코호트 통계기법의 변화

Ho Kim
School of Public Health,
Seoul National University

Mainly based on the article "Cohort Study" by Ross L. Prentice in the Encyclopedia of Epidemiological Methods(2000), Wiley, Edited by Gail and Benichou and on the article by Tager (1998)

Two categories by Tager*

1. Life table-type
2. Longitudinal
   * life table style
   - closely related to life-table methods
   - exposure, and person-time
   - (cumulative) incidence and their ratios
   -inferences are restricted to pop'n average effects

*Tager, Outcomes in Cohort Studies, Epidemiologic Reviews (1998) 20,1 15-27

Two categories by Tager*

* longitudinal cohort study
  - repeated measures characteristics
  - inferences on pop'n average effects and individual heterogeneity, changes in processes over time, and transitions between states of health and disease

Types of outcomes for cohort

Discrete events
  Single events
    Mortality
    First occurrence of a disease or health-related outcome
      Incidence (density)
      Cumulative incidence (risk)
      Ratios (incidence density and cumulative incidence)
  Multiple *occurrences:*
    Of disease outcome
    Of transitions between states if health/disease
    Of transitions between functional states

Level of a marker for disease or state of health

Change in a functional/physiologic/biochemical/anatomic *marker for disease or health*
  Rate if change
    Patterns of growth and/or decline
    " *Tracking* " of markers of disease/health
  Change in level with time (age)

Characteristics of "life-table-type" and "longitudinal" cohort studies

Characteristic	Comment
Subject selections	
Life table-type	
Exposures status at the inception of cohort	Need to assure that each exposure group is *representative* of those with similar exposure in the *target* population for the cohort
Longitudinal	
Exposure status at the inception of cohort	Same as for life-type cohort study
Can be unrelated to status on single, explicit exposure ("natural history" study)	Subjects should be representative of target population in general Target population *could* be subjects with defined *health* condition(s)
Time	
Life table-type	Cumulation of person-time to estimate cumulative incidence *and/or* average incidence *density* over the entire period of *follow-up* and/or specific age or time epochs encompassed by the cohort
Summarize	Typical *formulation* of time-dependence retains features of group person-time summarized over subintervals of follow-up
Longitudinal	
Summarize	Same as for life table-type *cohort* study
Effects evaluated explicitly	Can separate cohort and age effects Can evaluate/model time-related associations between observations over time

Exposure	
Life table-type	
"Baseline"	Exposure classification based on initial entry into the cohort ("baseline") often
Summarize/categorize	used to *characterize* the exposure *experience* over the entire follow-up of the cohort Updated exposure *summarized* into cumulative *measures* (continuous or ordinal), state measures (e.g. ,none, current, past), or combinations Averages *exposure* over the period or subperiods of obsevation
Longitudinal	
"Baseline"	Same as for life table-type cohort study
Summarize/categorize	Evaluate effects of *exposure* in addition to effects *attributed* to exposures
Time-*dependence* of more explicitly *formulated*	*accumulated* up to the time of *cohort* inception ("baseline") Distinguish effects on individuals from population average (cross-sectional) effects Evaluate exposure-effect association at any given time point in history of subjects Evaluate current exposure *conditional* on past exposure *profiles*

Outcome	
Life table-type	
The first recognized occurrence of a disease or other health-related state	Not well suited to outcomes that are "continuous" Such outcomes often reformulated into categorical (state) outcomes
Overall and/or cause-specific mortality	Measures: incidence density, cumulative incidence, and their respective *ratios* for "*exposed*" and "nonexpected"
Longitudinal	
The first *recognized* occurrence of a disease or other health-related state	Same as for the life-type cohort study
Overall and/or cause-specific mortality	Description of "*natural* history" of a health condition
Rate of change (growth, decline)	Description of physiologic processes that are precursors to discrete health conditions
Transitions between states of health	Eliminates restriction to first *occurrences* of health conditions
Tracking	
Inference	
Life table-type	
Population or *group* average effects	*Inference* properly limited to *groups* and not individuals
	Cumulative incidence refers to average individual risk and not relevant to any specific individual
Longitudinal	
Population or *group* average effects	Same as for life type cohort study
Inference at the level of individuals	Natural history *refers* to description of changes over time: in physiologic/biologic markers of disease pathogenesis; in expression of particular manifestation(s) of disease or general markers of health; in the relation between exposure and *outcome*; in growth and decline of anatomic/*functional*/physiologic characteristics with general health implications
Natural history of status of health and disease and important physiologic processes	

Table3. Examples of outcomes and exposures in two cohort studies analyzed as "life table-type cohort studies"

Health Professions Follow-up Study,1993	
Measure	Implementaion
Outcome	
Cumulative 4-year incidence	Person-years not assigned on age-specific basis
Relative risk (cumulative incidence) inferred from incidence rate ratios	Incidence rate ratio interpreted as relative risk across exposure categories Summary measure over 5-year age groups-appears to be age at last examination
Exposure	
Vitamin E, vitamin C, β-carotene intake from foods and supplements	Summed over the 3 years and categorized as quintiles of intake/day
Accessed on three occasions, 1986, 1988,1990	1. Table2: From food 2. Table2: From supplements 3. Table3: Years if supplement use

The "longitudinal" cohort study

- Estimation of the effects at the individual level, individual heterogeneity, cross-sectional versus longitudinal inference
- Estimation of the effects of risk markers over time on disease outcome
- Estimation of rates of change or change in level for outcomes that relate to disease natural history
- Natural history of states of health with multiple occurrences or that oscillate between states
- Separation of cohort or period effects from the effects of age or calendar time

Basic Cohort Study Elements

- Exposure Histories and Disease Rates
- Cohort Selection and Follow-Up
- Covariate History Ascertainment
- Disease Event Ascertainment
- Data Analysis

$$\lambda[t;Z(t)] = \lambda_0(t)\exp[z(t)^T\beta]$$

Cohort Study power and sample size, References

- Gail (1974) Biometrics 30, 231-37
- Casagrande et al (1978) Biocs 34, 483-6
- Fleiss et al (1980) Biocs 36, 343-6
- Whittemore (1981) JASA 76, 27-32
- Brown and Green (1982) AJE 115, 752-8
- Greenland (1985) J. Chronic diseases 39, 117-27
- Self et al (1992) Biocs 48, 31-9

Cohort Study power and sample size, Formula

$$n = [p_2(1-p_2)]^{-1}(\log\lambda)^{-2}Q$$

p_1 Prob of disease for exposed

p_2 Prob of disease for unexposed

$\lambda = p_1(1-p_2)/[p_2(1-p_1)]$ OR

$Q = [\gamma(1-\gamma)]^{-1}\{W_{\alpha/2} - W_{1-\eta}[\gamma + \lambda^{-1}(1-p_2+\lambda p_2)^2(1-\gamma)]^{1/2}\}$

γ exposure fraction at baseline

$W_{\alpha/2}, W_{1-\eta}$ upper $\alpha/2$ and $1-\eta$ percentile of N(0,1)

Cohort Study power and sample size, formula

Prentice (Sta'l Methods in Medical Research 1995;4: 273-92) displayed selected power calculation developed in Women's health initiative (JAMA 1995, 50-55)

Currently enrolling 100,000 post-menopausal American women in the age 50-79.

Measurement error -> loss of power

Cohort Study power and sample size, practical examples

- One approach

1. List hypothesis
2. Select cohort size that will yield acceptable power for all or most within a practical follow-up period

- Empirical approach 1
- Cardiovascular disease: 5,000-20,000
- Framingham Study, MRFIT (Multiple Risk Factor Intervention Trial; JAMA 1982;248:1465-77)

Cohort Study power and sample size, practical examples

- 5,000 for older persons, as in the Cardiovascular Health Study (Annals of Epi 1991;1:263-76) age 65 or older
- larger cohort size for younger persons' as in Royal College of General Practitioners' study of the health effect of the oral contraceptive use: 46,000 subjects

Cohort Study power and sample size, practical examples

- Empirical approach 2
- Diet and cancer: 50,000-100,000
- Nurse Health Study, the Canadian National Breast Screening Study, the Iowa Women's study
- Measurement error problem

Additional sampling strategies

- pop'n based case-control study
- Hospital based case-control study
- Case-cohort study
- Nested case-control study

Hospital based case-control study

- Cases: given disease, in a given hospital(s)
- Controls: pts with other diseases from those hospitals
- Pros against pop'n based c-c

1. More pts will agree (e.g.blood): reduce selection bias (nonresponse bias)
2. Reduce recall bias (similar quantity)

Population based case-control study

- All cases can be enumerated fin well-defined pop'n
- Controls: random samples w/o diseases
- Relative risk and absolute risk can be estimated (using additional information on source pop'n)

Case-cohort study

- All cases in the cohort
- Subcohort: random sample of the cohort
- Useful when

1. Too expensive to collect and process all covariate inf'n for all subjects.
2. Covariate inf'n at entry is banked and multiple disease status are of interest

Case-cohort study

- At the beginning of the study, all subcohorts' members analyzed
- Information on these cases can be processed as time passes
- Since subcohort data are not dependent on the disease, standard statistical analysis can be performed

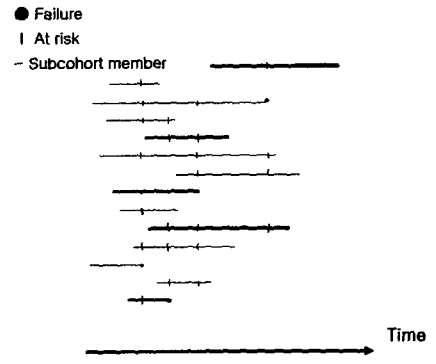

Figure Prentice pseudo-likelihood approach to the analysis of case-cohort data. Pseudo-likelihood contributions are conditional probabilities based on the case and the subcohort members at risk at the failure time

Case-cohort study
E.g. Study of Lung Cancer Mortality in Aluminum Production Workers in Quebec, Canada, AJE 1994;139: 150-62

- 16,297 men who had worked at least one year in manual jobs at a large Al production plant 1950-1988, coal tar pitch
- 338 lung cancer death identified
- Subcohort of 1138 subjects within year-of-birth strata -> 205 cases in subcohort
- Work and smk history for subcohort and 133 non-subcohort cases
- Lung cancer-coal pitch association observed

Nested Case-control study

- random sampling from eligible controls fro each case
- w/o possibility of selection or information bias unlike case-control study
- Closely related to pop'n based matched case-control study
- Restricting controls will typically result in biased estimation

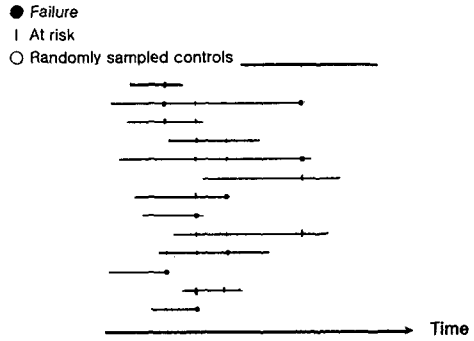

Figure Cohort of 15 subjects. Each failure determines a risk set. A single control is sampled for each case to form a one-to-one matched nested case-control study

Nested Case-control study
E.g. Occupational Cohort Study of TCDD Exposure and soft tissue Sarcoma(STS) and non-Hodgkin's lymphoma(NHL);
Lancet 1991;338: 1027-32

- 21,183 workers exposed to phenoxy herbicides, chlorophenols, and dioxins
- SMR=1.96 for STS and 1.29 for NHL
- 11 STS, 32 NHL cases, five controls from the same country, gender, year of birth
- 83% efficiency relative to the analysis of entire cohort

Comparison: Case-cohort study and nested case-control study

- Two main approaches to sample from assembled cohort studies
- Nested case-control: retrospective, time matched controls after the outcome occurs
- Case cohort: prospective and unmatched

Concluding remarks

- Careful plans in advance
- Hypotheses
- Design
- Analysis
- Post hoc
- Cohort studies properly play central role in epidemiologic research