

# Analytical Study Designs

Pages with reference to book, From 54 To 56

Melvyn A. Lobo, Asma Fozia Qureshi ( Department of Community Health Sciences, Aga than University, Karachi. )

## What is a Case-Control Study?

Case-control studies look for an association between a disease and exposure to a potential risk factor. The investigator first selects cases (persons with disease) and appropriate controls (persons without the disease). The researcher next obtains data on exposure to the risk factor(s) of concern in both groups.

When Is it appropriate to select a Case-Control Study design?

The case-control study is generally quick and Inexpensive to conduct. It is well suited for examining an association between a risk factor and a disease that is rare or has a long latent period. The researcher generally selects cases from persons already diagnosed to have the disease/outcome of concern. If the required number of cases for analysis are not already present, the researcher may have to recruit cases as they occur to obtain the necessary number of cases.

## How do I go about conducting a Case-Control Study?

**Conducting a case-control study is a 3 step process. The first step is to select cases. In doing so you have to pay particular attention to:**

**i. Case definition:** Cases i.e., diseases/outcome of interest should be defined using strict inclusion criteria to prevent an overlap between cases and controls and thus minimize results which are not easily explained. For the same reasons, it is also important to use exclusion criteria. For example, in identifying risk factors for cancer of the uterus, it would be appropriate to study cancer of the cervix and cancer of the corpus uteri separately, as risk factors for cervical cancer may differ from those for cancer of the corpus uteri.

**ii. Selection of cases:** Cases can be selected from records (hospital and other records) or identified through a community survey. Random selection of cases for study from a list of identified case makes the results from such a study more readily generalizable to the larger population. The second step is the selection of controls (those without the disease) which is necessary for comparison.

**i. Control definition:** Controls should be comparable to case except with respect to the exposure under study.

**ii. Selection of Controls:** Controls can similarly be chosen from the same hospital as the cases, or from the surrounding community. They should preferably be selected in a random manner, e.g., through the use of random tables, to avoid a systematic error. Hospital controls are those with conditions other than the disease being studied or diseases associated with the risk factor(s) under study. For example, if one wishes to study the relationship between tobacco use and lung cancers, inappropriate controls would be persons with chronic bronchitis. The proportion of smokers in both groups might be similar leading to the false conclusion that smoking is not related to lung cancer. The advantages of using hospital controls are that they are easily available. Controls chosen from the community are preferable to hospital controls. However, these persons are less likely (compared to cases) to recall information with respect to exposure- recall bias. Controls could thus be selected from relatives, friends or neighbours of cases as they would be more likely (because of interest in the case) to provide information on exposure to the risk factor(s) under study. Once the source of the control series has been determined, the next question to consider is the number of controls. Generally, the higher the ratio of controls to the case, the more likely the chance of detecting a difference in exposure to the risk factor. The ratio for cases-to-controls can vary from 1:1 to 1:4. The third step is ascertainment of exposure status. Any source of information must be carefully considered in terms of its accuracy and comparability of information for all study groups. Sources could be (i) interviews with study subjects or surrogates, e.g., spouses of subjects or mothers of children, or (ii) data recorded in records (physician records, hospital records,

death certificates, etc.). Procedures used to obtain information must be as similar as possible for cases and controls and free from bias.

### How do I analyze data from the Case-Control Study?

Cases and controls should first be compared to ensure similarity with respect to factors other than those being examined, e.g., age, sex, etc., that could be associated with the risk of developing the outcome under study. The measure of risk calculated from a case-control study is the odds ratio. This is a measure of association (not a measure of causality) which describes the odds of a case being exposed to that of controls being exposed to the suspected risk factor.

### How do I calculate the Odds Ratio?

If the researcher were to create a two-by-two table in a manner given below, four cells emerge: To calculate the odds ratio, multiply the numbers in cells A and D and divide this product by the product obtained from multiplying the figures in cells B and C.

	Disease (cases)	No disease (controls)
Exposed	Cell A Cases who were exposed	Cell B Controls who were exposed
Unexposed	Cell C Cases who were 'not exposed	Cell D Controls who were not exposed

### How Do I Interpret the Odds Ratio?

**There are three possibilities when calculating an odds ratio:**

1. An odds ratio of one means that the odds of exposure in cases and controls is the same, indicates that the risk factor is not associated with the disease.
2. An odds ratio of more than one signifies that the odds of the risk factor(s) in cases is higher than in controls, indicating a positive association.
3. An odds ratio of less than one signifies a protective (negative) association as the odds of exposure in cases is less than in controls.

**Assessing the Validity of the Results:** All analytic epidemiology studies require the researcher to evaluate how valid the results are. This requires consideration of chance, bias and confounding as possible alternative explanations.

**Assessing the Role of Chance:** In order to understand the extent to which the obtained odds ratio could have been due to chance, it is appropriate to calculate a confidence interval' around the odds ratio. The confidence interval is a range which contains the calculated odds ratio. For example, a 95% confidence interval around the odds ratio means that the researcher can be 95% sure that the true odds ratio lies within the range of values or conversely that there is only a 5% probability that this result could have been obtained by chance. **Understanding and Preventing Bias:** Bias refers to a 'systematic error' which affects study results. The researcher should ensure that there is as little bias as possible in conducting

the study. There are 4 important types of bias to consider in case-control studies:

**i. Selection Bias:** Those who agree to participate in a study are different from those who do not in ways possibly related to exposure and outcome. This can be overcome by selecting study participants by random sampling.

**ii. Observation Bias:** Knowledge of the disease status may influence how the data gatherer elicits and records information obtained from study subjects. This can be overcome by not disclosing the disease status of study subjects to the data gatherers.

**iii. Recall Bias:** Relates to differences in the ways exposure information is remembered by cases, who have experienced an adverse health outcome and by controls who have not. Choosing hospital controls or those related to or known to **the case may overcome this problem.**

**iv. Misclassification Bias:** Refers to errors in categorization of either exposure or disease status among cases and controls. This bias can be avoided by strict definitions for cases and controls and also for defining exposure and non-exposure. Assessing the Role of Confounding: A confounder is a factor associated with both the risk factor under study as well as the disease outcome. It could either mask an association or create one where none exists. For example, lung cancer is usually seen in the older age group. If one were to study the effect of the number of cigarettes smoked or the duration of smoking, age becomes a potential confounder, as older people are likely to have smoked for a longer duration. Confounding can be taken care of in the design phase or later in the analysis phase. In the design phase it can be dealt with by matching the cases and controls, e.g., age and sex (two common confounders). Confounding can also be dealt with in the analysis phase through stratification on the basis of the confounding factor or through more sophisticated analytical techniques. If designed and conducted well, a case-control study can be an effective means of identifying risk factors for a disease. Preventive strategies to deal with these risk factors can be designed to reduce the occurrence of the disease/outcome and thus the morbidity, mortality and financial burden associated with it.

## References

1. Hennekens, C.H. and Buring, J.E. Edited by S.L. Myrcent. Epidemiology in medicine. 1st ed. Little, Brown and Company. Boston, 1987.
2. Mausner, J.S. and Kramer, S. Epidemiology . an introductory text. 2nd ed. W.B. Saunders Company. Philadelphia, 1985.