1 Case Control Studies

1. Case-control (retrospective) study: Select cases, Select controls, Compare cases and controls

2. Cutaneous melanoma (Autier, 1996). 420 cases from 5 hospitals in Belgium, France and Germany. 447 controls local community served by the hospitals. 75% cases reported that they had not been protected against sunlight during their childhood. The percentage for controls is 69%. A case is more closely associated with lack of childhood protection of sunlight than a control. Exposure to sunlight in childhood may be a risk factor for Cutaneous melanoma.

3. Difference between Case-control studies and Cohort Studies. In Case Control Study, subjects are selected by disease status and we look back to see what, in the past, might have caused the disease. In Cohort study, A random population is selected in most cases and subjects are followed-up to see whether disease develops.

4. Advantages of Case-control studies:

   (a) Case Control studies are quicker and cheaper than follow-up studies - Good for disease with a long latency

   (b) Many risk factors can be studied simultaneously, many questions can be asked.

   (c) Case-control studies are particularly well suited to investigations of risk factors for rare diseases.

   (d) They require much smaller sample sizes than equivalent cohort studies.

   (e) Generally able to evaluate confounding and interaction more precisely for the same overall sample size than cohort studies.

   (f) Transient risk factors can be ideally studied e.g. contaminated food, pollution caused through industrial accidents, use of mobile telephones in cars.

5. Disadvantages of Case-control studies:

   (a) Case-control studies are not able to demonstrate causality. A case might reflect survival rather than morbidity.

   (b) Case-control studies can investigate only one disease outcome.

   (c) Case-control studies cannot provide valid estimates of risk or odds.

(d) Case-control studies are very likely to suffer from bias error.

(e) Problems may arise from the way controls are sampled. Differential quality of information: Cases may be researched more thoroughly.

(f) When asked to provide information, cases may be more likely to be accurate than the controls.

(g) Cases may also report biased information

(h) When the purpose is to investigate cause and effect, cohort studies are more reliable than case-control studies.

6. Application of case-control studies: Often used for pragmatic reasons. Case-control needs to be conducted very carefully, with full regard to possible sources of bias. They need to be reported with evidence of avoidance or minimization of bias resulting from i) Blindness of the analyst to the case or control status of individuals ii) Careful selection of cases and controls. A large odds ratio normally indicate an association between risk factor and disease.

7. Example During January 1984, six cases of Legionnaires disease were reported to the health authority in Reading, U.K, all of whom became ill between 15 and 19 December 1983. This cluster suggested a point source outbreak. Further investigation detected 7 more cases. Cases do not have obvious factor in common, but were found to have visited Reading town center just before their illness. A case-control study was conducted between the cases and 36 people without disease (controls). The result

Table 1. Number of people visiting parts of Reading town center in the 2 weeks preceding onset of Legionnaires' disease

Area of Reading	Cases	Controls
Abbey Square	9	19
Butts Centre	12	21
Forbury Gardens	3	6
Minster Street	9	21
South Street	4	9
Train station	3	9
Overall	13	36

suggests that the Butts Center might be a source of the legionella bacterium. A water sample from a cooling tower in one of the buildings in the Butts Center was found to have the legionella species, Legionella pneumophila.

8. Basic Methods of Analysis: Dichotomous Exposure: Consider the situation where the risk factor is dichotomous (yes or no). The table gives the data reported by Autier et al. (1996). We cannot estimate risk and relative risk. A sample that is stratified by case-control status. Cases generally have higher proportion in the case-control sample than in general population.

Table 2. Sun protection during childhood by case-control status for cutaneous melanoma.

Sun protection?	Cases	Controls	Total
Yes	99	132	231
No	303	290	593
Total	402	422	824

Table 3. Risk factor status by (a) disease status in the population and (b) case-control status in the sample (showing expected values).

	(a) Population values			(b) Expected values in the sample		
Risk factor status	Diseased	Not diseased	Total	Cases	Controls	Total
Exposed	A	B	$A+B$	f_1A	f_2B	f_1A+f_2B
Not exposed	C	D	$C+D$	f_1C	f_2D	f_1C+f_2D
Total	$A+C$	$B+D$	N	$f_1(A+C)$	$f_2(B+D)$	n

9. Suppose the case-control study samples a fraction of f_1 of those diseased and a fraction of f_2 of those without disease. The risk in the population is $A/(A+B)$ for exposed people and $C/(C+D)$ for unexposed.

10. Relative risk is $A(C+D)/(C(A+B))$.

11. The expected values of these quantities in the case-control sample are

$$f_1A/(f_1A+f_2B) \neq A/(A+B), \quad f_1C/(f_1C+f_2D) \neq C/(C+D)$$
$$\frac{f_1A(f_1C+f_2D)}{f_1C(f_1A+f_2B)} \neq \frac{A(C+D)}{C(A+B)}$$

12. The odds of the disease of the population are for exposed: $f_1A/(f_2B) \neq A/B$, for the unexposed: $f_1C/(f_2D) \neq C/D$.

13. However, the odds ration for the population and case-control study are the same

$$\frac{(f_1A)(f_2D)}{(f_2B)(f_1C)} = \frac{AD}{BC}$$

14. Polytomous exposure: The risk factor is measured at several levels. A base level is chosen and compared with other levels.
Ex. Case-control study of Escherichia coli by Fihn et al. (1996). Cases were women aged 18 to 40 years selected from the records of a health maintenance organization in Washington state, US. Controls were randomly sampled from the same database, chosen from those women without E. coli infections within the same age structure as the cases.

15. Odds ratios are given for ethnicity relative to the chosen base group.

Table 6.4. Ethnicity by case–control status in a study of *E. coli* in Washington state.

Ethnicity	Cases	Controls	Odds ratio (95% confidence interval)
Caucasian	514	541	1
African American	25	25	1.05 (0.60, 1.86)
Hispanic	13	5	2.74 (0.97, 7.73)
Asian	32	21	1.60 (0.91, 2.82)
Other	20	37	0.57 (0.33, 0.99)
Total	604	629	

16. Selection of Cases:

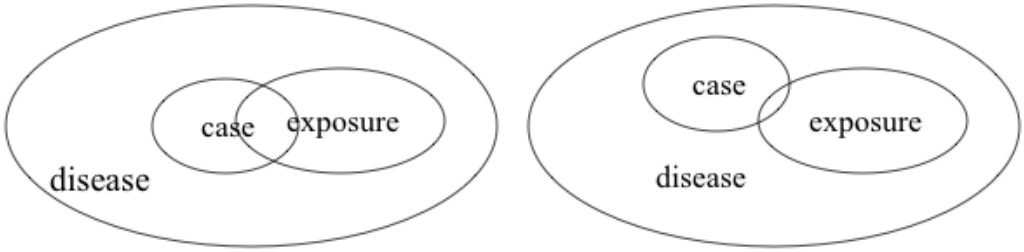
    (a) Definition: Before cases can be selected, disease needs to be precisely defined. Definition cannot be too broad.

    (b) Inclusion and exclusion criteria: Subjects with the disease are considered eligible only if they satisfy certain inclusion and exclusion criteria e.g. Father smoking and birth defect.

(c) Incident or prevalent?: Incident disease is generally a better criterion for case selection e.g. Diabetes and drinking.

(d) Source: Cases are usually selected from medical information systems, such as hospital admission records, Pathology department records, sickness absence forms and disease registers.

(e) Consideration of bias: Case selection is biased if $p_{case} \neq p_{disease}$, where p_{case} is the probability of exposure among cases and $p_{disease}$ is the corresponding probability for diseased in the whole population. Bias occurs when the chance of becoming a case depends on the fact of exposure to the risk factor.



17. Examples 1: Hormone replacement therapy (HRT) is considered as a risk factor for cervical cancer. Women patients registered with a particular health center who take HRT daily are required to attend at an annual HRT clinic as a condition of renewal of their prescription. At the clinic they undergo a cervical smear. For other female patients, this is done at intervals of several years. Undetected cervical cancer is more likely among those who are not receiving HRT and $p_{case} > p_{disease}$.

    Example 2: Pearl (1929) studied data from autopsies and found that cancer and tuberculosis (TB) were rarely found together. He suggested that cancer patients might be treated with tuberculin (the protein of the TB bacterium). It happened that people who died from cancer and TB were less likely to be autopsied than those who died from cancer alone. $p_{case} < p_{disease}$. This type of bias is also called Berksons bias (bias arises because of the source used).

18. Selection of controls: Control should be representative subgroup of members of the same base group that gave rise to cases, who have the particular characteristic that they have not (yet) developed the disease. Selection of controls is the most challenging aspect of case-control study design.

19. General principles:

(a) Controls should be drawn from among those who are free of the disease being studied.

(b) Controls should be drawn from the same general population of cases.

(c) The source from which controls are selected should not give rise to bias error. (Bias if $p_{control} \neq p_{undiseased}$).

(d) Controls should have some potential for the disease. More than one control groups can be selected and compared. e.g. Hospital controls, community controls

(e) Comparing two different controls. Female cases of hip fracture aged 45 years or more. Odds ratios (with 95% confidence intervals) in a study of hip fracture using two different control groups.

Odds ratios (with 95% confidence intervals) in a study of hip fracture using two different control groups.

Risk factor	Hospital controls	Community controls
Fall in past 6 months	1.08 (0.71, 1.53)	1.70 (1.22, 2.35)
Current (vs. never) smoking	1.30 (0.85, 1.98)	2.49 (1.61, 3.83)
Stroke	1.36 (0.87, 2.11)	2.51 (1.60, 3.94)
Poor vision	2.62 (1.27, 5.37)	1.42 (0.81, 2.48)

[a] Adjusted for several potential confounding variables.