Strategies for data analysis: Cohort studies

From research to practice: Postgraduate training in reproductive health/chronic disease

Dr Lale Say
March 2003
Cohort studies – Goal/utility

To measure and usually compare the incidence of disease in one or more study cohorts

To estimate average risks, rates or occurrence times

Cohort: a group of people who share a common experience or condition (eg. a cohort of smokers)
Analysis 1

Define the characteristics of the cohort
- decision to use case-non case data/person-time data
- Calculate risks/rates among groups accordingly
Risk

Proportion of people who develop the disease over a specified period of time

\[
\text{risk} = \frac{\text{N of sick people}}{\text{total population}}
\]

e.g. 1000 people observed for 5 years,
- 958 never became sick
- 42 became sick

\[
\text{risk} = \frac{42}{1000} = 0.042
\]
Risk of LBW in Denmark*

Subsequent liveborn infants of 11,069 women with previous LBW babies are evaluated in the subsequent pregnancy;

- 9,021 had normal birth weight babies
- 2,048 had LBW babies

risk = 2048 / 11069 = 18.5%

(Basso, 1997)*
Rate

Proportion of people who develop the disease during the total amount of observation time

rate = N of sick people / total amount of time people are observed (total time at risk)
Incidence of type 1 DM in Norway*

1 382 602 children were observed for 15 years
- 1 382 547 never became sick, 55 developed type 1 diabetes
- total observation period for all: 8 184 994 person-years

rate = 55 / 8 184 994 = 0.067 per 1000 person-years

(Stene, 2001)*
Disease Odds

Odds = probability of disease / probability of not disease
= probability of disease / 1 - probability of disease
= risk / 1-risk
Measures of Disease Frequency

RISK/RATE

- N of diseased
- N of total population/
  Total observation period

ODDS

- Risk
- 1-risk

Odds approximates risk when
risk is close to 1
## Risk/Rate Ratio, Odds Ratio

<table>
<thead>
<tr>
<th>Disease +</th>
<th>Disease -</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed (E+)</td>
<td>(a)</td>
</tr>
<tr>
<td>Unexposed (E-)</td>
<td>(c)</td>
</tr>
</tbody>
</table>

**Risk/rate ratio** = \(\frac{\text{risk in } E^+}{\text{risk in } E^-} = \frac{a}{a+b} / \frac{c}{c+d}\)

**Odds ratio** = \(\frac{\text{odds in } E^+}{\text{odds in } E^-} = \frac{a/b}{c/d} = \frac{ad}{bc}\)
Risk difference

Risk in exposed – Risk in unexposed
Analysis 2

- Check for sub-groups (strata)
  - low exposure, medium exposure, high exposure, etc.
  - Age, education, etc.

- Calculate risk/rate ratio in different subgroups (strata)

- Compare/adjust for other variables (confounders) between two groups
Induced abortion and low birthweight in the subsequent pregnancy*

Objective:
To examine whether induced abortion increases the risk of low birthweight in subsequent singleton livebirths

Methods

- Participants: all women who had their first pregnancies during 1980-82
- Exposed group: all primigravidae whose previous pregnancies were terminated by first-trimester induced abortion (n=11,394)
- Unexposed group: all primigravidae who had spontaneous termination of pregnancy (n=40,758)
- Follow-up: until subsequent deliveries
- Main outcome measure: Low birthweight baby in the subsequent delivery
## Results

<table>
<thead>
<tr>
<th></th>
<th>LBW +</th>
<th>LBW -</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abortion (E+)</td>
<td>570</td>
<td>10 824</td>
<td>11 394</td>
</tr>
<tr>
<td>Control (E-)</td>
<td>1427</td>
<td>39 331</td>
<td>40 758</td>
</tr>
<tr>
<td>Total</td>
<td>1997</td>
<td>50 155</td>
<td>52 152</td>
</tr>
</tbody>
</table>

Risk ratio = \( \frac{570}{11 394} / \frac{1427}{40 758} = 1.42 \)

Odds ratio = \( \frac{570 \times 39 331}{10 824 \times 1427} = 1.45 \)
Sub-groups

- Methods of abortion
- Age
- Inter-pregnancy intervals
- Gestational age of abortion
Confounders

- Previous spontaneous abortion
- Maternal age
- Residence
- Gender of newborn
Useful link

http://www.ccnmtl.columbia.edu/projects/episim/study2f.html

provides an example on the steps of analysing cohort design