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

risk = N of sick people / total population

e.g. 1000 people observed for 5 years,

- **958 never became sick**
- **42 became sick**

risk= 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;

- 9021 had normal birth weight babies
- 2048 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

$$\frac{\text{N of diseased}}{\text{N of total population/ Total observation period}}$$

ODDS

$$\frac{\text{Risk}}{\text{1-risk}}$$

Odds approximates risk when risk is close to 1

Risk/Rate Ratio, Odds Ratio

	Disease +	Disease -
Exposed (E+)	a	b
Unexposed (E-)	c	d

Risk/rate ratio=risk in E+ / risk in E- = a /a+b / c /c+d

Odds ratio=odds in E+ / odds in E- = a/b / c/d = 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

*Zhou W et al, Int J of Epidem, 2000;29:100-106

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

	LBW +	LBW -	Total
Abortion (E+)	570	10 824	11 394
Control (E-)	1427	39 331	40 758
Total	1997	50 155	52 152

Risk ratio= 570 /11 394 / 1427/40 758 = 1.42
Odds ratio= 570x39 331 / 10 824x1427= 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