Epidemiology (6) Chapter 5 Types of Epidemiologic Studies (1): Experiment and cohort study

Minato NAKAZAWA, Ph.D.
<minato-nakazawa@umin.net>

Types of epidemiologic studies

- Epidemiologic studies: measurement exercises to obtain estimates of disease occurrence and effect measures (Chapter 4)
- Two main types of epidemiologic study
  - Cohort study
  - Case-control study
  - Other specific studies (two-stage design, ecologic study, …) → see, Modern Epidemiology (Chapter 6 for outline, PART II esp. Chapter 7-11, Chapter 30 for Ecologic study)
- Cohort studies
  - Cohort: Any designated group of individuals who are followed or traced over a period of time
  - Typical cohort study: Within the cohort which comprises persons with a common characteristic (exposure/ethnicity), measuring disease occurrence. Compare two cohorts (exposed/unexposed)
  - Following a cohort to measure disease occurrence, there are many complications
    - Who is eligible to be followed?
    - What should count as an instance of disease?
    - How the incidence rates or risks are measured?
    - How exposure ought to be defined?

John Snow's natural experiment (1)

- When cholera outbreak occurred in London in 1854, several water companies supplied piped water.
- At that time, mainstream physicians believed miasma theory (bad air causes disease) as the cause of disease.
- John Snow knew the fact that in the outbreak in 1848, the first two patients used the same room of the hotel, after the occurrence of the third patient lived neighborhood, the cholera outbreak rapidly expanded, but the physician treated the first two patients did not get sick. This fact doesn't fit miasma theory.
- Snow found the higher cholera occurrence in Surrey Building than neighboring Truscott's court in 1849, where residents used different water pumps, then concluded that the cause of cholera exists in drinking water.
- However, the authority of public health in London, Chadwick and Farr believed miasma theory. They claimed the difference of cholera occurrence in 1849 attributable to the worse air in Surrey Building. They suggested necessity of two comparable population with only difference in drinking water quality.

John Snow's natural experiment (2)

- https://johnsnow.matrix.msu.edu/work.php?id=15-78-C1
- In 1854 outbreak, both S&V and Lambeth company supplied drinking water to the people living in the south bank of Thames river.
  - At that time, S&V fetched source water from the downstream, but Lambeth fetched the source water from upstream of the Thames river.
    (cf.) https://www.rmg.co.uk/stories/topics/pollution-river-thames-history
- The mixing of the supply was the most intimate kind. The pipes of each company went down all the streets and into nearby all courts and alleys.
  - Snow identified the water company which supplied the drinking water to each household by checking water salt concentration. S&V supplied the water containing much more salt than that of Lambeth.
- **Residents whose water came from the S&V had an attack rate 5.8 (=0.0154/0.0027) times greater than that of residents from Lambeth**. The circumstance naturally created conditions that emulated an experiment, in which people who were otherwise alike in relevant aspects differed by their consumption of pure or impure water.

Table 5-1. Attack rate of fatal cholera among customers of the S&V and Lambeth, 1854

Water company	S&V	Lambeth
Cholera deaths	4093	461
Population	266516	173748
Attack rate	0.0154	0.0027

Types of experiments (1)

- Experiment: IR or R of disease in 2 or more cohorts is compared after assigning the exposure to the people who constitutes the cohorts. The reason for the exposure assignment is solely to suit the objectives of the study (has to obey the study *protocol*).
- Typical experiments (trial is a synonym of epidemiologic experiment)
  - **Clinical trials**: In clinical setting, those aim to evaluate which treatment for a disease is better. Comparison of the IRs or Rs in cohorts with different treatments. Usually treatment assignment is done by *randomization*. It enables to assume the same distribution of any background factors over the all cohorts. Table 5-2 shows better prognosis by zidovudine. https://www.cancer.org/cancer/managing-cancer/making-treatment-decisions/clinical-trials/what-you-need-to-know/phases-of-clinical-trials.html
    - Sometimes the subjects may not be treated as assigned, because they react poorly to an assigned medication or otherwise ignore their assigned treatment (compliance violation). Even so, the standard approach to analyze data is to follow the principle of *intent to treat* (ITT, see Chap.13).
    - If randomized trial is intended to study adverse effects of treatment, underestimating the magnitude of those effects is a larger problem. In trials aimed at safety of a new treatment, the drawbacks of ITT may outweigh any advantages. Data analysis should be done on actual exposure rather than

Table 5-2. Randomized trial for the risk of opportunistic infection in HIV patients given zidobudine treatment or placebo

Treatment	Zidovudine	Placebo
Opportunistic infection	1	7
Total patients	39	38
Risk	0.026	0.184

- (Box1) Natural experiments are not experiments because in natural experiments the subjects were not randomly assigned to any exposure. Rather, it's just a cohort study that simulates what would occur in an experiment. (p.73)
- (Box2) Experiment is not perfect. (p.75)

Types of experiments (2)

- **Field trials**: The study participants are not patients. The goal is primary prevention of a disease. (eg.) Experiments of new vaccines to prevent infectious illness. The largest formal human experiment ever conducted, the Salk vaccine trial of 1954, was a field trial. As the result, polio vaccination is conducted all over the world.
- **Community intervention trials**: Exposure is assigned to the group of people. (eg.) Water fluoridation in 1940s and 1950s. Introduction of home care on neonatal death (Table 5-3).
  (cf.) Fortmann SP et al. (1995) Community Intervention Trials: Reflections on the Stanford Five-City Project Experience, *American Journal of Epidemiology*, 142(6): 576–586, https://doi.org/10.1093/oxfordjournals.aje.a117678

Table 5-3. Neonatal death after 3 years community intervention trial for home care (39 villages) compared to usual care (47 villages)

Group	Home care	Usual care
Neonatal deaths	38	64
Number of births	979	940
Risk	0.039	0.068

Population at risk

- Snow's study on cholera defined 2 cohorts on water supply (S&V and Lambeth). Any person in either of these cohorts could have contracted cholera. Snow measured the rate of cholera occurrence among the people in each cohort.
- To understand which people can belong to a cohort, basic requirement for cohort membership (eligibility) has to be considered.
  - The members must be at risk for disease (But not necessarily healthy, Box3, p.77).
  - The members to be followed is "population at risk".
  - It implies that all members of the cohort should be at risk for developing the specific diseases being measured.
- Standard requirement
  - Everyone must be free of the disease being measured at the outset of follow-up.
  - Everyone must be alive at the start of follow-up.
  - Other requirements may not be simple.
    - Are people with measles vaccination included in population at risk for measles occurrence? (vaccination efficacy is not perfect)
    - Should men be considered part of the population at risk for breast cancer?
      - Solution: Treating male's breast cancer and female's as different disease.
- If the disease occurs only once in a person, the person who suffered from the disease is removed from population at risk. For recurrent diseases (like urinary tract infection), after getting the disease may remove the patients from population at risk temporarily, and include again after the recovery.

Example: Cohort study of vitamin A during pregnancy on cranial neural-crest defects

- Interviewed more than 22000 pregnant women early in their pregnancies (*Note: maternal recall bias is avoided*)
- Original purpose was to study potential effect of folate to prevent neural tube defects
- Based on same population, the effect of dietary vitamin A on cranial neural crest defects was evaluated.
- Women were divided into cohorts by the amount of vit.A in food and supplement.
- Table 5-4 showed the prevalence (actually risk) of these defects increased steadily and substantially with increasing intake of vit.A supplements by pregnant women.
- If 2 cohorts divided by 8000 IU/Day, RR is 3.05 (95%CI 1.81-5.16) by

```
library(fmsb); riskratio(16, 105, 1080, 21668)

prop.test(c(51, 54, 9, 7), c(11032, 10531, 754, 310))
```

  gives p-value < 0.001

Table 5-4. Prevalence of cranial neural-crest defects among the offspring of 4 cohorts of pregnant women by their vit.A intake during early pregnancy

Vit.A intake (IU/Day)	0-5000	5001-8000	8001-10000	>10000
Affected infants	51	54	9	7
Pregnancies	11083	10585	763	317
Prevalence	0.46%	0.51%	1.18%	2.21%

In USA, multivitamin supplements typically contain 2500–10000 IU vitamin A, often in the form of both retinol and beta-carotene. About 28%–37% of the general population uses supplements containing vitamin A. (https://ods.od.nih.gov/factsheets/VitaminA-HealthProfessional/)

* One whole baked sweet potato contains more than 20000 IU vit. A.

Closed and open cohorts

- Closed cohorts
  - Fixed membership
  - After it's defined and follow-up begins, no one can be added to a closed cohort.
  - The initial roster may dwindle as people in the cohort die, are lost to follow-up, or develop the disease (Fig. 5-1).
- Randomized experiments are examples of closed cohorts.
- Framingham Heart Study, began in 1949 and still ongo, is closed cohort study.

- Open cohorts
  - a.k.a. Dynamic cohorts
  - It can take on new members at time passes.
  - As shown in Fig. 5-1, size of dynamic cohort does not change.
- Cancer registry of Connecticut, USA is an example of open cohort.
  - The population at risk at any given moment comprises current residents of Connecticut (as people move into Connecticut, they are newly added to the registry).

Miscellaneous issues of cohort study (1)

- Counting disease events
  - IR and R are calculated by dividing the number of new disease events by the appropriate denominator.
  - Some disease onsets are excluded due to “not first occurrence”
    - Cancer in right breast after cancer in left breast
    - Second myocardial infarction
  - Reasons: Difficult to distinguish between new case and recurrence or exacerbation of an earlier case, recurrent case may have a different set of causes from initial case.
  - It’s possible to include second or subsequent recurrence, when first IR, second IR and following IR should be separately calculated. The population at risk of second event is only those who had first event.
- Measuring incidence rates or risks
  - From a closed cohort, IR and R can be estimated. Because of competing risks, population at risk is not constant in size over time, but ignored due to the period of follow-up being short.
  - In open cohort or when we have to consider competing risks due to longer observation period, IR rather than R should be estimated, using the denominator being person-time.

Miscellaneous issues of cohort study (2)

- Example: Cohort study of X-ray fluoroscopy and breast caner (Table 4-**7** in Chapter 4)
  - Due to the wide variety of follow-up periods, IRR was used (It's possible to calculate risks by lifetable)
- Exposure and induction time (Figure 5-2)
  - Hiroshima and Nagasaki cohorts who are survivors of atomic bomb (several closed cohorts with different radiation exposure levels, due to distance and shielding) were followed-up for decades. It's known that cancer requires considerable time to develop cancer: Leukemia does not occur until the induction period (and probably latent period) after radiation exposure has passed. Researcher is not sure what the induction time is for a given exposure and disease. Scenario-based reanalysis or statistical method is used to estimate the most appropriate induction time.
  - In Fig. 5-2, in exposed group, if we ignore induction period, IR is 3/(12+20+15+2+10)=3/59=0.051 yr^{-1}. In unexposed group, IR is 1/(20+18+20+11+20)=1/89=0.011 yr^{-1}. IRR is 0.051/0.011=4.5… However, if we consider the induction period of 3 years (the disease cannot occur due to the exposure within 3 years), IR(E)=2/(9+17+12+0+7)=2/45=0.044 yr^{-1}. In unexposed group, there is no reason to exclude first 3 years and IR remains 0.011 yr^{-1}, then IRR=0.044/0.011=3.96 Or, first 3 years of exposed group can be added to unexposed group because of no exposure effect during that period. Then IR(U) becomes 2/103, IRR becomes 2.29.
  - Many epidemiologists ignore it, or assume zero induction period.

Miscellaneous issues of cohort study (3)

- Eligibility criteria, exposure classification, and time loops
  - In a prospective cohort, the investigator selects subjects who meet eligibility criteria, then assigns them to exposure categories as they meet the conditions that define those. In the study of smoking, the subjects who meet age and other entry criteria may be invited into the cohort and then classified into appropriate category. If a person classified as nonsmoker in the beginning start smoking later, the person should be reclassified as smoker. To the contrary, when the smoker gives up smoking, the person is reclassified as ex-smoker.
  - In a retrospective cohort study, the decision about eligibility and any exposure categorization have to be based on information that is known at the time to which these decisions or assignments pertain, rather than later. If this rule is not kept, *time loop* occurs: A decision is made to include or exclude or classify a subject at a point in time before the information is known that the decision is based on.
  - Misclassification of the subject by time loop causes immortal person-time. If we classify workers into the categories of working years, 20+ years workers passed through other shorter categories. The earlier observation than 20 years of them should be considered as shorter categories. Otherwise, it constitutes immortal person-time.

Miscellaneous issues of cohort study (4)

- Retrospective cohort studies (a.k.a. historical cohort studies)
  - The cohorts are identified from recorded information. An example of young women in Florence in 15th and 16th centuries entered into dowry fund showed milder epidemic of plague later over a period of 100 years.
- Tracing of subjects
  - If the study trace less than 60% of subjects, it's regarded with skepticism. Even 70 or 80% are traced, if the loss to follow-up is related with exposure, the result is unreliable.
- Special exposure and general population cohorts
  - Cohort studies focus on people who share a particular exposure → special-exposure cohort studies (eg.) soldiers exposed to Agent Orange in Vietnam, residents of the Love Canal exposed to chemical wastes, SDA adhering to vegetarian diets, atomic bomb survivors. Female offspring of women who took DES is special-exposure cohort.
  - Cohort studies focus on common exposure → general-population cohort studies (eg.) birth defects in pregnant women in relation to vit.A consumption (consumption levels were not used as eligibility criteria). Secondhand smoke or dietary intake of saturated fat may be common exposures, thus they are general-exposure cohort.