Hybrid Study Designs

- Nested Case-Control Study
- Case-Cohort Study

Outline

- Definition of Cohort
- Cohort Study VS. Case-Control Study
- Hybrid Studies (Nested Case-Control & Case-Cohort)
  - Selection of Case and Controls
  - Advantages and Disadvantages
  - Situations to use

Cohort

LATIN: Roman warriors tenth part of a legion containing 300-600 foot soldiers

A group of persons who share a common experience within a defined time period
- Student cohort
- Worker cohort
- Population cohort
- Patient cohort
- Military cohort
### Design of Cohort Study

**Cohort**
- Population at Risk
- Expose/Un-Expose

---

**Outcome**

---

**Time**

---

### Advantages of Cohort Studies

- Can get best assessment of exposure
  - Measurement of exposure is not biased by knowledge of disease status (Recall bias)
- Can estimate the incidence, prevalence
- Can examine multiple outcomes per exposure
- Can establish temporal relationship

### Major Issues in Cohort Study

- Expensive
- Long time commitment for follow-up
- Impractical for rare diseases
  - How many people is needed to study a disease that has incidence 1/10,000 population?
Case-Control Study

Design of Case-Control Study

Advantages of Case-Control Study

- Quick and inexpensive
- Well-suited to the evaluation of outcomes with long latent periods
- Optimal for the evaluation of rare diseases

Major Issues in Case-Control Study

- Selection of Cases
- Selection of Controls
Selection of Cases

Cases should be representative of all of disease people in the community (target population) in terms of risk factors or other characteristics.

Target Population

Selection of Cases

- Should be similar to the cases in all respects other than the disease in question
- Should be representative of all persons without the disease in the population from which the cases are selected
- Should have the potential to become cases

Selection of Controls

Common Sources of Cases and Controls

<table>
<thead>
<tr>
<th>Sources</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital</td>
<td>Patients with disease of interest</td>
<td>Patients with other diseases</td>
</tr>
<tr>
<td>Community</td>
<td>Registration of diseases</td>
<td>Random sampling</td>
</tr>
<tr>
<td>Community</td>
<td>Registration of diseases</td>
<td>Random sampling</td>
</tr>
</tbody>
</table>

New cases in the population

- No medical attention
- Seen elsewhere
- Undiagnosed or misdiagnosed
- Death or remission before diagnosis

Cases available for case-control study
Selection of Cases and Controls

• More severe
• Unique factors related to severe form of disease

Controls?

• High socio-economic status
• More educated

Selection of Cases and Controls

You would like to study the association between alcohol drinking and pancreatic cancer

Who should be your cases and controls?

Selection of Cases and Controls

Cases

- Lung cancer patients admitted in hospitals
- Lung cancer patients from cancer registration

Controls

- Patients admitted in the same hospital with diseases other than lung cancer
- Random sampling of people in a certain area
- Friends of lung cancer patients who do not have lung cancer

Study of Genetic Susceptibility to Colon Cancer

Cohort Study VS. Case-Control

- Study population, N
- Time conducted
- Biases
- Feasibility
- Advantages
Study of Vitamin Insufficiency and Colon Cancer

- Which study design would you choose?, Why?
  - Case-Control Study
  - Cohort Study

Cohort VS. Case-Control

<table>
<thead>
<tr>
<th>Issues</th>
<th>Cohort</th>
<th>Case-Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal relationship</td>
<td>Exposure measured at the beginning</td>
<td>Uncertain</td>
</tr>
<tr>
<td>Rare disease</td>
<td>Need large number of sample size</td>
<td>Compromised</td>
</tr>
<tr>
<td>Cost</td>
<td>Expensive</td>
<td>Cheaper</td>
</tr>
<tr>
<td></td>
<td>Need to analyze large number of sample</td>
<td>Analyze only selected number of sample</td>
</tr>
<tr>
<td>Representativeness</td>
<td>Make sure that case and non-case came from the</td>
<td>Cases and Controls likely drawn from different</td>
</tr>
<tr>
<td></td>
<td>same source</td>
<td>population</td>
</tr>
<tr>
<td>Duration of study</td>
<td>Long time follow-up</td>
<td>Usually Short duration</td>
</tr>
</tbody>
</table>

Ideal Observational Study

- Well-defined population – cohort
- Cases and Non-cases drawn from the same population - cohort
- Exposure measured at the beginning - cohort
- Collect and analyze data for only selected and appropriate number of samples – case-control

Establish study population follow cohort design
Selection of case and control follow case-control design

Hybrid Studies

A case-control study within a defined cohort
Hybrid Study Designs

- Selection of the participants carried out using a case-control approach
- Study takes place within a well-defined cohort

**Issues** | **Hybrid**
--- | ---
Temporal relationship | Exposure could be measured at the beginning
Sample size for rare disease | Could select certain number of cases and controls
Cost | Analyze only selected number of samples
Representativeness | Make sure that case and non-case came from the same source (same cohort)
Duration of study | Attached with the existing cohort

Example: the MORGAM Cohort

- MORGAM (MONICA, Risk, Genetics, Archiving, and Monograph)
- Aim to explore the relationship between development of cardiovascular diseases and their genetic risk factors


Example: the MORGAM Cohort

130,000 participants from various geographically defined areas

Should we analyze all these samples

52,446 available DNA

49,815 without CHD, Stroke
2,631 developed CHD, Stroke
Variation of Hybrid Studies

- Different in method of control selection
- Case-Cohort
  - Controls selected from the baseline cohort
- Nested Case-Control
  - Controls selected at each time when a case occurs

Case-Cohort Study

- Subcohort: Randomly selected from the baseline cohort, regardless their disease status is selected as "control"
- All cases observed in the cohort are selected as "case"
- Some cases may be a part of both case and control groups

Design of Case-Cohort Study

Example of Case-Cohort Design

- Chlamydia pneumoniae IgG and CHD
- The Atherosclerosis Risk in Communities (ARIC) Study

Cohort of 15,800

- Men and women
- Aged 45-67 years
- Recruited in 1986-1989
Cohort N=120,852 From 204 Dutch municipalities

Self-administered questionnaire (food & life-style)

Lost follow-up Unconfirmed Dx

Cohort N=120,852 From 204 Dutch municipalities

Subcohort N=5,000

Pancreatic Cancer N= 520 Controls N=4,438

16 years follow-up

Pancreatic Cancer N= 520 Controls N=4,438

Cases

None Smoker N= 1,734

Former Smoker N=1,771

Current Smoker N=1,435

Advantages of Case-Cohort Study

• Efficient
  - Analysis or data collection not needed for the whole cohort

• Reduce selection bias
  - Case and controls selected from the same population

• Reduce recall bias
  - Exposures collected at baseline

• Subcohort ~ baseline cohort
  - The same subcohort can be used to analyze several endpoints
    - Smoking and pancreatic cancer
    - Smoking and CHD
  - Use to estimate population parameters
  - Odds Ratio = Rate Ratio

What would happen if this study was case-control or cohort study?

Cancer Epidemiol Biomarkers Prev; 19(6) June 2010
Disadvantages of Case-Cohort Study

- Require a well-defined cohort
- Still affected by lost to follow-up
- Subcohort members cannot be matched to cases
- Require more complicated analysis
  - Weighting
  - Variance estimation
  - Less precision than cohort study
  - Loss to follow-up: corrected by using survival analysis

Design of Nested Case-Control

Cases

Defined Cohort

Controls

Example of Nested Case-Control Study

- Study of relationship of military rank and radiation exposure to brain tumor
- Military rank and radiation exposure: abstracted from occupational records

Cohort of US Air force members

N=880,000

230 brain tumor cases

920 controls still employed at the time case diagnosed

Controls are selected concurrently as case developed
- Control matched on time at risk to the case
- A control may be selected for more than one case
- A control may later become a case

Match on
- Year of birth
- Race

Szklo and Nieto. Epidemiology Beyond the Basic
Associations between carcinogen–DNA damage, glutathione S-transferase genotypes, and risk of lung cancer in the prospective Physicians’ Health Cohort Study

Frederica P.Perera, LaVerne A.Mooney, Moir Stumpe, David H.Phillips, Douglas A.Bell, Andrew Rundle, Stan Cho, Wei-Yuan Tsai, Jing Ma, Anne Blackwood and Deliang Tung

PHS cohort
Male healthy physician in USA
N= 22,071

Stored Blood Sample at baseline
N=14,916

Questionnaire
Smoking history

89 CA Lung cases
178 matched controls

Genotypes Analysis only
89+178 blood samples

13 years F/U

Advantages of Nested Case-Control Study

• Efficient
  • Analysis or data collection not needed for the whole cohort

• Reduce selection bias
  • Case and controls selected from the same population

• Matching case and controls on time at risk
  • Unit of person-time at risk
  • Can match on other potential confounders
  • Odds ratio = Rate ratio

• Reduce recall bias
  • Exposures collected at baseline

Disadvantages of Nested Case-Control Study

• Control series is highly selected (matching)
  • Less representative of the whole cohort
  • Same set of control cannot be used for different endpoints

Means and Standard deviations of aromatic-DNA adducts, by level of GSTM1 genotypes

<table>
<thead>
<tr>
<th></th>
<th>Cases Mean ± SD (n)</th>
<th>Controls Mean ± SD (n)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>8.32±16.57 (87)</td>
<td>7.39±12.92 (160)</td>
<td>0.42</td>
</tr>
<tr>
<td>GSTM1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild-type</td>
<td>12.02±24.03 (39)</td>
<td>8.20±15.33 (91)</td>
<td>0.08</td>
</tr>
<tr>
<td>Null mutation</td>
<td>5.36±3.93 (47)</td>
<td>6.31±8.78 (69)</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Carcinogenesis vol.23 no.10 pp.1641-1646, 2002

Carcinogenesis vol.23 no.10 pp.1641-1646, 2002
When the hybrid studies should be conducted?

- Assessment of exposures of the whole cohort is not efficient
  - Labor-intensive
  - Cost
  - Left-over samples can be used for future analysis

<table>
<thead>
<tr>
<th>Case-Cohort VS. Nested Case-Control</th>
<th>Case-Cohort</th>
<th>Nested Case-Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allow to analyze several endpoint for the same control (subcohort)</td>
<td>One subcohort can be used to analyze several diseases</td>
<td>Controls have to be re-selected for different endpoints</td>
</tr>
<tr>
<td>Statistical Analysis</td>
<td>More Complicated</td>
<td>Less Complicated</td>
</tr>
<tr>
<td>- Survival Analysis</td>
<td>- Variance estimate</td>
<td>- Logistic regression</td>
</tr>
<tr>
<td>Match case and control</td>
<td>Less likely</td>
<td>Possible</td>
</tr>
</tbody>
</table>

**Case-Cohort**

- All cases arise from baseline cohort
- Random sampling from the baseline cohort - cases in subcohort
- Subcohort: good representative of total cohort
- Yes
- Yes
- Controls: Less representative of total cohort

**Nested Case-Control**

- All cases arise from baseline cohort
- Population at approximate times when cases occur during follow-up
- Risk Ratio
- Yes
- Populations are less likely to be representative of the total population

Hybrid Studies

B Vitamin Plasma Levels and the Risk of Ischemic Stroke and Transient Ischemic Attack in a German Cohort
Cornelia Weikert, MD, MPH; Juana Dierkes, PhD; Kurt Hoffmann, PhD; Klaus Berger, MD, MPH; Dagmar Droge, MPH; Kerstin Klipstein-Grobusch, MSc, PhD; Joachim Strenger, MD; Matthias Möhlig, MD; Claus Loley, MD; Heiner Boening, MSPH, PhD

- A large-scale Europe-wide prospective cohort
- N~27,000
- Discuss study designs
  - Cohort
  - Case-Control
  - Case-Cohort
  - Nested Case-Control
Assignment

1. Search for a publication that used hybrid study design (either case-cohort or nested case-control)

2. Provide a brief introduction of the study, including rationale, methodology, important results

3. Discuss advantages and disadvantages of the study design, comparing with case-control study and cohort study

• Submit your work by October 4, 2017 (2 weeks)