

## 외국의 코호트 연구 현황

서울대학교 보건대학원  
조성일

## Cohort Studies Around the World

- In Medline 2002 database, 1941 abstracts [for English articles] included the term “cohort study.”
- In contrast, 1672 abstracts [for English articles] included “case-control study” or “case-referent study.”

## Cohort Studies with a Name

- From 600 “cohort study” abstracts dated for 6 months (Medline 2002 database), 65 distinct cohort names were identified.
- 65 named cohorts
  - 16 USA
  - 4 Japan
  - 4 Australia
  - 4 UK
  - 33 Other (Mostly Europe, plus New Zealand, Iceland, etc)

## Types of Cohort Studies (1)

- Time
  - Prospective vs. Retrospective
- Intervention
  - Observational vs. Experimental

## Intervention Cohorts

- Many trials need a large sample size, especially prevention trials
- Observational cohort study can be nested or combined
- Examples
  - Physician's Health Study (since 1982, N=22,071)
    - Treatment: aspirin and beta-carotene for prevention of cardiovascular disease and cancer
    - Nested observational studies
      - *Albert et al, Arch Int Med 2002; Nut consumption and decreased risk of sudden cardiac death in the Physicians Health Study.*
  - Women's Health Initiative Study (since 1991)
    - Clinical Trial (N=67,000)
    - Observational Study (N=100,000): ineligible/unwilling for trial
    - *Pradhan et al, JAMA 2002; Inflammatory Biomarkers, Hormone Replacement Therapy, and Incident Coronary Heart Disease: Prospective Analysis From the Women's Health Initiative Observational Study*

## Intervention Studies: Knowledge into Practice?

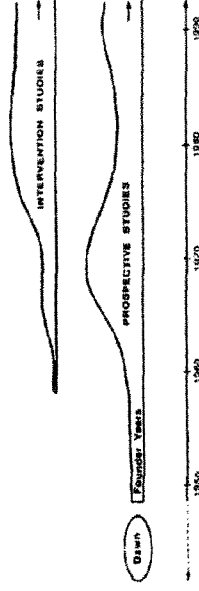


Figure 1. Cardiovascular disease epidemiology historical development. From: Epstein: Circulation, Volume 93(9), May 1, 1996:1755-1764

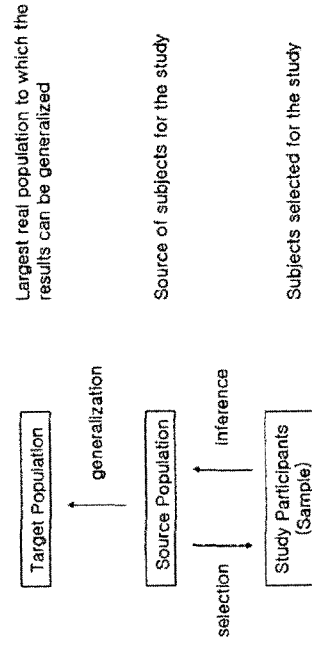
- For any risk factor, experimental evidence is deemed the best.
- Due to practical limitations (sample size, drop-out, etc), observational studies cannot be replaced.

## Types of Cohort Studies (2)

- Exposure
  - Atomic-bomb survivor cohort
  - PCB-exposed children cohort
  - Vietnam veterans cohort
  - Some occupational cohorts
- Disease
  - Disease as an outcome (cancer cohorts, Framingham Heart Study, etc)
  - Disease as a cohort definition (patient cohort: DM, ESRD)
- Source population (Natural or social grouping)
  - Birth cohort (esp. to study childhood diseases, genetic disease, etc)
  - Community cohort (defined by area of residence)
  - Occupational cohort
- Selection mechanism (Recruitment procedure)
  - Registry-based
  - Community-based
  - Hospital-based
  - Occupation or workplace-based
  - "Volunteer" e.g. internet advertisement

## Sample, Source, and Target

(Rothman & Greenland, Modern Epidemiology, 1998, p361)



## Population-Based vs. "Volunteer" Cohort (1)

- All cohort studies should only include volunteers
- Probability sample from the source population is an ideal
- Next best if the sampling frame (source population) is well-defined and the cohort sample is "representative":  
*population-based cohort*
  - Study participants are often not representative of the source population: e.g. a subset with higher motivation
  - Low participation does not necessarily mean a departure from representativeness: e.g. a small probability sample
- If either the source population is not clearly defined or the participants are not representative of the source population, we can call it a "*volunteer*" cohort

## Population-Based vs. "Volunteer" Cohort (2)

- The distinction can be vague, if "representativeness" is not clear-cut
- The degree of "representativeness" may range from null to perfect
- Aspects of (statistical) representativeness
  - Risk of the health outcomes
  - Distribution of exposure
  - Distribution of "important" covariates
- Examples of "good" volunteer cohort
  - Framingham Heart Study (60% of community population)
  - Nurses' Health Study
- Selection bias
  - Bias in the distribution or causal effect
  - Arises from discrepancy between study participants and the source population
- Generalizability
  - Validity of scientific generalization from source to target population
  - Generalizability  $\neq$  representativeness: E.g. experimental studies do not recruit representative rats from the households or sewage.

## Advantages of Population-Based Cohort Studies (Szklo, 1998)

- Direct application of the results to the source population
- Estimation of risk factor prevalence in the source population
- Comparability for time-trend assessment of risk factors
- Less bias for testing new hypotheses

## A Mixed Typology

- **Occupational cohort**
  - Usually representative of occupational source population
- **Population-based cohort**
  - Community cohort
  - Registry cohort (routinely maintained DB)
- **"Volunteer" cohort**
  - Source population is either undefined, or not represented by the participants
  - Recruitment can be based on registry, residence, occupation, hospital visit, or advertisement, etc.
  - Any group of individuals can be a volunteer cohort, as long as they are followed-up.
- **Other cohort**
  - Hospital-based cohorts, etc

## Occupational Cohort Studies

### 1. Based on Exposure Registry

- "National Dose Registry of Canada" Cohort Study [Sont et al., *AJE* 2001;115(4):309]
  - National Dose Registry: radiation-exposed workers with dose records between 1969- 1988 (N=191,333)
  - Linkage with Canadian Cancer Database (3,737 incident cancers)
  - Covariates: age, sex, year, exposure year, occupational group (No biomarkers)
  - Excess RR per 1 dose(Sv)\*year = 2.5 (all cancer), 64.8 (Hodgkin's disease), 5.4 (leukemia)

## Occupational Cohort Studies

### 2. Based on Workplace

- Incidence of lymphohaematopoietic malignancies in a petrochemical industry cohort: 1983-94 follow up (Huebner et al., *OEM* 2000;57(9):605-614)
  - 8942 workers employed since 1970 at refinery & chemical plants in Louisiana, USA
  - 672 incident cancer were identified by record linkage with Louisiana Tumor Registry
  - No increase in LH malignancies

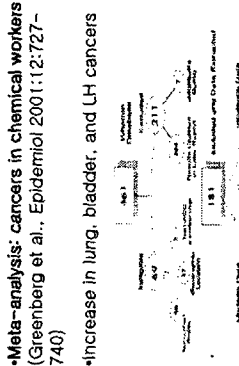


TABLE 1. Population Summary

	Number of Workers	Person-Years
All cases of death	1,573,000	15,945,027
Total cohort	117,251	1,585,138
All-cause cancer	42,285	3,335,173

## Occupational Cohort Studies

### 3. Studying Non-occupational Exposures

- Psychological distress as a risk factor for coronary heart disease in the Whitehall II Study (Stansfeld et al., *IJE* 2002;31(1):248-255)
  - 10,308 civil servants from 20 government department in London
  - BP, ECG, blood samples
  - Psychological distress at baseline was predictive of self-reported incident CHD and possible and probable ECG abnormalities.
- Incidence and Risk Factors for Stroke in an Occupational Cohort: The PROCAM Study (Berger et al., *Stroke* 1996;29(8):1562-1566)
  - 12,886 subjects from 52 workplaces in Germany
  - BP, ECG, blood samples
  - High SBP, smoking, DM history were independent risk factors for stroke

## Occupational Cohort Studies

### 4. Analysis Methods

- Survey of methods and statistical models used in the analysis of occupational cohort studies. (Callas et al., *OEM* 1994;51:649-655)
- 200 occupational cohort studies published in 1990-91,
- 154 (75%) reported SMRs
- 96 (45%) used internal comparison (RR)
- 54 (27%) used multivariate modelling
  - Proportional hazards, poisson regression, logistic regression
  - Logistic regression has no advantage, and valid only when F/U is short, outcome is rare, RRs are not large, and the underlying hazard is constant

## Framingham Heart Study

- Begun in 1948
- Boston Univ. + National Heart Institute (currently NHLBI)
- 5,209 participants
- Sampling plan
  - Population aged 30-59 10,000
  - Sampling ratio 2:3 6,600
  - 10% refusal 6,000
  - CHD free 5,000
- Measurements
  - Extensive medical exam for cardiovascular diagnosis
  - Biennial F/U exam
  - Information from death records, hospital, local physicians
- Participants characteristics
  - Considerably healthier than general population
  - Observed CHD prevalence < 10% of the expected
- Community study less than population-based, but better than just "volunteers"

## Why Framingham?

1. Large enough for needed sample size
2. Compact enough for convenient F/U
3. Sufficient variation in SEP, ethnicity
4. Relatively stable for long-term F/U
5. Near a medical center
6. Local physicians were highly cooperative
7. Only one hospital in the town
8. Vital statistics (residents, death) available
9. Previous TB study for 30 years

## Framingham: Next Generations

- Framingham Offspring Cohort
  - 5,124 of the original participants' adult children and their spouses
- Generation III Cohort
  - A Third Generation (the children of the Offspring Cohort) is currently being recruited and examined
  - New study purpose: genetic factors of CVD
  - Goal: 3,500 grandchildren of the original cohort

## Nurses' Health Study

- Harvard University + NIH
- NHS 1: since 1976
  - The primary study aim: long term consequences of OC
  - Married registered nurses (aged 30 to 55) in the 11 states
  - 122,000 (72%) nurses out of the 170,000 mailed responded
  - Every two years cohort members receive a follow-up questionnaire
  - 1980, 1984, 1986 and every four years since.
  - 68,000 sets of toenail samples between the 1982 and 1984 questionnaires.
  - 33,000 blood samples were collected in 1989.
  - Second blood collection: 2000/2001.

## NHS 2

- Since 1989
- Primary study aims: oral contraceptives, diet and lifestyle risk factors
- Younger population than the original NHS 1.
- More detailed information on type of oral contraceptive used, not obtained in NHS 1.
- Mailing to RNs in 14 states
  - The overall response rate to the baseline mailing was approximately 24% (123,000 of 517,000.)
  - Eligible participants: 116,686
- Optically scannable questionnaires
- F/U questionnaire every two years
- FFQ: every 4 years since 1991
- Blood and urine samples from approximately 30,000 nurses were collected in the late 1990's.

## Health Professionals Follow-Up Study (HPFS)

- Male counterpart for NHS, started in 1986 by Harvard University
- Study aims: Nutrition and men's health
- 51,529 men in health professions
  - 29,683 dentists
  - 4,185 pharmacists
  - 3,745 optometrists
  - 2,220 osteopath physicians
  - 1,600 podiatrists
  - 10,098 veterinarians.
- Questionnaire every two years
- FFQ every 4 years
- Blood samples
- Cheek cell collection (N=15,000) during 1998 - 2000

## European Prospective Investigation of Cancer (EPIC)

- Large multi-centre study looking at the connection between diet and cancer.
- 9 European countries

Subject recruitment in the European Prospective Investigation into Cancer and Nutrition (EPIC) (September 1999)

	Subjects included in the study with:		Completion of subject recruitment
	Questionnaire	Blood collection	
Spain	41,448	40,040	1998
Ibiza	63,087	63,077	1998
UK	98,171	43,490	1998
Netherlands	40,110	36,367	1997
France	86,321	24,371	1993
Germany	53,130	50,719	1998
Greece	27,863	26,632	1999
Sweden	53,830	53,830	1996
Denmark	57,054	56,800	1987
<b>Total</b>	<b>484,042</b>	<b>397,266</b>	

## EPIC-Norfolk (UK)

- Considered as population-based
- 30,000 people living in rural Norfolk
- Recruitment 1993 - 1997.
- Invitations were sent to all 45-74 year olds on the list of collaborating GPs.
- Baseline questionnaire, 24 dietary recall
- Appointment for a health check and FFQ.
- Health check: examined by a nurse
- Anthropometry, BP, PFT, blood and urine samples.
- Over 25,000 people attended for health checks.

## EPIC-Oxford (UK)

- Approximately 65,000 participants during 1993-1999.
- Diet and lifestyle questionnaire which consists of a food-frequency questionnaire (FFQ)
- Two methods of recruitment were used: General Practice (GP or nurse) recruitment and postal recruitment.
  - General Practice recruitment
  - Postal recruitment (to add more vegetarians)
- all members of The Vegetarian Society of the UK
- names and addresses of relatives and friends interested in receiving a questionnaire ("snowballing").
- a short questionnaire (or insert) was distributed to all members of The Vegan Society, enclosed in health/diet-interest magazines, and displayed on health food shop counters.
- Main questionnaire: 57,496 participants
- Blood samples: 19,500 participants. Blood samples were taken using a specially designed blood kit, either by Research Nurses working in General Practice surgeries, or by General Practitioners or their Practice Nurses.
- 7-day food diary: 32,000
- F/U after 5 years: 70% completed
- Second blood samples: 1,000

## Japan Public Health Center (JPHC) Study

- Population-based prospective cohort study on cancer and CVD
- Cohort 1: since 1990 (N=62,000)
- Cohort 2: since 1993 (N=79,000)
- F/U every 5 years
- Data collection
  - Questionnaire
  - Blood sample
  - Health check-ups

## Alcohol and Mortality: JPHC Study

TABLE 5. Relative risks (RR) (95% confidence interval (CI)) for all-cause and cancer mortality by smoking status and alcohol intake categories in 18,908 men, Japan, 1990-1996  
 From: Tsugane. Am J Epidemiol. Volume 150(11). December 1, 1999. 1201-1207

Type of mortality	Weekly alcohol intake (grams)			
	Non-drinkers	1-45	150-295	300-445
<b>All-cause mortality (n = 533)</b>				
Non-drinkers*	63	33	25	23
No. of deaths	1.00	1.03	0.47	0.51
Multivariate RR†		(0.68, 1.57)	(0.26, 0.74)	(0.32, 0.61)
95% CI			(0.44, 1.17)	(0.69, 1.67)
<b>Current smokers</b>				
No. of deaths	59	23	40	83
Multivariate RR†	1.20	0.99	1.06	1.25
95% CI	(0.83, 1.73)	(0.61, 1.60)	(0.71, 1.59)	(0.87, 1.79)
<b>Total cancer mortality (n = 210)</b>				
Non-drinkers*	23	13	8	10
No. of deaths	1.00	1.12	0.41	0.54
Multivariate RR†		(0.57, 2.23)	(0.18, 0.81)	(0.28, 1.14)
95% CI			(0.38, 1.74)	(0.58, 2.38)
<b>Current smokers</b>				
No. of deaths	20	6	14	24
Multivariate RR†	1.10	0.70	1.02	1.30
95% CI	(0.59, 2.04)	(0.28, 1.73)	(0.52, 2.00)	(1.18, 3.60)

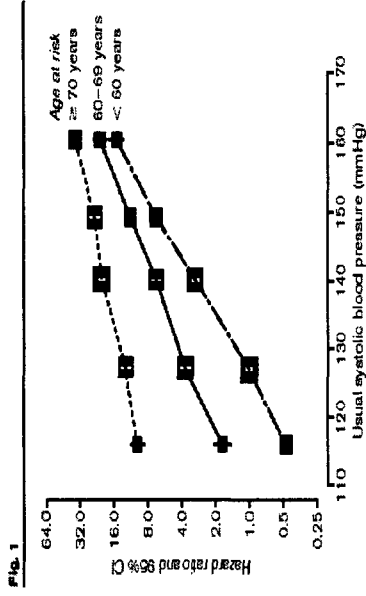
\* Includes ex- and never smokers  
 † Relative risk of study area (Hirota, Akita, Nagano, Ohtsawa), age in 1990 (40-44, 45-49, 50-54, 55-59 years), and number of cigarettes smoked per day.

## Asia Pacific Cohort Studies Collaboration (APCSC)

**Blood pressure and cardiovascular disease in the Asia Pacific region**  
 Asia Pacific Cohort Studies Collaboration  
 Journal of Hypertension 2003, 21:707-716

- Data from 37 cohort studies in Australia, China, Hong Kong, Japan, New Zealand, Singapore, South Korea, and Taiwan
- 425,325 participants with 3 million yrs
- 5178 strokes, 3047 IHD events, 6899 CVD deaths

## Asia Pacific Cohort Studies Collaboration (APCSC)



## Summary

- Cohort study became the major approach to study of chronic diseases such as CVD and cancer
- Cohort can be population-based or volunteer-based
- Types of cohort study may be categorized by source population and selection mechanism
- More and more cohort studies involve biological specimens, such as blood, urine, toenails, cheek cells, etc.
- Multi-center and multi-national collaboration is an effective way to increase sample size.
- Current statistical method typically use time-to-event analysis by Cox proportional hazard model.