Introduction to Cohort Studies: design and key points

BSc Global Health – Module 2 – November, 2011

Marc Chadeau-Hyam

Department of Epidemiology and Biostatistics

Imperial College London

Learning objectives

- Understand the design of a cohort study
- Critically appraise the use of a cohort study design to answer a specific question
- Calculate, understand and interpret crude RR

Main type of studies in Epidemiology

Objective of epidemiological studies: to assess the impact of Exposures ($e.g.$ risk factor) on the risk of a certain Outcome ($e.g.$ disease onset).

- Randomized controlled trials:
  Quantify the effect of a treatment on a certain condition.

  $\Rightarrow$ specific to one outcome, and few treatments

- Case-Control studies:
  Identify features discriminating diseased/healthy individuals.

  $\Rightarrow$ Specific to one outcome, possibly several exposures

- Cohort Studies:
  Follow-up in time the evolution of the health status of a population of interest.

  $\Rightarrow$ Possibly several outcomes, and several exposures

Cohort (longitudinal) study: Definition

Question of interest: do a certain set of exposures play a role in the development of a certain condition?

$\Rightarrow$ what is the role of nutrition in carcinogenesis (EPIC)?

- Studied population:
  Defining a group of individuals in which information about the exposure of interest will be collected

  $\Rightarrow$ collection of data on dietary habits, quantification of food/nutrient intake

- Follow-up:
  Health conditions are ascertained forward in time in the population

  $\Rightarrow$ Identification of disease(s) onset ($.e.g.$ cancers)

$\Rightarrow$ exposure prior to onset can be related to subsequent disease(s) experience

Cohort (longitudinal) study: Main steps

- Definition of a scientific question of interest

  $\Rightarrow$ identify outcome(s) and exposure(s) of interest

- Recruitment:

  $\Rightarrow$ what population would enable to answer the question of interest?

- Data Collection:

  $\Rightarrow$ measuring the exposures of interest in the enrolled population

- Follow-up:

  $\Rightarrow$ monitoring the outcome(s) of interest

- Statistical analyses: measure of association

  $\Rightarrow$ quantify the effect of exposure(s) on outcome(s)

Population recruitment

We are looking into the role of exposure A in the risk of developing a condition B.

$\Rightarrow$ What are the specifics of the population we should enroll?

Population recruitment

We are looking into the role of exposure A in the risk of developing a condition B.

- Aim of the study: disentangle if A plays a role in the subsequent occurrence of B

  $\Rightarrow$ need to consider exposed and unexposed subjects

- General characteristics:
  - All included subjects should be disease-free at enrollment
  - All included subjects MUST be at risk of developing B ($.e.g$ women with hysterectomy should be excluded from an endometrial cancer study)
  - Exposure can be quantitative (not necessarily binary)

Population recruitment

We are looking into the role of exposure A in the risk of developing a condition B.

- Aim of the study: disentangle if A plays a role in the subsequent occurrence of B

  $\Rightarrow$ need to consider exposed and unexposed subjects

- General characteristics:
  - All included subjects should be disease-free at enrollment
  - All included subjects MUST be at risk of developing B ($.e.g$ women with hysterectomy should be excluded from an endometrial cancer study)
  - Exposure can be quantitative (not necessarily binary)
- Unexposed population:
  - As similar as possible to the exposed population $w.r.t.$ covariates other than studied exposure(s) ($e.g.$ similar age structure)

Data Collection

- How to measure the exposure(s) and covariates of interest?
- Exposure is measured at recruitment (baseline):
  - Pre-existing records
  - Self reporting (questionnaire, interviews)
  - Proxy measurement (job title, biomarkers)

  $\Rightarrow$ ideally unbiased, complete set of information

- Including a temporal component
  Level of exposure may change over time

  $\Rightarrow$ possible underestimation of the true association

  $\Rightarrow$ possible reassessment along the study (reexamination, resurvey)

Follow-up

- Aim: monitor the occurrence of the disease(s) of interest, the vital status, and the cause of death in time

  $\Rightarrow$ the health status of each enrolled subject should be assessed (registries, medical records, self/family-reports)

- Misclassification of the outcome: Failure to ascertain disease incidence and/or vital status

  $\Rightarrow$ potential misleading conclusions

- Lost to follow-up: enrolled people that were lost

  $\Rightarrow$ for some subjects, outcome/vital status is censored

  1. Is there a pattern in exposure and/or outcome of the lost population?

     $\Rightarrow$ bias and results may be affected

  2. Is the follow-up process is comparable within each exposure class?

     $\Rightarrow$ comparison of the disease experience in subgroups should be unbiased

Assessing the quality of a cohort

- Completeness: does it include as many eligible subjects as possible?

  $\Rightarrow$ potentially overlooked or omitted subject may differ w.r.t exposure and/or vital status

- Healthy Worker Effect: selection bias
  - Working enrolled subjects may experience lower mortality/ disease incidence compared to the general population

  $\Rightarrow$ underestimation of the relative risk (ill people move in non-exposed)

- Generalisablility
  - If selected population is different from the eligible group, external validity of the findings may be questioned
  - However, participation is not likely to depend on exposure AND risk of the outcome: internal validity holds.

$\Rightarrow$ need to carefully check for these features while analysing the data

Measure of association

- How to summarise data collected?
  We study the role of exposure (e.g. 'smoking') in the occurrence of a certain type cancer. Available data from follow-up are:
  - Unexposed: 76 did not developed that cancer, 115 did
  - Exposed: 94 did not developed cancer, 225 did

$\Rightarrow$ how can we answer the question from these data?

Measure of association

- Defining a 2×2 table

	Non-Cancer	Cancer	Total
Non-Smokers	76	115	191
Smokers	94	225	319
Total	170	340	510

1. Describe the table
2. Can we answer the question from the table?
3. Define the risk of cancer in smokers
4. Define the risk of cancer in non-smokers

Measure of association

	Non-Cancer	Cancer	Total
Non-Smokers	76	115	191
Smokers	94	225	319
Total	170	340	510

1. Risk of cancer in non-smokers:

$$R = \frac{115}{76 + 115} = 60.2\%$$

2. Risk of cancer in smokers:

$$R = \frac{225}{94 + 225} = 70.5\%$$

3. What can you tell from these risks? What would the risk ratio mean?

Measure of association

	Non-Cancer	Cancer	Total
Non-Smokers	76	115	191
Smokers	94	225	319
Total	170	340	510

1. Define the risk ratio:

$$RR = \frac{70.5\%}{60.2\%} = 1.17$$

On average, smokers have 1.18 times more chances of getting that cancer.

WARNING: statistical significance has not been assessed!!!!!

Measure of association

	Non-Diseased	Diseased	Total
Non-Exposed	a	b	(a+b)
Exposed	c	d	(c+d)
Total	(a+c)	(b+d)	(a+b+c+d)

1. Define the formula for the risks of cancer
2. Define the formula for the risk ratio

Measure of association

	Non-Diseased	Diseasde	Total
Non-Exposed	a	b	(a+b)
Exposed	c	d	(c+d)
Total	(a+c)	(b+d)	(a+b+c+d)

1. In unexposed:

$$R_{\overline{E}} = \frac{b}{(a+b)}$$

2. In exposed:

$$R_E = \frac{d}{(c+d)}$$

3. Relative Risk:

$$RR = R_E / R_{\overline{E}} = \frac{(a+b) \times d}{(c+d) \times b}$$

Strengths of cohort studies

- Ability to look at multiple outcomes

  $\Rightarrow$ possibility to nest a case-control study within the cohort

- Ability to elucidate temporal relationship between exposure and disease outcome
- Incidence can be calculated in both exposed and unexposed
- Less prone to exposure bias (prospective design)
- Ability to handle rare exposures

Weaknesses of cohort studies

- Time consuming design (follow-up over years/decades)
- Expensive procedure (recruitment, data collection, follow-up of large numbers of participants)
- Subject to potential bias through loss to follow-up
- subject to Healthy Worker Effect in occupational epidemiology
- Rare diseases?