Cohort Studies

Jerrilyn Cambron, DC, PhD
Department of Research
Quiz: Clinical trials

1. What is the difference between random selection and random allocation?
2. Which one of the two above is always involved in a randomized clinical trial?
3. What is the difference between internal validity and external validity (generalizability)?
Hierarchy of scientific evidence: observational

- Meta-Analysis, Systematic Reviews
- Best Evidence / Evidence Guidelines
- Randomized, controlled trials (RCT)
- Clinical trials, Cohort Studies, Case Control
- Case series
- Case study / case report
- Animal studies, in vitro studies
- Expert opinions, editorials
Other names for cohort study

- Incidence study
- Longitudinal study
- Forward-looking study
- Follow-up study
- Concurrent study
- Prospective study
Cohort studies

- Observational study
  - No randomization
  - No control over intervention or risk factor exposure
- Follow groups prospectively from exposure to outcome of interest
- All participants in the cohort of subjects must be at risk of developing the outcome of interest
Cohort study: Prospective nature of design

Study investigators begin with people who were exposed or not exposed, and follow them forward to see if they get the outcome of interest.
Subjects must *not* start with outcome

- If subjects join a cohort study with the outcome of interest, the investigators cannot measure the point at which it develops because it is already present.
Historical cohort

- Still prospective in nature in that the subjects are followed from exposure to outcome
- Start with a study sample. *Sample is from the past.*
- Exposure or non-exposure to risk factors for each subject is determined based on previous records.
- Subjects are followed forward over time to determine outcome development in present time or in future.
1. ‘Prospective’ or ‘concurrent’ cohort

2. ‘Historical’ or ‘retrospective’ cohort

3. ‘Ambidirectional’ cohort

Figure 2: Schematic diagram of concurrent, retrospective, and ambidirectional cohort studies
Historical cohort

- Most historical cohort studies are occupational studies in which workers complete the same occupational tasks over a period of years.
- Must have specific definition of past exposure, therefore only possible if have way to measure exposure status.
- Exposed and unexposed groups may be from the same company but in different positions.

Ex: Plant workers for whom you have previous personnel records regarding exposure (ie: department or job title) and you contact them to find out about current disease status.
Selection of subjects: 2 methods

1. Start with group and then measure exposure

2. Start with exposed and unexposed groups
   - Who do you choose as the unexposed group?
FIGURE 8–5. Design of a cohort study beginning with a defined population.
FIGURE 8–4. Design of a cohort study beginning with exposed and nonexposed groups.
Start of study: Who is exposed or NOT exposed?

- Cohort studies need a clear and unambiguous definition of the exposure.
- All participants need to be assessed for the exposure based on the definition.
- Ex: Definition of smoking may include ‘bar smokers’ or ‘second hand smoke’
** It is important to recognize that the exposed and unexposed groups may have other baseline differences that affect the results!
End of study:
Who does/not have the outcome of interest?

- Cohort studies need a clear and unambiguous definition of the outcome
- All outcomes should be assessed equally between the exposed and non-exposed groups
- Duration of follow-up is based upon the length of disease latency so that outcome can truly be assessed

- Ex: Cancer defined through colonoscopy (all subjects get colonoscopy regardless of symptoms), then biopsy of suspicious lesions
Blinding issues

- If investigators know who was exposed and who was not, there may be bias in determination of the outcome. Therefore, the clinician determining the outcome, should be blind to the exposure status.

- However, because subjects know if they were exposed or not exposed (and are not blind), they may describe their outcomes differently.
Losses to follow-up

- Drop out rate may be differential, meaning related to exposure, outcome, or both.
- People who have disclosed that they were exposed to something embarrassing or illegal may be less likely to complete the study than those who were not exposed to it.

Drug users

Do they acquire HIV over time?

Drug users

Did the lost drug users die of HIV?

Die of other causes?

Stop using drugs?

Not drug users

Not drug users

This might be an unfair comparison.
Incidence

- The proportion of a new disease (outcome) in a population
- Indicates the magnitude of risk in a group of people with a certain exposure

\[
\text{Incidence} = \frac{\text{# people with new disease}}{\text{# people who were at risk}}
\]

- Incidence can be measured in both clinical trials and cohort studies because they are both prospective!
- aka: Absolute risk
Incidence vs. Prevalence

Incidence = Number of NEW cases  
(cohort) Total population

Prevalence = Number of ALL cases  
(survey) Total population

Cohort studies look at NEW disease, so they utilize incidence.
Relative Risk (RR)

- A measure of strength of the exposure-disease association in COHORT STUDIES. Equals the ratio of the risk in the exposed group to the risk in the unexposed group

- RR=(incidence in exp)/(incidence in unexp)

- Interpretation
  - RR=1 means no association
  - RR>1 means positive association (causal)
  - RR<1 means negative association (protective)

- AKA: Rate ratio
Basic analytical approach in a cohort study.
Example

- Incidence (risk)
  - Car accidents when on phone: 25/100 drivers
  - Car accidents when not on phone: 5/100 drivers

- Relative risk
  \[
  \text{Relative risk} = \frac{\text{Car accidents on phone}}{\text{Car accidents not on phone}} = \frac{25/100}{5/100} = 0.25/0.05 = 5
  \]

- There is a 5 times greater risk of a car accident when on the phone versus when not on the phone.
Confidence Intervals

- A range of values (interval) in which the ‘true value’ falls with a certain percent of confidence

- Ex: RR = 5.0 (95% CI: 1.2 - 9.5) means that the relative risk is 5.0, and that we are 95% confident that the true value falls between 1.2 and 9.5
Statistical significance of relative risk measure based on CIs

- Confidence intervals that include 1.0 equate with no statistical significance (ex: 0.5-3.2)
- Confidence intervals that do NOT include 1.0 equate with statistical significance (ex: 1.4-4.3)

Ex: RR = 3.4 (95% CI: 1.2 - 9.5) means that the relative risk is 3.4, is statistically significant, and that we are 95% confident that the true value falls between 1.2 and 9.5.
Calcium supplementation in pregnancy for preventing pre-eclampsia (Atallah et al. The Cochrane Library. 2000)

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Calcium n/N</th>
<th>Placebo n/N</th>
<th>RR (95%CI Fixed)</th>
<th>Weight %</th>
<th>RR (95%CI Fixed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Villar1</td>
<td>1987</td>
<td>1 / 25</td>
<td>3 / 27</td>
<td></td>
<td>1.0</td>
<td>0.36[0.04,3.24]</td>
</tr>
<tr>
<td>L-Jaramillo1</td>
<td>1989</td>
<td>2 / 55</td>
<td>12 / 51</td>
<td></td>
<td>4.4</td>
<td>0.15[0.04,0.66]</td>
</tr>
<tr>
<td>Villar2</td>
<td>1990</td>
<td>0 / 90</td>
<td>3 / 88</td>
<td></td>
<td>1.2</td>
<td>0.14[0.01,2.67]</td>
</tr>
<tr>
<td>L-Jaramillo2</td>
<td>1990</td>
<td>0 / 22</td>
<td>8 / 34</td>
<td></td>
<td>2.4</td>
<td>0.09[0.01,1.48]</td>
</tr>
<tr>
<td>Belizan</td>
<td>1991</td>
<td>15 / 579</td>
<td>23 / 588</td>
<td></td>
<td>8.0</td>
<td>0.66[0.35,1.26]</td>
</tr>
<tr>
<td>S-Ramos</td>
<td>1994</td>
<td>4 / 29</td>
<td>15 / 34</td>
<td></td>
<td>4.9</td>
<td>0.31[0.12,0.84]</td>
</tr>
<tr>
<td>Purvar</td>
<td>1996</td>
<td>2 / 97</td>
<td>11 / 93</td>
<td></td>
<td>4.0</td>
<td>0.17[0.04,0.77]</td>
</tr>
<tr>
<td>CPEP</td>
<td>1997</td>
<td>158 / 2163</td>
<td>168 / 2173</td>
<td></td>
<td>59.0</td>
<td>0.94[0.77,1.16]</td>
</tr>
<tr>
<td>L-Jaramillo3</td>
<td>1997</td>
<td>4 / 125</td>
<td>21 / 135</td>
<td></td>
<td>7.1</td>
<td>0.21[0.07,0.58]</td>
</tr>
<tr>
<td>Crowther</td>
<td>1999</td>
<td>10 / 227</td>
<td>23 / 229</td>
<td></td>
<td>8.1</td>
<td>0.44[0.21,0.90]</td>
</tr>
</tbody>
</table>

Total (95%CI) 196 / 3412 287 / 3452

Test for heterogeneity chi-square=28.67 df=9 p=0.0007
Attributable risk (AR)

- The extent (proportion) to which the incidence of a disease can be attributed to a certain exposure
- This demonstrates public health significance: “If we eliminate this risk factor, we can then prevent this much disease.”

AR = (incidence in exp) - (incidence in unexp)
FIGURE 11-1.  A, Total risks in exposed and nonexposed groups. B, Background risk. C, Incidence attributable to exposure and incidence not attributable to exposure.
Example

- **Incidence (risk)**
  - Car accidents when on phone: 25/100 drivers
  - Car accidents when not on phone: 5/100 drivers

- **Attributable risk**
  \[
  \text{Attributable risk} = \frac{25}{100} - \frac{5}{100} = \frac{20}{100}
  \]

- If we eliminate talking on the phone while driving, we would prevent 20 car accidents per 100 people.
Advantages of cohort studies

1. Can calculate an incidence rate and relative risk
2. Can examine many exposures
3. Efficient for the study of rare exposures
4. Establishes correct exposure-disease association
5. Treatment is not withheld from subjects, and subjects are not artificially exposed to hazards or treated with new therapies
Disadvantages of cohort studies

1. A large number of people is needed
2. May be hard to find control (non-exposed) subjects if most people have been exposed to risk factor
3. Drop out due to long follow-up is a threat to validity
4. Changes over time in exposure status may confound exposure data
5. Changes in diagnostic techniques may lead to inaccurate previous outcomes
6. Expensive and time consuming, if prospective
7. Requires availability of adequate records, if retrospective
8. Inefficient for evaluating rare diseases
Well-known cohort studies

- Framingham Study
- Nurses’ Health Study
- Women’s Health Initiative
- Physicians’ Health Study
Example: Framingham study

- In the 1930s and 1940s, coronary heart disease was considered a natural consequence of aging.
- By the late 1940’s, the US Public Health Service questioned the possible risk factors.
- The objective of the Framingham Heart Study was to identify the common factors or characteristics that contribute to CVD by following its development over a long period of time in a large group of participants who had not yet developed overt symptoms of CVD or suffered a heart attack or stroke.
Defining the cohort (1948)

- Framingham, MA was a town of 28,000 people and was located 20 miles from Boston.
- It was a relatively self-contained community with sufficient numbers in the proper age range.
- From a previous study, there was evidence that the town folks were cooperative.
- The town was small enough that everyone could go to one examination center, and follow up would be easy because there was only one central hospital.
Study duration to be 20 years

- Long term follow-up was necessary in this study due to the slow development of CHD
- The study would include ages 30 to 62 years old
  - <30 years, the incidence of CHD was too low
  - >62 years, CHD would be too advanced
Target sample size

- The target sample size was 5,000 but investigators chose 6,600 to compensate for refusals.
- A list of the town members was developed according to the number of eligible people in the family.
- 2 out of every 3 families were selected.
- 6,507 were recruited.
- Only 4,469 (69%) participated; at baseline, 1.5% had existing heart disease and were excluded.
Examination for study

- Medical history
- Physical exam
- Lab tests
  - EKG
  - Chest x-ray
  - Serum lipids
Follow-up of subjects

- Subjects were followed up every 2 years after their initial examination.
- The follow up exams were primarily directed at detecting:
  - CHD
  - Stroke
  - Peripheral vascular disease
Data analysis and publication

- There have been ~1,200 articles published since 1951
- Risk factors for CHD have been identified with associated RRs
CHD risk and related factors

- Risk factors: high blood pressure, high cholesterol, smoking, obesity, diabetes, and physical inactivity
- Related factors: blood triglyceride and HDL cholesterol levels, age, gender, and psychosocial issues.
<table>
<thead>
<tr>
<th>Risk Group</th>
<th>Characteristics of Risk Group</th>
<th>Risk</th>
<th>Risk Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowest</td>
<td>All of the following factors:</td>
<td>2.2%</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>- Non-smoker</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- No glucose intolerance</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- No hypertrophy of left ventricle</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Low systolic blood pressure (&lt;105 mm Hg)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Low cholesterol level (&lt;185 mg/dL)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest</td>
<td>All of the factors listed below</td>
<td>77.8%</td>
<td>35.4</td>
</tr>
<tr>
<td>Intermediate</td>
<td>One of the following factors:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Smoker</td>
<td>3.8%</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>- Glucose intolerance</td>
<td>3.9%</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>- Hypertrophy of left ventricle</td>
<td>6.0%</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>- Severe hypertension (systolic blood pressure ≥195 mm Hg)</td>
<td>8.4%</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>- High cholesterol level (≥335 mg/dL)</td>
<td>8.5%</td>
<td>3.8</td>
</tr>
</tbody>
</table>

Since the original study...

- Since 1948, the subjects have continued to return to the study every two years for a detailed medical history, physical examination, and laboratory tests.
- In 1971, the study enrolled a second generation - 5,124 of the original participants' adult children and their spouses - to participate in similar examinations.
- In April 2002, the Study entered a new phase: the enrollment of a third generation of participants, the grandchildren of the original cohort. Exam 1 of the Third Generation Study was completed in July 2005 and involved 4,095 participants.
Just released on 01/25/08…

- Risk assessment tool for estimating 10-year risk of developing hard CHD (myocardial infarction and coronary death)

Dr. Claire Johnson: JMPT editor

How to Publish Your Case Report

Wednesday, June 4th 12–1 PM in SC 1

RSVP to ebpfaculty@nuhs.edu by 10:00 AM Tuesday, June 3rd to receive a voucher for lunch

If you go, you will receive 5 extra credit points on your midterm exam

Objectives that will be covered:

- Know the benefits and limitations of case reports.
- Understand how case reports are developed
- Become familiar with the essential components of a case report
- Become aware of what to include when submitting a case report for publication
- Recognize what to avoid when writing and submitting a case report for publication
- Appreciate the amount of time and effort that is needed to complete a publishable case report
- Know where to submit case reports for publication
Any questions?