

Chapter 6

Type of Studies - II : Case - Control Studies

When one wishes to find out whether exposure to X is associated with a disease Y one would need a study with the following features:

- a). A group of people who have the disease and a comparison group of healthy persons free of the disease to serve as controls.
- b). The two groups must be comparable in all respects except for the factor of interest. The reason is that if one wishes to show that an associated factor is a cause, it is necessary to control for all important differences other than the exposure factor of interest.
- c). There must be enough people in the study (i.e. amongst 'cases' and 'comparison' groups) so that chance does not play a large part in the observed differences.

Case reports and case series do not meet these requirements. However, in a large cross-sectional study it is often possible to separate out a group who have the condition of interest and compare them with a healthy group for a number of possible risk factors. But the chances of spurious associations are high. Positive findings, if any, can only be speculative. They can, however, serve as useful pointers for future case-control studies.

Studies that compare the frequency of a possible risk factor in a group of cases and a group of controls are called case control studies (see figure 6.1).

The five main steps in the design of case-control studies are: hypothesis development; definitions of cases and controls; selection of cases; selection of controls and determination of exposure status. Clear cut and unambiguous definitions, unbiased selection procedures, and well thought out ways of measuring exposure help to make successful case-control studies.

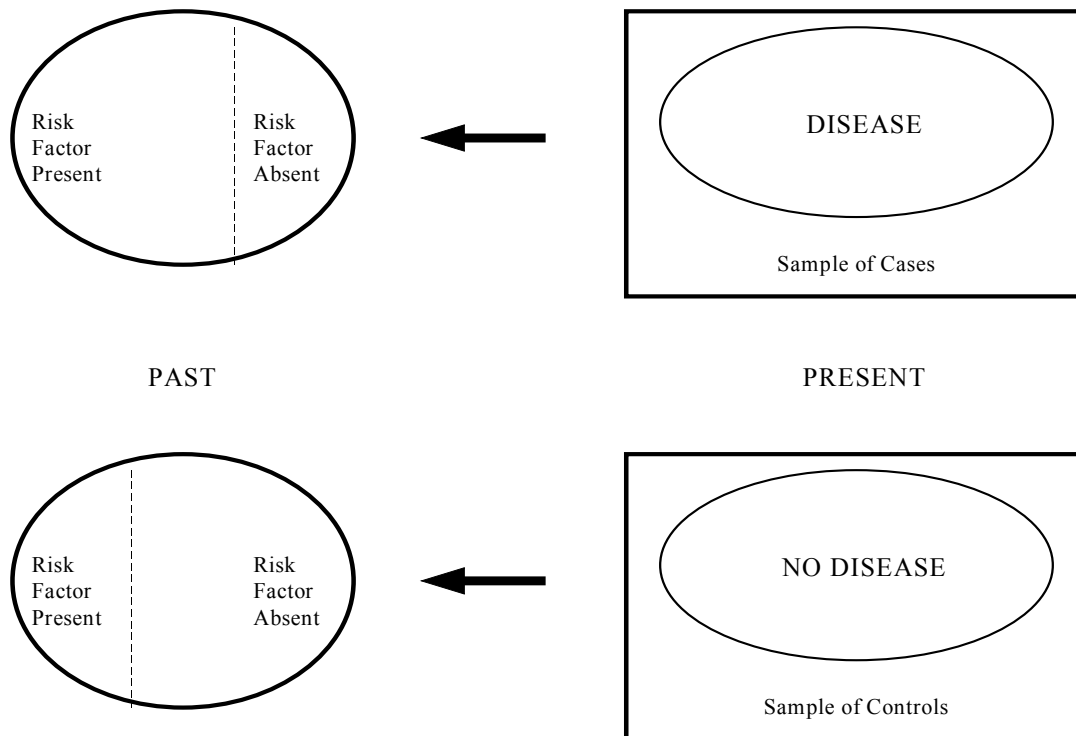


Figure 6.1: Case-Control Studies

Case-control studies are particularly susceptible to bias because of either the selection process for cases and controls being different or because of differences in the reporting or recording of information about exposure due to prior knowledge of the disease status of the subjects. The strength of case-control studies lies in being able to draw conclusions about diseases that occur several years following exposure in a cost-effective manner.

In planning case-control studies one has to pay careful attention to selection of cases, of controls, and about defining the exposure.

Selection of Cases

It is clear that difficulties would arise if the study subjects do not represent the larger population of those with the outcome of interest. Patients admitted to specialist referral hospitals might not reflect accurately the world of patients with the disease. On the other hand if the diagnosis is incorrect e.g.

patients with equivocal symptoms being included the true association between exposure and disease may be missed. In such a case it may be advisable to perform separate analyses for cases according to diagnosis categorized as definite, probable or likely. Besides clear case definition the other points to consider are the following:

- are cases to be "incident" (newly diagnosed) or "prevalent" (existing at commencement of the study). Inclusion of prevalent cases saves time and costs, but a problem may arise in interpretation because prevalent cases represent determinants of duration in addition to cause of the disease. Secondly, the temporal sequence in the case of "incident" studies can be more clearly established than in "prevalent" studies. Hence whenever possible it is better to enroll cases which have been newly diagnosed in accordance with clear case definitions.
- how are cases to be identified? Cases can be selected from a variety of sources. The commonest approach is to select newly diagnosed cases from a hospital because it is relatively easy and inexpensive to recruit a sufficiently large number of subjects in the hospital setting. The second approach is to recruit patients from amongst the general population of a defined area. The advantage in this latter case is that one can avoid the influence of factors that may have led individual subjects to seek medical care from a particular health facility. A population based study also allows one to describe the rates of the disease being studied in the population. Population based studies are not routinely undertaken because of their costs.
- what would be the mechanisms by which cases are identified?
- what should be the inclusion and exclusion criteria? Definitions and diagnostic criteria are important so that true cases do not get diluted by wrongly diagnosed ones.

Selection of controls

In the selection of controls the objective is to derive the general rates of exposure. The guiding principle is that the controls should be from amongst the non-diseased population such that they would have stood an equal chance of inclusion as cases had they developed the disease. The selection of controls is perhaps the most critical issue in designing case-control studies. Results can become biased by selectively including those who either under-estimate or over-estimate exposure.

The points to consider whilst selecting controls are the following:

- **source of controls** (hospital patients without the disease under consideration, neighbourhood controls, random sample of population, siblings, and so on). Controls are selected not to represent the entire non-diseased population but the people who would have been identified as cases if they had the disease. They should be comparable to the cases and any exclusion criteria applying to cases must equally apply to them also. Different sources for recruiting controls must be considered with these requirements in mind according to the nature of the study and the characteristics of cases. The commonest source of controls are hospital patients who came in with similar symptoms as cases but were diagnosed as not having the disease. This has the advantage of convenience besides the fact that the controls are likely to have undergone the same selection procedures for referral to hospital as the cases. Controls drawn from the general population may not be very co-

operative, or their recall of exposure may not be that reliable. Friends, neighbours, family members are a third source of controls. The advantage with this group is that a number of confounding factors like socioeconomic status, ethnic background, housing and so on can be avoided. On the other hand they are likely to be too similar to the cases to help bring out any differences. Thus, each type of control group has its advantages and disadvantages. If after careful consideration, one control group may be thought to have a specific drawback another different group may be enrolled. If similar results are consistently obtained with each different type of control group it adds credibility to the findings. Such an approach will end up with having more controls than cases. Does it matter? The optimum ratio of cases to controls is 1:1. But if cases are difficult to enroll because of expense or because of the rarity of the disease the number of controls per case may be increased to the ratio of 1:4. This increases the statistical power. Any further increase raises the statistical power only slightly and is not much helpful.

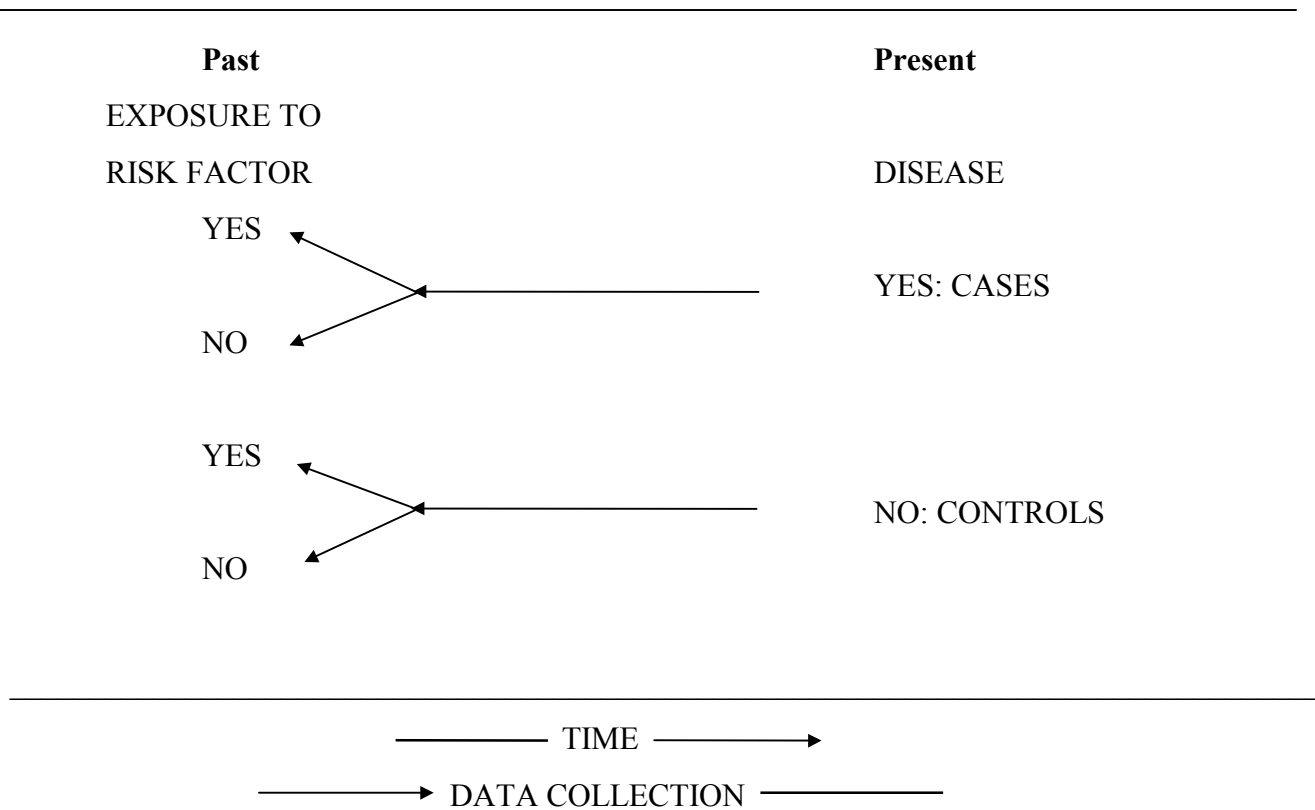
- **inclusion and exclusion criteria** for the controls. These should be the same as in the cases.
- **matching with cases** e.g. by age, sex, social class, residence and so on. This is especially the approach to use in population-based studies. In hospital based studies a common approach is to do pairing. Each case is paired with the subsequent admission. For example, in studies on risk factors for low birth weight each case is paired with a subsequently born normal baby.

Information about exposure

The following should be clarified at the outset:

- The risk factors of interest. How would the exposure be defined and its intensity graded? How the information is to be collected? Would it be by means of a questionnaire, from existing medical records, employer's records as in studies of industrial toxins, or from some other source?
- Whatever the procedure decided upon it should be the same for cases and controls. As far as possible the place and circumstance for interview should be similar to avoid a situation where cases are interviewed in hospital and controls at home.
- If possible, the interviewer should be unaware ("blind") as to the identity of cases and controls as well as the hypothesis being tested. This is to avoid undue probing of cases for exposure information.

The basic design of case control studies is as follows:



Information about exposure to the risk factor depends upon past memory, and therein lies a potential for bias in all case control studies. It is difficult not to interpret the past in the light of one's present condition.

Case control studies should not be confused with cohort studies, even though they are both observational studies concerned with risk factors. A distinguishing feature of case control studies is that cases have the **outcome of interest** (i.e. disease or death) at the time when information on risk factors is sought. In the case of cohort studies, people are free of disease at the **beginning of observation**, but are exposed or unexposed to risk factors.

As we have seen cases and controls can be selected in two ways relative to the course of disease. One is to select new (incident) cases, that is cases of disease as they arise. Controls are selected from those without the disease in the same setting at the same time. The other way is to select existing (prevalent) cases from a defined population. In the latter case, a population is first defined and then at a point in time cases and a much larger number of controls are identified and their exposure determined. If prevalent cases and controls are to be used, they should all be from defined, and relatively unselected population, so that there is less opportunity for selection bias.

Using incident cases is a stronger way of conducting case control studies. This is obvious because both the diagnosis and exposure to the risk factor are carried out under the watchful eye of the investigator.

One fundamental problem with case-control studies is that the question addressed by such studies is usually about incidence. The research question is “Does exposure to X result in new cases of Y? ” When prevalent cases are used the question becomes: “Is exposure to X a risk factor for having disease Y?” And as we have seen, having Y is determined both by its incidence and duration.

Analysis

The main purpose in analysis is to measure the chances (odds) of exposure to the risk factor amongst cases as compared to the control group. The way of doing this is set out below:

| | Cases | Non-cases |
|-------------------------------|--------------|------------------|
| Exposed to risk factor | A | B |
| Not exposed | C | D |
| | A + C | B + D |

$$\text{Odds ratio} = AD \div BC$$

The stronger the association between the exposure and disease, the higher the odds ratio. On the other hand, if the frequency of exposure is lower among cases, the odds ratio will be less than one indicating protection.

The calculation of the odds ratio is a simple procedure. During analysis it is tempting to calculate odds ratios for a variety of likely risk factors. It is important to draw a distinction between testing a hypothesis specified in advance before designing a study and other associations emerging after data has been collected. Such data derived associations must be looked upon as spurious. They may serve as hypotheses for future studies, and no more. For each case-control study the hypothesis for which the study was designed is the one that matters.

Adjustment for confounding

By virtue of their association with the exposure variable confounding variables can distort the result and lead to a mistaken assessment of risk. For example, if one is studying the effect of parents' income and the nutritional status of their children, mothers' education may be a confounding factor. Low parental income is associated with malnutrition in children, and so also is lack of education of the mother. Mothers in the low-income group are more likely to have less education. Stratification is a way of handling such confounders. One can stratify the subjects by level of education. Then if the association between parental income and malnutrition in their children holds for all levels of mothers' education then low income as a risk factor for malnutrition is proven.

Advantages of case control studies.

1. The investigator can identify cases not constrained by the natural frequency of disease, and can yet make comparisons. When the outcome under study is relatively rare, a large number of exposed individuals would have to be observed for a prolonged period of time for the outcome to happen. It is more economical to start with a number of individuals who have already experienced the outcome, and look back into the past to obtain a history of exposure. If the outcome occurs after a long period of exposure the observer might not be able to follow subjects for a prolonged period. Moreover, dropouts are very likely during a long period of follow-up. By starting at the end of the process these difficulties are avoided. Case-control studies are also useful for examining multiple risk factors causing the same outcome (e.g. smoking, lack of exercise, fatty diet, and high blood cholesterol causing coronary artery disease). Here care must be taken to recruit a large enough sample size for the effects of each of the factors to be assessed satisfactorily.
2. The waiting time for the answer becomes relatively short. Many diseases have a long latency period between exposure and first symptoms. In the case of some cancers it may be as long as 20 to 25 years.
3. Case-control studies are useful for generating hypotheses in conditions where the etiology is not known. A large number of possible predictor variables can be examined by the case-control approach. The drawback is that only one outcome can be studied, and that outcome is the basis of drawing the samples. Cross-sectional studies and cohort studies, on the other hand, can examine several outcomes.

Because of their ability to address important questions rapidly and effectively, case control studies have come to play an increasingly prominent role in the medical literature. The case control approach is a relatively strong method to find answer to a cause-effect relationship in a short time.

Disadvantages of case-control studies

Case-control studies also suffer certain disadvantages. Firstly, the result is an approximation, as compared to cohort studies. What the case-control study gives is those with exposure ÷ those with the outcome. (It does not give those with the outcome ÷ those exposed to risk). And that is the crucial difference between Odds Ratio and Relative Risk, (to be described in the following chapter). Instead of relative risk of developing a disease when exposed to the given risk factor, what the researcher obtains is relative odds of being exposed to the risk factor if disease is present. Both are measures of the strength of relationship between exposure and outcome, but are numerically equal only in rare diseases when disease rates in unexposed people are of a magnitude of 1/100 and less. In practice this does not pose a major problem. Most diseases that are studied by case-control studies have lower rates than this.

Secondly, determining an individual's past experience of exposure is not as efficient as determining it in the present. For the former the researcher must rely on the history as provided by the subject or get it from the records. In the case of the latter the measurement of exposure is under the researcher's scrutiny.

Thirdly, the size of the Odds Ratio is governed by the type of the control group, and so is open to manipulation. Thus skeptics are not satisfied because the reasoning is progressing from an effect (disease) backwards in time to a possible cause (previous exposure).

Bias in case control studies.

An increased susceptibility to bias is a major difficulty with case-control studies. Manipulation is possible in the case control approach. For example, the investigator can manipulate the comparison groups and obtain a high or low odds ratio without altering the cases or the measurement of exposure. Bias can arise from two sources viz. i). separate sampling of cases and controls and ii). retrospective assessment of exposure.

Sampling bias

Cases and controls are comparable if they are equally likely to have been exposed to the factor of interest. Both the groups must seem to have had an equal chance of being exposed to the factor of interest. And so, ensuring comparability between cases and controls requires the careful consideration of the circumstances under which an individual becomes exposed.

Secondly those cases which were not correctly diagnosed, or have not yet come forth for treatment, or have die are unlikely to be included in the sample. Thus the sample does not represent truly all cases. Secondly, all those among the cases would have had a full diagnostic work-up performed. But not so the controls since they are considered healthy. This can be a potential source of bias. If possible, similar diagnostic procedures should be used on the controls as in cases.

Measurement bias

There is a very strong likelihood of differences in recall of exposure between cases and controls. Those with a disease are more likely to remember exposure to risk factors compared to those without it.

Secondly, interviewers are likely to question cases more searchingly for exposure than controls if they know who is a case and who is not.

Strategies for avoiding bias

Selecting cases.

In case control research the cases should preferably be new (incident), rather than already existing ones (prevalent). This is because both incidence and duration govern prevalence. Duration is determined by the rate at which patients leave the disease state because of death or recovery, or persist in it because of its slow course. Thus, risk factors for prevalent cases are risk factors for both incidence and duration, and the relative contribution of each cannot be distinguished.

Selecting controls.

The selection process can be a major source of bias. The fundamental question is the following: Are controls as likely to be exposed as cases? Any difference between cases and controls that might be related to becoming exposed could distort the odds ratio and make it an inaccurate estimate of the true risk.

Several strategies are available for choosing the right controls.

1. Selecting both cases and controls from unbiased samples of the same population can minimize selection bias. If cases are a sample of all cases arising in a defined population, then controls should be a random sample of all the other people in the same population. Alternatively, cases and controls can be unbiased sample of the same cohort i.e. a selected group of people who are being observed over time. This strategy is called a population-based or nested case control study. Nested because it is inside a cohort. Selection of cases and controls from a defined population or cohort is ideal, but in practice selection from hospitals is easier. The problem is that such selection is open to error, because hospital patients are a biased sample of the people in the community. And the results are to be eventually applied to the community where the risk factor is operating.
2. Another strategy is to do careful matching. This poses the greatest challenge. In matching, for each case one or more controls are selected who possess characteristics in common with the case. The usual way of matching is by age, sex, race, occupation, education and socio-economic status. These are commonly related to disease. But matching has to extend beyond these attributes when other factors are known to be important. In such cases errors can occur if predictors that can undergo change over time are chosen like for example, income; blood urea or cholesterol levels.

When performed properly matching maximizes the information available from a set of cases and controls, because it reduces the differences between the groups with regard to a number of determinants of disease except for the one being looked for. In other words at each level of the matching variable a balance is obtained in the number of cases and controls. For example, if matching is done by say age group, then equal number of cases and controls are obtained for each age group. Age If race was an additional matching variable, then cases and controls would be balanced by age group in each racial category.

But there is also a problem. If the investigator happens to match for a factor which is itself related to exposure, then there is a strong chance that cases and controls would end up with the same history of exposure. This is called **overmatching**, and can result in a falsely low estimate of risk.

If the prime objective of matching is to avoid bias then analysis should take it account. Matching is only the first step in a two step process viz. i). Matched design, followed by ii). matched analysis. An analysis that ignores the matching will typically result in a disease-exposure odds ratio that is biased towards unity. Therefore, unless there are very good reasons to match one is better off avoiding it. How to take matching into account during analysis is demonstrated below:

An example of analysis by matching

Say matching is done by income into three groups as high, medium and low. Then for each of the three groups a table is set out in the usual manner as for measuring the odds ratio.

| | High | | Medium | | Low | |
|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| | Case | Non-case | Case | Non-case | Case | Non-case |
| Exposed | a ₁ | b ₁ | a ₂ | b ₂ | a ₃ | b ₃ |
| Not exposed | c ₁ | d ₁ | c ₂ | d ₂ | c ₃ | d ₃ |
| Number In group | n ₁ | | n ₂ | | n ₃ | |

$$\text{Odds ratio} = [a_1d_1/n_1 + a_2d_2/n_2 + a_3d_3/n_3] \div [b_1c_1/n_1 + b_2c_2/n_2 + b_3c_3/n_3]$$

3. Another approach is to choose more than one control group, and try to draw them from different sources. For example, if cases are drawn from a hospital, one control group may be drawn from other patients in the hospital and a second control group from the neighbourhoods in which cases live. If similar odds ratios are obtained using different control groups it provides evidence against bias because it is unlikely that bias would affect dissimilar groups to the same extent. But if the estimates of relative risks are different, it is a signal that one or both controls are biased, and one then needs to investigate where bias lies.

A somewhat different approach is to choose one set of controls to probe for possible bias in another. If the researcher suspects that a high odds ratio obtained from one set of controls may be wrong, then he might select another group of controls such that they are likely to produce a low odds ratio. If calculations of risk from the two control groups are not very different, the concern is unfounded.

One rule to remember is that if cases are all those patients living in a defined population, or a representative sample of them, then controls should also be likewise. If cases are a biased sample of all cases, as for example referred cases in a hospital, then controls should also be selected with similar biases. Regardless of how the investigator chooses the case and control groups, the method of selection should have been established before the research is commenced. It is also important to bear in mind that subjects without the outcome of interest (i.e. controls) may differ in other ways than exposure from cases.

Measuring Exposure

Having dealt with selection bias, the investigator must next deal with avoiding bias in measurement of exposure. In other words there is a need for validly measuring exposure after the disease or outcome has occurred.

Three kinds of *measurement bias* can occur:

1. The presence of the outcome directly affects the exposure. For example, people who have chronic lung disease tend to move out of the city environment into the country. A higher proportion of lung disease may then be found if the sample comprises a predominantly rural population.
2. The presence of the outcome affects the subject's recollection of the exposure. Attempt should be made to equalize recall by using multiple ways of inquiring about exposure, or by checking the answer with alternative source of data.
3. Knowledge of the outcome affects the investigator's measurement or recording of the exposure. This is likely to occur when the interviewer knows the disease status of the subject and the research question. Cases would then be questioned more thoroughly about exposure.
4. The researcher may use data about measurements which were made before the outcome occurred. The difficulty is that the measurements may not have been recorded; or if recorded, they may not be very reliable.

A way of avoiding bias is to define exposure by type, duration and dose. A careful note should be made as to whether exposure preceded the disease, and by how long. Exposure should be recorded without bias related to the disease status. This can be achieved by keeping the interviewer "*blind*" as to the status (whether case or control) of the subject.

Appendix 6.1

Criteria For a Sound Case Control Study

1). Were the cases

a). entered in the study at onset of disease?

(If prevalent cases then consider how differences in duration of disease, if related to exposure, could affect estimate of relative risk).

b). described concerning criteria for diagnosis?

2). Were controls comparable to cases as likely to be exposed for reasons other than being a case?

One or more strategies can be used:

a). Select samples of cases and controls from the same defined population.

b). If cases are biased samples of all cases in the population, select controls with similar biases i.e. opportunity for exposure, similar setting, inclusion/exclusion criteria, etc.

c). Control for the effects of known extraneous variables by matching, stratified analysis, and so on.

d). Compare results from more than one control group, selected to uncover potential bias.

3). Have cases and controls undergone similar efforts to detect the disease (particularly if the disease is known to have a long "silent" phase?).

4). Was exposure:

a). Defined by type, dose, duration etc?

b). Known to precede disease?

c). Recorded without bias related to disease status?

Appendix 6.2

Determination of sample size for Case-control Studies

1). Formula for calculating size of **each group**

π_1 = proportion of controls exposed

OR = Odds Ratio

π_2 Proportion of cases exposed, calculated from

$$\pi_2 = \frac{\pi_1 \cdot \text{OR}}{1 + \pi_1 (\text{OR} - 1)}$$

$$\text{Required sample size} = \frac{\{\mu \sqrt{[\pi_1(1-\pi_1) + \pi_2(1-\pi_2)]} + v \sqrt{[\pi_m(1-\pi_m)]}\}^2}{(\pi_2 - \pi_1)^2}$$

$$\text{where } \pi_m = \frac{\pi_1 + \pi_2}{2}$$

and the values of μ and v as in the table below:

| Power | μ | Significance Level | v |
|-------|-------|--------------------|------|
| 80% | 0.84 | 10% | 1.64 |
| 90% | 1.28 | 5% | 1.96 |
| 95% | 1.64 | 2.5% | 2.23 |
| 97.5% | 1.96 | 1% | 2.58 |
| 99% | 2.33 | | |

2). Tables for sample size in case-control studies.

Tables 6.A1 and 6.A2 show the number of cases required to be included in order to have a specified power of detecting a significant difference for a given risk ratio and a given proportion in the control group who are exposed to the risk factors. In each cell three numbers are shown. They indicate the ratio of controls to cases in a descending order of 1,2 and 4. For example in **Table 6.A1**, for a power of 80% and a chance of detecting a significantly increased risk at 5% ($\alpha = .05$) level, when the true risk ratio is 4 and 50% of the population is exposed to the risk factor of interest (say smoking), then from the table we find the three numbers are 37, 28 and 23. If the ratio of cases to controls chosen is 1:2 then we would need 56 controls for 28 cases. If the ratio chosen is 1:3, then the number of controls needs to be 69 for 23 cases.

In **Table 6.A2** for a power of 80% for detecting a significant risk at 1% level ($\alpha = 1\%$), when the true risk ratio is 4 and 50% of the population is exposed to the risk factor of interest, we find from the table that the three numbers are 56, 43 and 36. If the ratio of cases to controls is chosen as 1:2 then we need to recruit for 43 cases 86 controls. If the ratio is 1:3, then we need 108 controls for 36 cases.

Table 6.A1. Alpha = 0.05 and Power of 80%

Proportion in the Control Group who are exposed to the risk factor

| Risk Ratio | 0.01 | 0.1 | 0.1 | 0.2 | 0.2 | 0.3 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 |
|-------------------|-------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| 1.5 | 6672 | 1413 | 763 | 550 | 448 | 390 | 356 | 325 | 325 | 352 | 419 | 571 |
| | 4901 | 1041 | 563 | 407 | 332 | 290 | 265 | 243 | 244 | 265 | 317 | 434 |
| | 4009 | 854 | 462 | 335 | 274 | 240 | 219 | 203 | 203 | 222 | 266 | 365 |
| 2 | 2087 | 243 | 135 | 101 | 84 | 76 | 71 | 69 | 72 | 82 | 102 | 146 |
| | 1512 | 125 | 98 | 74 | 62 | 56 | 53 | 51 | 54 | 62 | 78 | 112 |
| | 1220 | 140 | 79 | 60 | 51 | 46 | 43 | 42 | 45 | 52 | 66 | 95 |
| 2.5 | 1114 | 243 | 135 | 101 | 84 | 76 | 71 | 69 | 72 | 82 | 102 | 146 |
| | 799 | 189 | 98 | 74 | 62 | 56 | 53 | 51 | 54 | 62 | 78 | 112 |
| | 638 | 140 | 79 | 60 | 51 | 46 | 43 | 42 | 45 | 52 | 66 | 95 |
| 3 | 732 | 151 | 91 | 69 | 58 | 53 | 50 | 49 | 53 | 61 | 78 | 112 |
| | 521 | 115 | 66 | 50 | 43 | 39 | 37 | 37 | 40 | 47 | 59 | 86 |
| | 412 | 92 | 53 | 40 | 35 | 32 | 31 | 31 | 33 | 39 | 50 | 73 |
| 4 | 420 | 94 | 55 | 42 | 37 | 34 | 33 | 33 | 37 | 43 | 56 | 82 |
| | 296 | 57 | 39 | 31 | 27 | 25 | 24 | 25 | 28 | 33 | 43 | 64 |
| | 231 | 53 | 31 | 25 | 22 | 20 | 20 | 21 | 23 | 28 | 36 | 54 |
| 5 | 290 | 66 | 39 | 31 | 27 | 26 | 25 | 26 | 29 | 35 | 46 | 69 |
| | 203 | 47 | 28 | 22 | 20 | 19 | 19 | 20 | 22 | 27 | 36 | 54 |
| | 157 | 37 | 22 | 18 | 16 | 15 | 15 | 16 | 19 | 23 | 30 | 46 |
| 7.5 | 161 | 39 | 24 | 20 | 18 | 17 | 17 | 19 | 22 | 27 | 36 | 55 |
| | 112 | 27 | 17 | 14 | 13 | 13 | 13 | 14 | 17 | 21 | 28 | 43 |
| | 85 | 21 | 13 | 11 | 10 | 10 | 10 | 12 | 14 | 17 | 24 | 36 |
| 10 | 111 | 28 | 18 | 15 | 14 | 14 | 14 | 16 | 19 | 24 | 32 | 50 |
| | 77 | 20 | 13 | 11 | 10 | 10 | 11 | 12 | 14 | 18 | 25 | 39 |
| | 58 | 15 | 10 | | | | | 10 | 12 | 15 | 21 | 33 |
| 15 | 69 | 19 | 13 | 11 | 11 | 11 | 12 | 13 | 16 | 21 | 29 | 45 |
| | 48 | 18 | | | | | | 10 | 12 | 16 | 22 | 35 |
| | 36 | 10 | | | | | | | 10 | 13 | 19 | 29 |
| 20 | 51 | 15 | 10 | 10 | | 10 | 10 | 12 | 15 | 19 | 27 | 42 |
| | 35 | 10 | | | | | | | 11 | 15 | 21 | 33 |
| | 26 | | | | | | | | | 12 | 17 | 28 |

Table 6.A2. Alpha = 0.01 Power = 80

Proportion of the control group who are exposed to the risk factor

| Risk ratio | 0.01 | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 |
|-------------------|-------------|-------------|------------|-------------|------------|-------------|------------|------------|------------|------------|------------|------------|
| 1.5 | 10583 | 2245 | 1211 | 873 | 711 | 620 | 565 | 515 | 515 | 559 | 664 | 906 |
| | 7698 | 1638 | 887 | 642 | 524 | 458 | 419 | 385 | 387 | 422 | 505 | 693 |
| | 6247 | 1332 | 724 | 525 | 430 | 377 | 346 | 319 | 323 | 354 | 425 | 585 |
| 2 | 3266 | 703 | 386 | 283 | 234 | 207 | 192 | 181 | 186 | 207 | 253 | 354 |
| | 2328 | 504 | 278 | 206 | 171 | 153 | 142 | 135 | 140 | 158 | 194 | 274 |
| | 1851 | 403 | 224 | 166 | 139 | 125 | 117 | 112 | 117 | 133 | 165 | 234 |
| 3 | 1128 | 249 | 140 | 106 | 90 | 82 | 78 | 76 | 82 | 95 | 119 | 173 |
| | 784 | 175 | 100 | 76 | 65 | 60 | 57 | 57 | 62 | 73 | 93 | 136 |
| | 606 | 136 | 79 | 61 | 52 | 48 | 47 | 47 | 52 | 61 | 79 | 116 |
| 4 | 641 | 144 | 84 | 64 | 56 | 52 | 50 | 51 | 56 | 66 | 85 | 126 |
| | 439 | 100 | 59 | 46 | 40 | 38 | 37 | 38 | 43 | 51 | 67 | 100 |
| | 333 | 77 | 46 | 36 | 32 | 31 | 30 | 31 | 36 | 43 | 57 | 86 |
| 5 | 440 | 101 | 50 | 47 | 41 | 39 | 38 | 40 | 45 | 54 | 70 | 105 |
| | 298 | 70 | 42 | 33 | 30 | 28 | 28 | 30 | 34 | 42 | 56 | 84 |
| | 223 | 53 | 32 | 26 | 24 | 23 | 23 | 25 | 29 | 35 | 47 | 72 |
| 7.5 | 243 | 58 | 36 | 29 | 27 | 26 | 26 | 28 | 33 | 41 | 55 | 83 |
| | 162 | 40 | 25 | 21 | 19 | 19 | 19 | 21 | 25 | 32 | 44 | 67 |
| | 119 | 30 | 19 | 16 | 15 | 15 | 16 | 18 | 21 | 27 | 37 | 58 |
| 10 | 167 | 42 | 27 | 23 | 21 | 21 | 21 | 24 | 28 | 36 | 48 | 74 |
| | 111 | 28 | 19 | 16 | 15 | 15 | 16 | 18 | 22 | 28 | 39 | 60 |
| | 80 | 21 | 14 | 13 | 12 | 12 | 13 | 15 | 18 | 24 | 33 | 52 |
| 15 | 104 | 28 | 19 | 17 | 16 | 16 | 17 | 20 | 24 | 31 | 43 | 67 |
| | 68 | 19 | 13 | 12 | 12 | 12 | 13 | 15 | 19 | 25 | 34 | 54 |
| | 59 | 14 | 10 | | | 10 | 10 | 12 | 16 | 21 | 29 | 47 |
| 20 | 76 | 22 | 16 | 14 | 14 | 14 | 15 | 18 | 22 | 29 | 40 | 63 |
| | 50 | 15 | 11 | 10 | 10 | 11 | 12 | 14 | 17 | 23 | 32 | 51 |
| | 35 | 11 | | | | | | 11 | 14 | 19 | 28 | 44 |