Chapter 13. Epidemiology in Clinical Settings

February 1, 2024
Epidemiology (15)

Minato Nakazawa
<mailto:minato-nakazawa@people.kobe-u.ac.jp>

Focus of clinical epidemiology

- Application of epidemiologic principle on questions related to diagnosis, prognosis, and therapy.
- Screening and preventive medicine in population and individuals.
- Pharmacoepidemiology largely affected therapeutic thinking, and lead to outcomes research and comparative effectiveness.
  - Outcomes research: Epidemiologic methods combined with clinical decision theory to determine which therapeutic approaches are the most cost-effective
  - Comparative effectiveness: The effect of different interventions against one another in a variety of settings.
- Diagnosis
  - The process of diagnosis may appear to involve intuition, conviction, guesswork, processes that are opaque to quantification and analysis.
  - Formal approaches to understanding and refining the steps have helped to clarify the foundation for diagnostic decision making.
  - Formulating a diagnosis is based on the data from signs, symptoms, and diagnostic test results that distinguish the patient from nonpatient.

The Gold Standard

- Diagnosis is not perfect. Any sign or symptom or combination of those rarely distinguish completely between those with and those without a disease (overlapping exists).
  - Diagnosis seems to be established when a specific combination of signs and symptoms posed as the criterion for disease is present.
  - A diagnosis meeting this standard may be “definitive”.
  - Another definitive diagnosis can be done by expert’s judgment.
- Two different approaches to the same disease will not necessarily lead to the same classification for every patient.
- Anyway, even if arbitrary, “gold standard” is needed to judge the findings.

Sensitivity and Specificity (1)

- For long years, diagnosis of TB was the detection of Mycobacterium tuberculosis from smears of acid-fast bacilli and from culture. Lower detection limit was 10000 bacteria/mL.
- Catanzaro et al. (JAMA, 2000) investigated how well smear predicted the diagnosis of clinical TB (Table 13-1). [https://doi.org/10.1001/jama.283.5.639]
  - Clinical TB diagnosis is used as gold standard.
  - How many of clinical TB patients show positive smear = 43/72 = 60% (**Sensitivity**)
  - How many of those who don't have TB show negative smear = 244/266 = 92% (**Specificity**)
- A test with 100% sensitivity and 100% specificity would be positive for everyone with disease and negative for everyone without disease. It's ideal test. However, usually no such test exists.

Table 13-1. Distribution of patients with suspected active pulmonary tuberculosis, by diagnosis and by results of acid-free bacillus smear testing.

	Tuberculosis		
Smear	Present	Absent	Total
Positive	43	22	65
Negative	29	244	273
Total	72	266	338

- (Note) As in the case of RT-PCR test for COVID-19, if the test itself is used as definitive diagnosis, its performance cannot be evaluated.
  - To assess the performance of such test method, alternative definitive diagnosis is necessary such as including clinical symptoms or fever or oxygen saturation to determine the patients and controls by "gold standard method".

Sensitivity and Specificity (2)

- Combination of 2 tests can be considered.
  - Test A and B (both sensitivity 80%, specificity 90%), are to be independent each other.
  - If “positive for both A and B” is used to judge the patient as positive, sensitivity becomes the multiplication of each 80%, thus 64%.
  - In this case, negative means “negative in either A or B”. Specificity will improve. In A, 90% of non-patients was correctly judged as negative. In B, among remaining 10%, 90% is additionally judged as negative. As a result, 0.9+(0.1x0.9)=0.99 (99%) specificity can be achieved.
  - The reverse occurs if a positive result on either test is taken to indicate the presence of disease. Sensitivity is 0.8+0.2x0.8=0.96. Specificity is 0.9x0.9=0.81.
- (eg.) Pap smear test to detect cervical cancer has high sensitivity and low specificity. A sequence of repeated Pap smear can improve specificity. However, single smear combined with relatively newly developed DNA detection of HPV gives much better performance.
- (Note) If the test has continuous value to be judged as positive by some cutoff value, ROC (Receiver-Operating-Characteristics) analysis is conducted to determine an optimal cutoff for the best point showing higher sensitivity and specificity as possible.
  - In R, ROC() of Epi package, roc() of pROC package and roc() of fmsb package can automatically detect an optimal cutoff value and calculate AUC (Area Under the Curve), of which higher value (close to 1) means higher performance of the test.

Predictive Value

- Sensitivity and specificity indicate the performance of the test to classify those who have and don't have a disease.
- However, even if both sensitivity and specificity are high, there may be many false-positive under low prevalence setting. The indicator to show the proportion of correct judgment among positives is predictive value positive (PV+). Similarly, the proportion of correct judgment among negatives is predictive value negative (PV-).
  - In Table 13-1 (prevalence is 72/338 = 21%), PV+ = 43/65 = 66% and PV- = 244/273 = 89%.
  - Under the condition of much lower prevalence (72/838 = 9%, Table 13-2), PV+ = 43/106 = 41%, PV- = 703/732 = 96%.
- Sensitivity and specificity are not affected by prevalence. The data in Table 13-2 give almost same sensitivity and specificity as Table 13-1 (Sensitivity is completely same because the data for TB patients are same, and specificity is 703/766 = 92%, almost same as 90%).

Table 13-2. Results from Table 13-1 augmented with data from 500 additional people without TB.

	Tuberculosis		
Smear	Present	Absent	Total
Positive	43	63	106
Negative	29	703	732
Total	72	766	838

- If the screening is conducted in the population with no disease, PV+ = 0, PV- = 100%. Alternatively, if the screening is conducted in the population of everyone in which has disease, PV+ = 100%, PV- = 0.
- (Note) One of the reasons why screening test for breast cancer is not conducted for young women nor men is the low prevalence of breast cancer in them.

Screening and its bias (1)

- For many diseases, screening is expected as the measure of early detection. However, many reasons including cost, actual implementation of screening test is limited.
- (Note) Principles for screening are suggested by WHO (Wilson and Jungner, 1968; Andermann et al., 2008), as shown later.
- Lead-time bias
  - Screening advances the date of diagnosis of the disease.
  - Comparing with the situation when screening is not done, the period from detection to the natural occurrence of symptoms (the lead-time) is added to survival time. It make the survival time apparently longer.
  - If only detection is done by screening, with no treatment, **the lead-time can be assessed by the comparison of survival times between screened and not screened groups**, but it's not allowed ethically.

Screening and its bias (2)

- Prognostic selection bias
  - Another bias comes from self-selection of subjects who decide to be screened.
  - Since screening programs are voluntary, those who are willing to be screened are more health conscious than those who refuse to be screened.
  - To avoid this bias, the subjects to be screened or not screened should be randomly assigned (but it's difficult to force the people to be screened).
  - In non-experimental studies, this bias should be considered.
- Length-biased sampling
  - There is natural variability in the progression rate of disease.
  - Screening is done during the latent (asymptomatic) period.
  - The patients with slow progression rates have longer latent period than those with fast progression rates.
  - Slow progression patients have higher chance of screening than the fast progression patients. It's length-biased sampling.
  - Patients with slow progression rate have longer survival time even if the early detection by screening had no effect to make survival time longer.

(Note) 10 Principles of Screening (Wilson and Jungner, 1968)
https://apps.who.int/iris/handle/10665/37650

- The condition sought should be an important health problem.
- There should be an accepted treatment for patients with recognized disease.
- Facilities for diagnosis and treatment should be available.
- There should be a recognizable latent or early symptomatic stage.
- There should be a suitable test or examination.
- The test should be acceptable to the population.
- The natural history of the condition, including development from latent to declared disease, should be adequately understood.
- There should be an agreed policy on whom to treat as patients.
- The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole.
- Case-finding should be a continuing process and not a “once and for all” project.

(Note) Synthesis of emerging screening criteria proposed over the past 40 years (Andermann et al., 2008) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2647421/

- The screening programme should respond to a recognized need.
- The objectives of screening should be defined at the outset.
- There should be a defined target population.
- There should be scientific evidence of screening programme effectiveness.
- The programmes should integrate education, testing, clinical services and programme management.
- There should be quality assurance, with mechanisms to minimize potential risks of screening.
- The programme should ensure informed choice, confidentiality and respect for autonomy.
- The programme should promote equity and access to screening for the entire target population.
- Programme evaluation should be planned from the outset.
- The overall benefits of screening should outweigh the harm.

PROGNOSIS

(https://www.nejm.org/doi/full/10.1056/nejm200001203420301)

- Prognosis: Qualitative/quantitative prediction of the outcome of an illness.
- Simplest epidemiologic measure of prognosis is case-fatality risk (CFR: most textbook including this refers “rate”, but I recommend to use “risk”).
  - The proportion of people with newly confirmed cases who die from the disease.
  - Only used over a fixed and stated time period, such as 3 months or 12 months.
  - Usually used for acute infectious diseases, such as typhoid fever’s CFR is 1%, paralytic polio’s CFR is 5%, ebola disease’s CFR is 75%.
  - (Note) Seasonal influenza’s CFR is 0.01-0.1%, Spanish flu’s CFR was >=2.5% [https://wwwnc.cdc.gov/eid/article/12/1/05-0979_article]. 2019-nCoV’s CFR is estimated as 2-3% in early 2020 [https://www.who.int/news-room/detail/23-01-2020-statement-on-the-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)].
  - The presumption of CFR is that essentially all of the deaths that occur promptly after disease onset are a consequence of the disease.
- For diseases with a long clinical course, CFR is difficult to apply. Instead, 5 year survival rate (it’s also not rate, the proportion of patients surviving for 5 years after diagnosis) is used. Common method to obtain a survival curve is Kaplan-Meier product-limit method. Example is Figure 13-1 (see, the graph shown right).

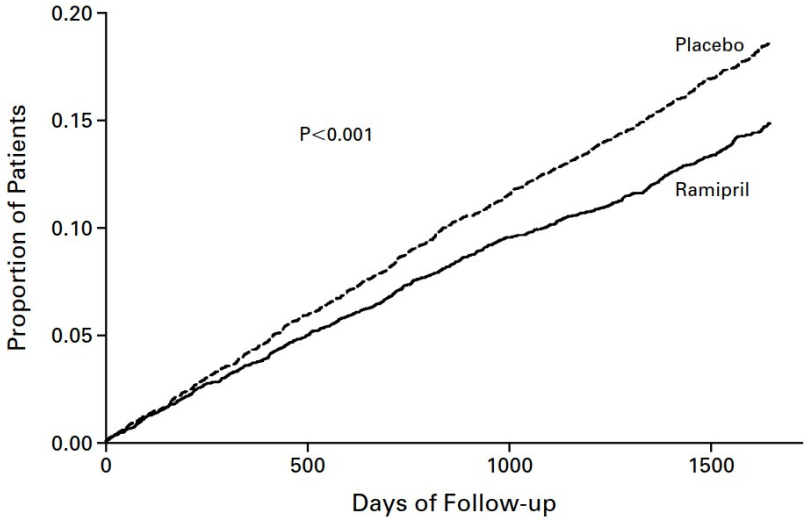


Figure 1. Kaplan–Meier Estimates of the Composite Outcome of Myocardial Infarction, Stroke, or Death from Cardiovascular Causes in the Ramipril Group and the Placebo Group.

The relative risk of the composite outcome in the ramipril group as compared with the placebo group was 0.78 (95 percent confidence interval, 0.70 to 0.86).

Table 3. Incidence of the Primary Outcome and of Deaths from Any Cause.

Outcome	Ramipril Group (N=4645)	Placebo Group (N=4652)	Relative Risk (95% CI)*	z Statistic	P Value†
	no. (%)				
Myocardial infarction, stroke, or death from cardiovascular causes‡	651 (14.0)	826 (17.8)	0.78 (0.70–0.86)	−4.87	<0.001
Death from cardiovascular causes§	282 (6.1)	377 (8.1)	0.74 (0.64–0.87)	−3.78	<0.001
Myocardial infarction§	459 (9.9)	570 (12.3)	0.80 (0.70–0.90)	−3.63	<0.001
Stroke§	156 (3.4)	226 (4.9)	0.68 (0.56–0.84)	−3.69	<0.001
Death from noncardiovascular causes	200 (4.3)	192 (4.1)	1.03 (0.85–1.26)	0.33	0.74
Death from any cause	482 (10.4)	569 (12.2)	0.84 (0.75–0.95)	−2.79	0.005

*CI denotes confidence interval.

†P values were calculated with use of the log-rank test.

‡In the substudy, 34 of 244 patients (13.9 percent) assigned to take a low dose of ramipril (2.5 mg per day) reached the composite end point, as compared with 31 of 244 assigned to take 10 mg of ramipril per day (12.7 percent) and 41 of 244 assigned to placebo (16.8 percent). The inclusion of the data from the low-dose group did not change the overall results (relative risk of the primary outcome, 0.78; 95 percent confidence interval, 0.70 to 0.86).

§All patients with this outcome are included.

Kaplan-Meier method

Let the times of event happening since beginning of time at risk $t_1, t_2, ...$, the numbers of events at each time $d_1, d_2, ...$, and the size of population at risk just before the each time $n_1, n_2, ...$ The size of population at risk decreases not only by the event occurrence but also by censoring such as moving out or loss to follow up or death by competing risks. When the censoring and event occurred at the same time, usually the censoring occurred just after the event happening.

Here the Kaplan-Meier's product-limit estimates $\hat{S}(t)$ can be defined as follows.

$$\hat{S}(t) = (1 - d_1/n_1)(1 - d_2/n_2)... = \prod_{i<t}(1 - d_i/n_i)$$

Clearly this value means the probability of survival and is numerically 1 at first (nobody has experienced the event) and 0 in the end (after the everybody experienced the event).

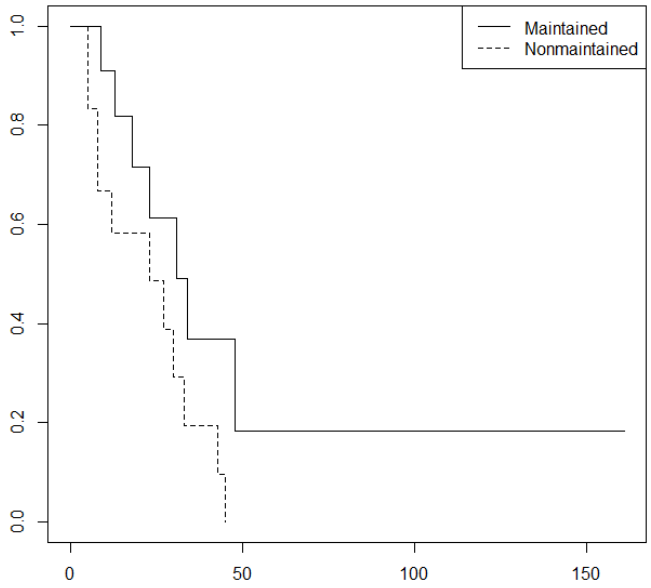
The standard error of $\hat{S}(t)$ is given by the Greenwood's formula shown below.

$$\mathrm{var}(\hat{S}) = \hat{S}^2 \times \sum_{i<t} \frac{d_i}{n_i(n_i - d_i)}$$

Estimated $\hat{S}(t)$ is usually plotted as survival curve with 95% confidence intervals[*33].

In R, it can be applied using survfit() of survival package.

```
> library(survival)
> data(leukemia)
> plot(survfit(Surv(time, status) ~ x, data=leukemia), lty=1:2)
> legend("topright", lty=1:2, legend=c("Maintained", "Nonmaintained"))
```



Clinical trials

https://www.cancer.org/treatment/treatments-and-side-effects/clinical-trials/what-you-need-to-know/phases-of-clinical-trials.html

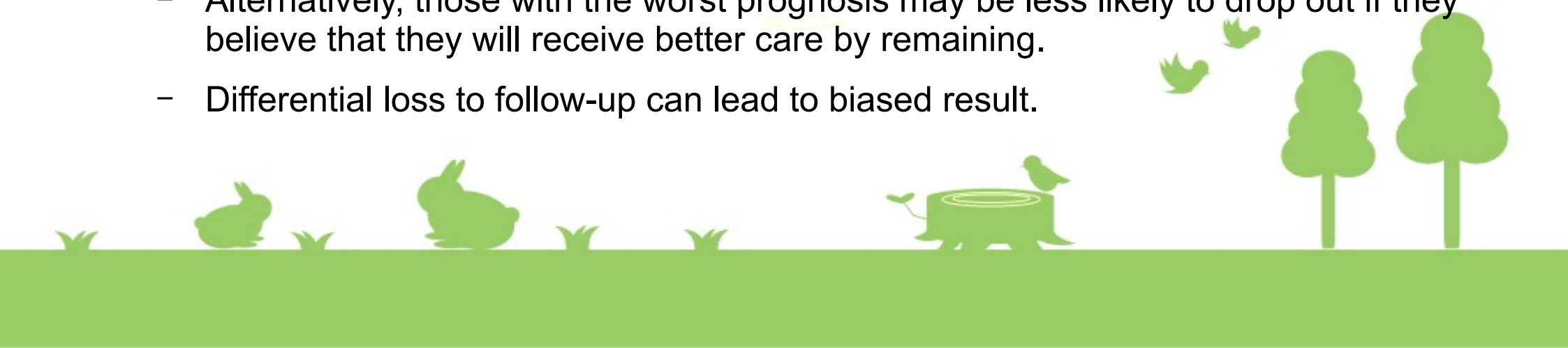
- Phases to get the new drug approval by the government (PMDA in Japan, FDA in USA)
  - 0: Exploring if and how a new drug may work
  - 1: Is the treatment safe?
  - 2: Does the treatment work?
  - 3: Is it better than what's already available?>**RCT (Randomized Controlled Trial)**
    → After phase 3, government official approves the new drug if it's better
  - 4: What else do we need to know?>**Post-marketing trial**
- RCT is the centerpiece of clinical epidemiology, which is considered to show the highest level evidence.
- RCT's advantages over nonexperimental studies
  - Effective control of confounding
  - Confounding by indication (due to inherent differences in prognosis between patients given different therapies, highly effective but costly and with common adverse effect new drugs may be more prescribed for the patients who face high risk of death than for the patients with lower risk of death) may inevitably occurs in nonexperimental study, but is avoidable in RCT.

Use of placebos in clinical trialls

- Blinding and use of placebos
  - Blinding: Hiding information about treatment assignment from the key participants in a trial.
  - Since the physician's decision about hospitalization is affected by knowledge about which treatment was assigned to a given patient, the physician should be blinded for assignment.
  - Since the patient's compliance and recovery may be affected by the knowledge, the patient should be blinded.
  - Sometimes double (both physician and patients) blind or even triple (including administers) blind is conducted.
  - One method to facilitate blinding is using placebo. Placebo pills typically contains sugar, with the same shape and color with therapeutic drug.
- Ethics of placebo use in randomized trials (box, p.244)
  - Before, it was common to prescribe placebos so that patients could benefit from improved expectations.
  - Today, such practice is rare because it's unethical. Instead, new drug is compared with the standard drug ever prescribed due to Declaration of Helsinki (and WMA Declaration of Lisbon on the Rights of the Patient: https://www.wma.net/policies-post/wma-declaration-of-lisbon-on-the-rights-of-the-patient/).

Threats to validity in trials (1)

- Incomplete Follow-up
  - Differential follow-up of the treatment groups may occur.
  - Ideally there to be no subjects lost to follow up.
  - Usually some patients are not followed to the end.
  - Reasons of lost to follow-up are the same as other cohort studies.
  - To deal with this potential source of bias, investigators may analyze the data under the assumption that the experience of those who were lost to follow-up is similar to the of those who remained, though this assumption is not always reasonable.
  - Alternatively, those with the worst prognosis may be less likely to drop out if they believe that they will receive better care by remaining.
  - Differential loss to follow-up can lead to biased result.



Threats to validity in trials (2)

- Intent(ion)-to-Treat Analysis (ITT)
  - As described in Chapter 5, ITT is often employed.
  - In ITT, the patients are classified by assignment, regardless of whether they adhered to that assignment or not.
  - This approach maintains the benefits of random assignment for the comparison of a new treatment against an older treatment, but at cost of misclassification of actual treatment.
  - To the extent that the misclassification is independent of the study outcome, the misclassification will be nondifferential and will lead to underestimation of the effect of actual treatment.
  - Underestimation of the actual treatment effect is often considered acceptable.
  - See also
    - https://dx.doi.org/10.5811%2Fwestjem.2017.8.35985
    - https://www.youtube.com/watch?v=mZp2KomA3Ws

Threats to validity in trials (3)

- Confounding Imbalances
  - Baseline risk factors are prognostic, measured at the time of random assignment.
  - If randomization succeeds, the frequency of the those factors will be similar in the various treatment groups – balanced between groups.
  - However, by chance, by comparing the distribution between groups, imbalance may be proved to exist.
  - Any imbalance in a baseline risk factor represents confounding.
  - As shown in Chapter 7 (example of University Group Diabetes Program, Table 7-7 and 7-8), randomly assigned but age distribution was imbalanced.
  - Distributions of baseline factors are rarely identical, so how can we tell when the imbalance in a baseline risk factor is severe?
  - The best way to assess the confounding is to use the same approach that epidemiologists use in other situations: Comparison of crude estimate with unconfounded estimate.
  - **A common mistake is to use statistical significance testing to assess imbalances in baseline risk factors**.

Threats to validity in trials (4)

- An unrejectable null hypothesis (box, p. 248)
  - Statistical significance testing is to test the null hypothesis that observed difference is only due to by chance.
  - In the randomly assigned data, we have already known that the difference is purely due to "by chance". Thus such null hypothesis is unrejectable.
  - That's why statistical testing makes no sense.
- Example: The Alzheimer's disease cooperative study of selegiline and alpha-tocopherol
  - As baseline, MMSE mean score was 11.3 and 13.3 in alpha-tocopherol group and placebo group, respectively.
  - If this difference is ignored, RR was 0.7, after adjusting this difference, RR was 0.47.
  - Correspondent criticized the existence of baseline difference (by statistical test) and "no true effect of treatment has been proved". But the correspondent was wrong.

Pharmacoepidemiology

- Active area of research
  - Pharmaceutical companies invest huge money for this area
- Postmarketing surveillance
  - Japan's PMDA guideline
    https://www.pmda.go.jp/files/000226080.pdf
  - FDA encourages the voluntary reporting of suspected adverse drug effects (*spontaneous reports*), but it's difficult to interpret by at least 2 reasons.
    https://www.fda.gov/vaccines-blood-biologics/biologics-post-market-activities/postmarketing-clinical-trials
    - Only a small, but unknown proportion of suspected adverse drug effects are reported – serious events are more likely reported than mild events, but even so, only a part of adverse events are reported
    - By knowing the all numbers in 2 x 2 table of exposed (taking drug) – unexposed (didn't taking drug) and disease (adverse event) – health (no adverse event), it's possible to evaluate the excess occurrence of adverse events, but the spontaneous reports only include the one cell (disease in exposed) of possible 4. Therefore, in principle, it's impossible to assess the extent of excess occurrence of adverse event by exposure.

When the disease definition includes an exposure (box, p.251)

- Disease definitions may refer to exposures.
- If a clear understanding of causal relation exists, it's possible to refine the definition to reflect this insight.
- If "insight" is only presumption, the disease definition has to be independent from exposure, but many disease definitions are such cases.
  - Analgesic nephropathy, Asbestosis, Berylliosis, Food poisoning, Frostbite, Heatstroke, Hypervitaminosis D, Iron-deficiency anemia, Motion sickness, Protein-calorie malnutrition (Protein-energy malnutrition), Radiation sickness, Silicosis, Smoker's cough, Strep throat, Tennis elbow, Tuberculosis, and so on.
  - Most infectious diseases (syphilis, malaria, influenza, and so on) could also be included.

HEALTH OUTCOMES RESEARCH

- Related to pharmacoeconomics
- New research area
- Medical research typically focuses on a primary endpoint (survival or disease occurrence). Therapeutic evaluations based on narrowly defined endpoints are criticized as not considering patient's overall QOL as a result of therapeutic outcome and unintended effects of treatment.
- The economic costs of treatment are paid by patients (out-of-pocket payment) and/or insurers and/or government.
- Therapies offering desirable results for patients at costs that are attractive to patients or only relative to the therapeutic alternatives. Using meta-analysis, cost-effectiveness analysis, decision analysis, and sensitivity analysis in addition to more traditional epidemiologic analysis, health outcome research seeks to find such desirable therapies.
- See, Petitti DB (2000) Meta-analysis, Decision Analysis and Cost-effectiveness Analysis. OUP. DOI:10.1093/acprof:oso/9780195133646.001.0001