	Infinity
Units (dimensionality)	None	1/time
Interpretation	Probability	Inverse of waiting time

- The risk increases with time (See, Fig 4-1)
  - Pattern A: Risk climbs rapidly early to reach plateau
  - Pattern B: Risk climbs at slowly but steadily rate
- Possible conditions
  - A: Susceptible people in the beginning become immunized after recovery
  - B: Exposure to DES in the beginning gradually develop vaginal adenocarcinoma / Chronic NCDs with aging

Pros and Cons of risk

- Cumulative measure
  - For a given person, risk increases with the length of observation
  - For a given period, risk**s** for a person can rise or fall with time
    - 1-year risk of dying in an automobile crash for a driver: For any one person, risk accumulates, but 1-year risk is greater for most drivers in their teenage years than 50s
- Drawback as a tool for assessing the disease occurrence
  - Impossible to measure risk over any appreciable time interval
  - During sufficient time, some people in the population will die from causes other than the outcome under study (**competing risks**)
    - (eg.) DV in 10000 married women over 30 years: Some of them will die before the completion of 30 years by cardiovascular disease, cancer, infection, vehicular injury, and other causes. If a woman dies by cancer after 5 years, we don't know whether the woman will become a victim of DV during the subsequent 25 years if she wouldn't suffer from cancer
    - If the number of person died due to competing risks is included in denominator N, the risk of DV will be underestimated
  - When we assess the all-causes mortality, no competing risks. When we assess the risk within short time period, the influence of competing risks is small
    - Evaluation of the efficacy of Salk vaccine for polio in schoolchildren in 1954 was done by 1-year follow up and during that period competing risks are negligible, it was reasonable
  - **Loss to follow-up** is another issue.

(Column) Attack rate and case fatality *rate* (Note: case fatality **risk** is preferred now)

- Attack rate: term for risk used in connection with infectious outbreaks
  - Risk of contracting a condition during an epidemic period
  - (eg.) When flu epidemic has 10% attack rate, 10% of the population will develop the disease during the epidemic
  - Time reference is not stated but implied by biological nature of the disease
    - Usually short, typically a couple of months, sometimes less
- Secondary attack rate: Attack rate among susceptible people who come into direct contact with primary cases (the cases infected in the initial wave of an epidemic) (See, Chapter 6)
- Case-fatality rate (it's older term, so hereafter I use **Case-Fatality Risk**): The proportion of people dying of the disease (fatality) among those who develop the disease (case). In general, the denominator is the number of confirmed cases. Sometimes, the numerator is the number of all deaths in the cases, but usually the number of deaths caused (succumbed) by that disease (See, Chapter 13).
  - It indicates the severity of the disease.
  - Same as attack rate, time reference is usually implicit
  - CFR of measles in USA is 1.5/1000. This time reference is much shorter than other outcomes than death by measles infection
  - For the diseases with long term process like MS, the interpretation of CFR is difficult and thus other measures may be used.

Incidence Rate

- The issues of competing risks can be addressed by changing the denominator from the population observed to the total person-time observed.
- Incidence rate = A/Time = (Number of subjects developing disease)/(Total time experienced for the subjects followed)
- The denominator is the sum of the time that each person is followed for every member. If 5 people are followed for 30 years, the denominator is 150 person-years. If among those 5 people, 4 people were followed for 30 years but 1 died after 5 years from the beginning of the observation, the denominator is 125 person-years.
- For people who don't die during follow-up, 2 methods of counting
  - If the disease can recur (like upper respiratory tract infection), the numerator includes all recurrence, the denominator includes all the time during which each person was at risk of getting infected
  - If the disease can occur only once (or outcome is death, in which the incidence rate is the mortality rate), the person ceases from the population at risk after the event occurrence
  - If a person is lost from follow-up, or dies from a competing risk, the person also ceases from the population at risk

Examples

- Hypothetical example of leukemia (Fig 4-2)
  - No value for time is given in the text, but here I assumed as follows
  - Among 5 people, only the 1st died of leukemia after 2 years. 2nd died of competing cause after 3 years. 3rd lost to followup after 1 year. 4th and 5th are censored (completed 5 years followup period without event)
- Incidence rate of leukemia death is 1/(2+3+1+5+5) = 1/16 (/year)
- Comparison between risk and incidence rate (Table 4-1)
  - Risk is probability, no unit, ranges [0, 1]
  - Incidence rate has dimensionality of 1/time, ranges from 0 to infinity

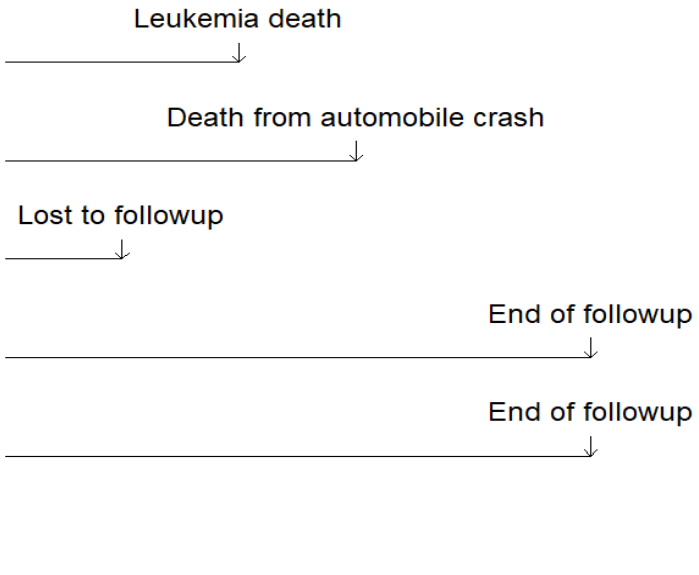

Figure 4-2. Time at risk for leukemia death for 5 people.

- Calculate an incidence rate in a population as 47 cases occurring in 158 person-months (=13.17 person-years).
  - 47 cases/158 person-months
  - 0.30 cases/person-months
  - 47 cases/13.17 person-years
  - 3.57 cases/person-years
- The change of unit results in the change of values

Annual incidence and waiting time

- Incidence rates commonly are described as **annual incidence**, in the form of "50 cases per 100,000"
  - Actually 50 cases per 100,000 person-years or 50/100000 yr^{-1}
  - Negative 1 in the exponent means inverse
- Different from risk, incidence rate may not easily understood
- Dimensionality of an incidence rate is that of the reciprocal of time
  - Under steady-state conditions, the reciprocal of incidence rate is **waiting time**, the average time until an event occurs
  - (eg.) If incidence rate is 3.57 cases per person-year, the waiting time is 1/3.57 = 0.28 years
  - (another eg.) A mortality rate of 11 deaths per 1000 person-years means the average waiting time until death is 1000/11=90.9 years, referred as expectation of life or expected survival time
    - Mortality rate changes with time, so that we cannot assume steady-state conditions. Taking reciprocal of mortality rate is not useful as the method to estimate expectation of life (Usually life table or survival analysis is used)

(Column) Chicken and egg

- Riddle
  - If a chicken and one half lay an egg and one half in a day and one half, how many eggs does one chicken in 1 day?
- Intuition
  - 1 egg?
  - It's wrong!
- Calculation considering "rate"
  - The rate of laying eggs = 1.5 eggs / 1.5 chickens / 1.5 days
    = 1.5 eggs / (1.5 x 1.5) chicken-days
    = 1 eggs / 1.5 chicken-days
    = 2/3 eggs / days

The Relation between Risk and Incidence Rate

- Converting the incidence rate measures to risk measures is convenient
  - The simplest formula: Risk ≈ Incidence rate x Time [4-1]
  - Confirming dimensionality is a good habit
    - Risk is a proportion, no dimension
    - Incidence rate [/time] x Time [time] → no dimension
  - Checking the range of measures is also useful
    - Risk ranges [0, 1]
    - Incidence rate ranges [0, infinity]
    - Time ranges [0, infinity]
    - Products of [0, infinity] and [0, infinity] ranges [0, infinity], does not always match with [0, 1]
  - Only for the risk<20%, equation [4-1] can work as approximation
- (eg.) When incidence rate of lung cancer is 8/10000 person-years and we followed the population for 1 year, risk is 8/10000 for 1 year period. If the same rate is applied for 0.5 year, the risk is 4/10000 for a half year period.
- (eg.2) If mortality rate is 11/1000 person-years and we followed the population for 20 years, the risk of death over 20 years will be 0.22 (22%). It means 220 deaths among 1000 people occurring within 20 years. However, it neglects the fact that the size of population at risk shrinks as deaths occur. Tables 4-2 (a kind of life table) shows 197 deaths instead of 220 deaths will occur within 20 years (It's refereed as **exponential decay**, as shown in Figure 4-3).

Simplified life table

- In a hypothetical cohort of 100,000 people followed for 85 years
- Initial number at risk = 100,000
- Assume **no competing risk** (the number at risk at the start of each age group is reduced only by deaths from motor vehicle injury), **no lost to followup**.
- The qx is obtained by dx/lx for each age group
- The px is 1 – qx for each age group
- The cumulative survival probability is the product of px up to that age
- 1 minus final cumulative survival probability gives the total risk for 85 years
  - 1 – 0.98378 = 0.01622 (1.6%)

Table 4-3. Life table for death from motor vehicle injury from birth through age 85

Age (x)	Number at risk (lx)	Deaths in interval (dx)	Risk of dying (qx)	Survival probability (px)	Cumulative survival probability
0-14	100,000	70	0.00070	0.99930	0.99930
15-24	99,930	358	0.00358	0.99642	0.99572
25-44	99,572	400	0.00402	0.99598	0.99172
45-64	99,172	365	0.00368	0.99632	0.98807
65-84	98,807	429	0.00434	0.99566	0.98378

More realistic life table

(See, demography https://minato.sip21c.org/demography-special/demography-2020-06.pdf)

- In a hypothetical cohort of 100,000 people followed for 85 years
- Initial number at risk = 100,000
- Considering **competing risk** and **lost to followup** (censored).
- Effective number at risk = lx – (lost to followup / died of other causes)/2
  - Censoring is assumed to occur uniformly throughout each age interval.
- The qx = dx/lx' is approximately same as Table 4-3.
- The px is 1 – qx for each age group
- The cumulative survival probability is the product of px up to that age
- 1 minus final cumulative survival probability gives the total risk for 85 years
  - 1 – 0.98378 = 0.01622 (1.6%)

Table 4-4. Life table for death from motor vehicle injury from birth through age 85

Age (x)	At risk (lx)	MVI deaths in interval (dx)	Lost to followup / died of other causes	Effective number at risk (lx')	Risk of dying (qx)	Survival probability (px)	Cumulative survival probability
0-14	100,000	67	9,500	95,250	0.00070	0.99930	0.99930
15-24	90,433	301	12,500	84,183	0.00358	0.99642	0.99572
25-44	77,632	272	20,000	67,632	0.00402	0.99598	0.99172
45-64	57,630	156	30,000	42,360	0.00368	0.99632	0.98807
65-84	27,204	64	25,000	14,704	0.00435	0.99565	0.98377

Point-source and propagated epidemics

- Epidemic: An *unusually high* occurrence of disease
  - *"Unusually high"* depends on the circumstances. No clear demarcation between epidemic and minor fluctuation
- Outbreak: A sudden increase in the occurrence of a disease that is usually absent or nearly absent (As seen in Fig 4-4, Epidemic curve of cholera in London in 1854 around Broad Street Pump)
- Point-source epidemic: An epidemic stems from a single source of exposure to a causal agent
  - (eg.) Food poisoning of restaurant patrons who have been served contaminated food
  - (eg.2) Cancer occurrence among survivors of the atomic bomb blasts in Hiroshima and Nagasaki
  - The typical shape of epidemic curve for a point-source epidemic: Initial steep increase in the incidence rate followed by a more gradual decline (This asymmetry partly reflects less variability in the direction of the zero point than in the other direction: (eg.) The distribution of recovery times for healing of a wound is log-normal)
- Propagated epidemic: An epidemic in which the causal agent is transmitted through a population
  - (eg.) Flu epidemic are propagated by person-to-person transmission of the virus
  - The typical shape of epidemic curve: More symmetric than point-source propagation, because the causes spread gradually through the population (See Chapter 6 in detail)
  - (eg.2) Propagated epidemics can occur over extremely short time spans, as in the case of epidemic hysteria in elementary school (Fig 4-5)

Prevalence proportion

- Incidence proportion (risk) and incidence rate are measures that assess the frequency of disease onset.
- **Prevalence proportion** (often referred as prevalence) does not measure disease onset, but disease status.
  - Disease status: Considering disease as being either present or absent
  - Prevalence: Proportion (P/N) of people who have disease at a given time (P) in a population (N).
  - (eg.) On July 1, 2001, among 10,000 women residents of a town, 1,200 had hypertension. Prevalence of hypertension was 1200/10000 = 0.12 (12%).
- The factors affecting prevalence
  - Disease occurrence
  - Disease duration: Length of time that a person has disease
- Prevalence measures the **disease burden** on a population
- In a steady state, P/(1 – P) = ID. P/(1 – P) is known as the **prevalence odds**.
  - P: Prevalence
  - I: Incidence rate
  - D: Average duration of disease
- If prevalence is 0.75, the prevalence odds is 0.75/(1 – 0.75) = 3. If prevalence is 0.20, the prevalence odds is 0.20/(1 – 0.20)=0.25
  - For small prevalences (eg. <0.1), the value of the prevalence and the prevalence odds become close, P ≈ ID
- P/(1 – P) = ID <=> P = ID/(1+ID)
- Usually measured for public health administration, but it also can be measured for causal inference.

  (eg.) The proportion of infants who are born alive with a defect of ventricular septum of the heart (see, https://www.mayoclinic.org/diseases-conditions/ventricular-septal-defect/symptoms-causes/syc-20353495) is prevalence. Measuring the incidence rate or risk of ventricular septum defects needs ascertainment of a population of embryos at risk and measurement of the defect's occurrence: Usually impossible to get such data

(Column) Prevalence of characteristics

- Prevalence measures status
- Sometimes used to describe the status of characteristics or conditions other than disease in a population
- (eg.) The proportion of a population that engages in cigarette smoking often is described as the prevalence of smoking
- The proportion of a population exposed to a given agent is often referred to as the **exposure prevalence**
  - The number of exposed to a given agent (E) among total population (N) gives the exposure prevalence (E/N)
  - Similarly, **exposure odds** can be defined as E/(N – E)
- Prevalence can be used to describe the proportion of people in a population who have brown eyes, type O blood, or an active driver's license. For causation, it is sometimes useful.