

7096

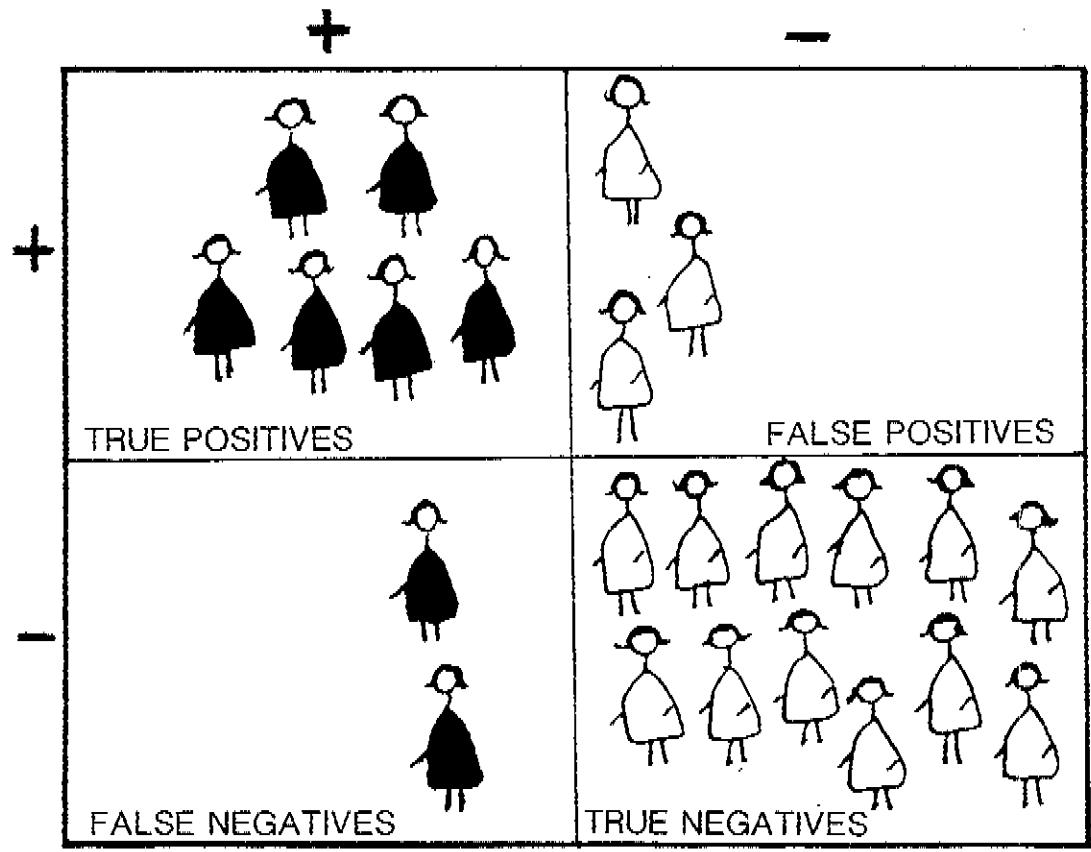
WHO/FHE/86.3



*not kept in service
since 1986
from file
file
based on*

**A WORKBOOK ON HOW TO PLAN AND CARRY OUT
RESEARCH ON THE RISK APPROACH
IN MATERNAL AND CHILD HEALTH
INCLUDING FAMILY PLANNING**

CASE-CONTROL STUDIES



981198 OMS



A WORKBOOK ON HOW TO PLAN AND CARRY OUT
RESEARCH ON THE RISK APPROACH
IN MATERNAL AND CHILD HEALTH
INCLUDING FAMILY PLANNING

* SECTION 5.10	CASE-CONTROL STUDIES
5.10.1	Structure of studies
5.10.2	Risk quantification
5.10.3	Types and sources of error
5.10.4	Conducting the study
5.10.5	Data analysis
5.10.6	Links with population surveys
Appendix	Estimation of the relative risk from the odds ratio for unwanted outcome
Annex 1	Exercises

*This module on case-control studies supplements the material contained in the document A Workbook on How to Plan and Carry Out Research on the Risk Approach in Maternal and Child Health including Family Planning (FHE/MCH/RA/84.1) which is available free of charge upon request to: Division of Family Health, World Health Organization, 1211 Geneva 27, Switzerland.

ACKNOWLEDGEMENTS

Financial support for the preparation of this document was provided by the United Nations Fund for Population Activities (UNFPA) (Project No. INT/83/P48).

The issue of this document does not constitute formal publication. It should not be reviewed, abstracted or quoted without the agreement of the World Health Organization. Authors alone are responsible for views expressed in signed articles.

Ce document ne constitue pas une publication. Il ne doit faire l'objet d'aucun compte rendu ou résumé ni d'aucune citation sans l'autorisation de l'Organisation mondiale de la Santé. Les opinions exprimées dans les articles signés n'engagent que leurs auteurs.

Summary

A case-control study compares risk factors for a selected number of cases with the unwanted outcome and a selected number of controls without the unwanted outcome. The study population is not, as a group, representative of the target population because of the way cases and controls are chosen. It is therefore not possible to calculate the incidence of the unwanted outcome.

Despite these problems, case-control studies are useful for the widespread application of the risk approach as they provide a quick, easy and low-cost method for conducting risk studies, especially when there is a severe lack of epidemiological data.

5.10.1 Structure of studies

The previous sections of this chapter* used a cohort study to compare individuals with the unwanted outcome to those with the wanted outcome.

If the unwanted outcome under investigation is fairly rare, we may end up with fewer cases with the unwanted outcome than we need for our analysis and too many cases with the wanted outcome. This problem may be overcome by treating those with wanted and unwanted outcomes as two separate populations and taking comparatively more of those with the unwanted outcome than with the wanted outcome.

Thus, we start the study by identifying two population groups: those with the unwanted outcome and those without the unwanted outcome. Then we determine what the exposure to the risk factor was. This type of study is called a case-control study. Those with the unwanted outcome are the cases whereas those with the wanted outcome are the controls. Alternatively, the study can be designed in such a way that those with the risk factor are considered as the cases and those without that risk factor as the controls. This alternative design is rarely used and will not be dealt with further in this section.

The price we pay for getting a large number of cases of the unwanted outcome is that, unlike cohort studies, we cannot deduce the incidence of the unwanted outcome and the prevalence of the risk factor in the target population and must therefore obtain that information from other sources.

Case-control studies using past records are known as retrospective case-control studies whereas prospective case-control studies use data collected as the study progresses. Case-control studies are usually retrospective in that the information on risk factors is obtained after the occurrence of the outcome in the individuals in the study population.

Advantages

The case-control methodology is particularly suited for conducting risk approach studies using unwanted outcomes that are rare. As a large number of cases with an unwanted outcome can be included in the study, valuable information is generated.

Case-control studies require relatively small samples of cases and controls compared to cohort studies where the low incidence of unwanted outcomes may mean that very large samples need to be assembled.

Case-control studies provide a simple, rapid, low-cost and effective method of testing hypotheses.

Disadvantages

The investigator does not have the opportunity to examine the whole population at risk as only selected groups of those with the unwanted outcome and without the unwanted outcome are available for study. Therefore, it is not possible to calculate directly the incidence of the unwanted outcome or the prevalence of risk factors in the whole population. Estimates of attributable risk have to be derived by indirect methods.

*This refers to Chapter 5 of A Workbook on How to Plan and Carry out Research on the Risk Approach in Maternal and Child Health including Family Planning (FHE/MCH/RA/84.1). All page numbers and sections mentioned in this module refer to this chapter.

Susceptibility to bias presents another major disadvantage of case-control studies: selection bias leads to a situation whereby cases and controls cannot be compared. The criteria for the selection of controls are often debatable.

Bias can also result when the information gathered from cases and controls is not fully comparable. Recall bias may occur when those having suffered an unwanted outcome recall preceding events in greater detail and/or with explanations to account for the occurrence of the unwanted outcome. Also, influenced by his knowledge of the outcome, the researcher may examine the medical records in a different way while looking for preceding events.

The delay between exposure to a risk factor and the implementation of the research study may make it difficult to obtain certain types of information: not only can people forget but records may get lost. The latter may occur selectively for cases and controls depending on the professional interest of the health workers in the unwanted outcome.

Research in developing countries

For research on the risk approach in MCH/FP, the case-control design may have advantages over prospective cohort designs in developing country settings where relatively few women receive antenatal care and where routine medical records may be deficient. A prospective cohort study of unwanted pregnancy outcomes which are relatively rare would require the identification of a large number of women during pregnancy and follow-up until delivery. Since only a small proportion of women may come for antenatal care, most of the population would be excluded from observation. Also, those women who receive antenatal care may be of higher socioeconomic status or be referred for complicated pregnancies, and thus would be unrepresentative of the general population of pregnant women. Furthermore, information on relevant risk factors may not be available if medical records contain insufficient data or are incomplete.

A case-control approach overcomes some of these problems since a moderate number of cases with the unwanted outcomes and controls can be identified after delivery, and information on the course of pregnancy or pre-pregnancy risk factors obtained from the mother at interview and from medical records. Thus the case-control approach is more economical and logistically simpler because of the smaller sample size required and the absence of prospective follow-up. The study population used may be more representative of the target population since selective antenatal care is avoided. Furthermore, a case-control study does not depend upon available records and can provide higher quality and more detailed information through standardized interviews or measurements. Although a case-control study cannot directly provide data on the incidence of the unwanted outcomes, the latter can be obtained from other sources.

5.10.2 Risk quantification

The main aim of case-control studies is to assess the importance of risk factors through a comparison of their prevalence among those with and without the unwanted outcome. If the prevalence of the risk factor is greater among cases than controls, an association between the risk factor and the unwanted outcome is likely. A similar prevalence of the risk factor among cases and controls suggests that there is no association. If the prevalence of the risk factor is less among cases than controls that presumptive risk factor is in fact likely to have a protective effect for that unwanted outcome.

The relation between risk factor and outcome in a risk study, whether of the cohort or case-control type, can be represented in tabular form as below:

Risk factor	Outcome		Total
	Unwanted	Wanted	
Present	a	b	a + b
Absent	c	d	c + d
All	a + c	b + d	a + b + c + d

In a cohort study, the incidence of the unwanted outcome in the population $(a + c)/(a + b + c + d)$ and the prevalence of the risk factor is $(a + b)/(a + b + c + d)$.

In a case-control study whereby cases and controls are specifically chosen according to the status of the outcome, the ratio between unwanted and wanted outcomes $(a + c)/(b + d)$ becomes distorted and not representative of the target population. Therefore, one can only speak of the prevalence of the risk factor in those with the unwanted outcome $(a/(a + c))$ or in those with the wanted outcome $(b/(b + d))$. This is because one has in fact chosen specified numbers of those with and those without the unwanted outcome while not knowing the ratio of the latter in the target population.

Using the method described in section 5.7, the relative risk can be calculated for a cohort study as follows:

$$\text{Relative risk} = \frac{\text{probability of an unwanted outcome in the presence of the risk factor}}{\text{probability of an unwanted outcome in the absence of the risk factor}}$$

$$= \frac{a/(a + b)}{c/(c + d)} = \frac{a(c + d)}{c(a + b)}$$

When the incidence of the unwanted outcome is low, a is negligible compared to b so that $(a + b)$ approximates to b and also c is negligible compared to d so that $(c + d)$ approximates to d.

Therefore, the above formula for the estimation of relative risk approximates to:

$$\text{Relative risk} = \frac{ad}{bc}$$

In a case-control study, the column for unwanted outcome in the above table represents the cases and the column for wanted outcome represents the controls.

The odds of the presence of the risk factor among the cases is a/c whereas the odds of the presence of the risk factor among the controls is b/d . The odds ratio of the presence of the risk factor among cases compared to controls is:

$$\text{Odds ratio} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

Hence, the relative risk approximates to the odds ratio in a case-control study.

The approximation between the relative risk and the odds ratio is reasonable, provided that the unwanted outcome is relatively uncommon - an incidence of less than 10% is usually acceptable. However, with more common unwanted outcomes, the odds ratio as calculated from a case-control study will overestimate the relative risk in the population. The extent of this overestimation can be calculated and the odds ratio can be corrected as shown in the Appendix. It is therefore important to note that the odds ratio from a case-control study is always a usable estimate of the relative risk in that population.

The odds ratio, like the relative risk, measures the strength of the association between the presence of the risk factor and the unwanted outcome. From the point of view of the community, the importance of a risk factor is largely dependent on its prevalence. The public health importance of the risk factor is measured by its attributable risk which indicates what might be expected to happen to the overall unwanted outcomes in the community if the risk factor were eliminated or its adverse effect prevented (section 5.7.4).

Attributable risk (%) can be calculated from the following equation:

$$\text{Attributable risk (\%)} = \frac{b(r - 1)}{b(r - 1) + 1} \times 100$$

where r = relative risk and

b = proportion of population with characteristic.

The relative risk (r) can be estimated from the odds ratio. The proportion (b) of the population with the risk factor cannot be calculated directly in case-control studies as the ratio of cases to controls in the study is not representative of the target population itself. However, the prevalence of the risk factor among the controls provides a fair estimate of the proportion of the population with the risk factor when the unwanted outcome is relatively uncommon. Alternatively, a population survey can be carried out solely to determine the prevalence of risk factors.

5.10.3 Types and sources of error

Types of error

We may erroneously conclude that there is an association between a risk factor and an unwanted outcome when, in reality, an association does not exist. This is called a Type I (or alpha) error. The likelihood of such an error occurring can be measured by tests of statistical significance. These tests measure the probability, alpha, that a result occurred by chance. Usually, one accepts a value of alpha of 5 per cent ($\alpha = 0.05$).

We may also erroneously conclude that there is no association between a risk factor and an unwanted outcome when, in reality, an association exists. This is called a Type II (or beta) error and can be measured by the probability (beta) of missing an association if one indeed exists. Usually, one accepts a value of beta of 10 per cent ($\beta = 0.10$).

These two types of error are discussed in section 5.4.3.

Sources of error

Case control studies are, like other epidemiological studies, subject to a variety of errors (see pages 119 and 140). However, case-control studies are particularly prone to systematic errors such as selection bias, misclassification bias and confounding bias.

Selection bias is minimized by trying to ensure that controls are representative of the population from which the cases with the unwanted outcome are drawn. Both cases and controls should therefore be selected from the same source over a similar period of time. However, cases with the unwanted outcomes are often selected from those admitted to hospitals and other health facilities and may not be representative of all cases of the unwanted outcome in the target population. For example, mothers giving birth in an urban hospital may predominantly come from urban areas or be of higher socioeconomic status and have greater access to antenatal care whereas rural women of lower socioeconomic status may be less likely to deliver in a hospital and have less access to antenatal care. If hospital cases were compared to controls drawn from the general population, bias could arise in a comparison of the use of antenatal care. To avoid selection bias, cases and controls must be chosen in such a way that cases with a presumptive risk factor (such as lack of antenatal care) have an equal chance of entering the study as do controls with that risk factor.

Misclassification bias occurs with the inaccurate determination of either presence of the risk factor or occurrence of the unwanted outcome. Misclassification bias may be divided into two distinct categories, differential and non-differential, depending on the existence of a relation between the error in one variable and the other variable - that is, unwanted outcome when there is an error in determining the presence of the risk factor and vice versa.

With non-differential misclassification bias, the error is not related to the other variable. For example, maternal age may be wrongly calculated using a calendar of local events that contains certain mistakes. Some mothers would therefore be assigned to the wrong age group. This misclassification would not be related to the occurrence of the unwanted outcome under study. As a result of a non-differential misclassification, the observed odds ratio would be an underestimate of the actual odds ratio.

With differential misclassification bias, the error is related to the other variable and the observed odds ratio can be either an overestimate or an underestimate depending on the direction of the misclassification. Interview bias and recall bias are the most common forms of misclassification bias.

Interview bias occurs when information is not obtained from cases and controls under similar conditions such as the location and the structure of the interview, as well as the manner in which the questions are asked and whether the subjects and interviewers are aware of the specific research hypotheses.

Recall bias occurs when the recollection of events by subjects is influenced by the occurrence of the unwanted outcome.

It is well recognised that there is a tendency for cases to associate unwanted outcomes with a variety of events which would therefore be overreported. When recall of a risk factor is better among those cases with the unwanted outcome than among controls, a spurious association arises.

Misclassification bias can be minimized by having standard questionnaires and interview techniques and, if possible, the interviewer should not know whether the mother is a case or a control. Also, subjects in the study and interview personnel should not be informed of the specific research hypotheses being tested. It is furthermore advisable to verify responses with information from other sources such as health record cards.

Confounding bias results from an indirect or non-causal association between a risk factor and an unwanted outcome. It is due to the effects of another variable which is independently related to both the risk factor and the unwanted outcome. Confounding is further discussed on page 161. Confounding bias may be dealt with in case-control studies by either matching cases with controls at the design stage of the study or by stratification and mathematical modelling at the analysis stage. However, mathematical modelling is beyond the scope of this workbook and will not be dealt with in any greater detail.

5.10.4 Conducting the study

The steps in conducting a case-control study closely follow those for a cohort study as detailed in Chapter 5. This section will only describe those aspects of the study which are specific to the case-control methodology.

Choice of cases and controls

The following basic principles govern the choice of cases with the unwanted outcome and controls:

- (a) Cases and controls should be representative of the same population over the same period of time.
- (b) The choice of subjects should not be based on the presence or absence of risk factors which are included in the research hypotheses as this would forfeit the subsequent analysis of these risk factors.
- (c) The cases should be representative of all the cases with the unwanted outcome and the controls should be representative of all those without the unwanted outcome in the target population.

The cases with the unwanted outcome can be selected from hospitals and health facilities, or from the community.

Cases with the unwanted outcome can be selected from a single hospital or group of health facilities. The advantage of this method is that the diagnosis is usually more accurate and it is easier to identify cases with the unwanted outcome in a clinical setting. The disadvantage is that these cases may not be representative of all the cases with the unwanted outcome in the community. Various factors such as hospital catchment area, admission policies, referral patterns and selective use of health services may affect the representativeness of cases chosen from hospitals and health facilities.

Cases with the unwanted outcome can be selected from all such cases in the community. These cases may be identified from readily available sources such as registers of vital statistics where all births and deaths are recorded. However, special surveys or arrangements may need to be made so as to detect all cases with that unwanted outcome in a

given area. Usually those cases occurring within a particular time span are chosen but sampling may be done if the number of eligible cases is greater than required. Even though case identification on a community basis is difficult, it has the great advantage of avoiding selection bias.

The choice of controls presents the most difficult aspect of conducting a case-control study. The controls should be those without the unwanted outcome and chosen from the same sources over the same time span as the cases. As the potential controls are more numerous than cases, some type of sampling procedure is used to choose a representative sample of controls (see page 117).

Controls chosen from a hospital are called hospital controls. Their advantage is that they are easier to identify and selective factors influencing the use of health services should apply equally to cases and controls allowing internally valid comparisons. However, the disadvantage of hospital controls is that they may not be representative of the community and this can seriously limit the generalization of the results.

Controls chosen from the community are called community controls. They can be chosen by taking a sample of the population of a geographical area (see page 116).

One or more controls may be selected for each case. A larger number of controls will increase the statistical significance of the results but will also increase the work and cost of the study. Thus, the decision on the ratio of cases to controls must be based on availability of resources.

The classical case-control study is set up in such a way that confounding variables are controlled by matching cases with controls at the design stage of the study. Controls are therefore chosen so that they differ from cases for those risk factors being investigated, known risk factors for the unwanted outcome being controlled for in the matching process.

The better the match between cases and controls, the more valid are the conclusions about the risk factors being investigated. This is usually crucial when one is testing the effect of a drug on the course of a disease. However, by matching for a particular risk factor, that difference between cases and controls is eliminated and it will not be possible to examine the importance of that risk factor at the data analysis stage.

In a risk study, one looks at many risk factors in order to find those that enable the prediction of likelihood of an unwanted outcome. If any of these risk factors are used for matching purposes, they are forfeited as possible risk factors before they have even been investigated. Therefore, matching is not advisable for most research studies on the risk approach.

Sample size

As explained in section 5.4, the estimation of the sample size for a study takes into account the prevalence of the risk factor, the magnitude of the relative risk we wish to detect, and the degree of confidence we wish to be able to place on the results.

Steps for determining the necessary sample size for a case-control study

The following method for the estimation of sample size is a simplified approximation suitable for most case-control studies.

- Step 1 Estimate the prevalence, p_1 , of the risk factor among controls.
 Step 2 Specify the level of the relative risk, r , that we would like to detect.
 Step 3 Calculate the prevalence, p_2 , of the risk factor among cases:

$$p_2 = \frac{p_1 r}{1 + p_1(r - 1)}$$

- Step 4 Calculate the average prevalence, p , of the risk factor among cases and controls:

$$p = \frac{1}{2}(p_1 + p_2)$$

- Step 5 The values of alpha and beta are then used through the standard normal deviates (Z-alpha and Z-beta as on page 129) to calculate the sample size N for cases and an equal number of controls:

$$N = \frac{2p(1 - p)(Z\text{-alpha} + Z\text{-beta})^2}{(p_2 - p_1)^2}$$

With alpha of 0.05, the corresponding value of Z-alpha is 1.96 and with beta of 0.10, the corresponding value of Z-beta is 1.28. Therefore the above formula for calculation of sample size simplifies to :

$$\begin{aligned} N &= \frac{2p(1 - p)(1.96 + 1.28)^2}{(p_2 - p_1)^2} \\ &= \frac{2p(1 - p)(3.24)^2}{(p_2 - p_1)^2} \\ &= \frac{21p(1 - p)}{(p_2 - p_1)^2} \end{aligned}$$

This simplified formula is conservative but adequate for most studies. For practical purposes, the prevalence of the risk factor among controls should be either estimated from available data or specified by assuming the lowest level of interest.

Example

- Step 1 We estimate the prevalence, p_1 , of the risk factor among controls to be 0.10.
- Step 2 We would like to detect a relative risk, r , of 2.0.
- Step 3 The prevalence, p_2 , of the risk factor among cases is:

$$\begin{aligned} p_2 &= \frac{p_1 r}{p_1(r-1) + 1} \\ &= \frac{0.10 \times 2}{0.10(2-1) + 1} \\ &= \frac{0.2}{0.1 + 1} \\ &= \frac{0.2}{1.1} \\ &= 0.18 \end{aligned}$$

- Step 4 The average prevalence, p , of the risk factor is:

$$\begin{aligned} p &= \frac{1}{2}(p_1 + p_2) \\ &= \frac{1}{2}(0.10 + 0.18) \\ &= \frac{0.28}{2} \\ &= 0.14 \end{aligned}$$

- Step 5 For alpha of 0.05 and beta of 0.10, the required sample size, N , for cases and an equal number of controls is:

$$\begin{aligned} N &= \frac{21p(1-p)}{(p_2 - p_1)^2} \\ &= \frac{21 \times 0.14 \times (1 - 0.14)}{(0.18 - 0.10)^2} \\ &= \frac{21 \times 0.14 \times 0.86}{0.08^2} \\ &= 395 \end{aligned}$$

Therefore, about 400 cases and 400 controls are necessary.

5.10.5 Data analysis

Data analysis for case-control studies should proceed from the simple to the complex, and from the general to the detailed. Each variable recorded can be treated as a dichotomy and the data should be transformed into one of two possible responses: presence or absence. The frequency of the responses can then be tabulated for each variable, and the odds ratio estimated.

The data in the following table will be analysed as an example of case-control studies using the methodology described in section 5.10.2.

Risk factor 1	Outcome		Total
	Unwanted	Wanted	
Present	110	20	130
Absent	190	280	470
All	300	300	600

As the above data was obtained by selecting 300 cases with the unwanted outcome and 300 controls from the target population, the relative risk cannot be calculated using the formula on page 157 although the structure of the table on that page and the above table are similar.

The odds ratio can be calculated using the formula in section 5.10.2:

$$\begin{aligned} \text{Odds ratio} &= \frac{110 \times 280}{20 \times 190} \\ &= 8.1 \end{aligned}$$

The proportion of the controls with the risk factor is 20/300 i.e. 6.7% or 0.067.

As the incidence of the unwanted outcome in the target population is relatively uncommon, it will be assumed that the odds ratio is an acceptable approximation for the relative risk and that the proportion of the controls with the risk factor is an acceptable approximation for the prevalence of the risk factor in the target population.

The attributable risk (%) can therefore be calculated:

$$\begin{aligned} \text{Attributable risk (\%)} &= \frac{b(r - 1)}{b(r - 1) + 1} \times 100 \\ &= \frac{0.067 (8.1 - 1)}{0.067 (8.1 - 1) + 1} \times 100 \\ &= 32 \end{aligned}$$

Significance testing

Significance tests are described in detail on page 158. A simple test for the odds ratio is a Chi-square with one degree of freedom, as shown below using the notation from the table in section 5.10.2.

$$\text{Chi-square} = \frac{(ad - bc - \frac{a+b+c+d}{2})^2 \times (a + b + c + d)}{(a + b)(c + d)(a + c)(b + d)}$$

Using the data from the above table:

$$\begin{aligned} \text{Chi-square} &= \frac{((110 \times 280) - (20 \times 190) - \frac{600}{2})^2 \times 600}{130 \times 470 \times 300 \times 300} \\ &= \frac{(30800 - 3800 - 300)^2 \times 600}{130 \times 470 \times 300 \times 300} = \frac{26700^2}{130 \times 470 \times 150} \\ &= 77.8, \text{ which is statistically significant.} \end{aligned}$$

Interdependence

The topic of interdependence was considered in section 5.7.2. Whenever data analysis leads to the suggestion that certain risk factors are associated with the unwanted outcome, one should consider the possibility of confounding bias. For example, if young maternal age and primiparity are both found to have increased odds ratios for perinatal mortality, confounding between maternal age and parity should be examined as younger mothers are more likely to be primiparous.

This can be done easily by stratifying the data for primipara and multipara into two separate tables and calculating the odd ratios associated with young maternal age. For example, consider the odds ratios for those analyses of perinatal mortality which are set out in the following table:

<u>Population</u>	<u>Risk factor</u>	<u>Odds ratio</u>
All	Young maternal age	8.1
All	Primipara	6.1
Primipara only	Young maternal age	4.8
Multipara only	Young maternal age	3.5

It is clear that maternal age is important regardless of parity but the risk associated with young maternal age is somewhat reduced once parity is controlled. Young primiparous women seem to be those at highest risk.

5.10.6 Links with population surveys

Case-control studies are often conducted in hospitals or other health facilities because this is usually the easiest method of identifying cases of the unwanted outcome and controls. However, such studies have two major limitations in that they cannot provide information on the population who fail to use the health services, and there may be selection bias among those who choose to deliver at home and never contact the formal health sector.

The only way one can obtain information on the population outside the health sector is by special surveys which use data collected by interview or other means to estimate the prevalence of risk factors in the target population. The prevalence of risk factors can be used to estimate attributable risk more accurately.

The prevalence of risk factors among the controls in the case-control study can be compared to the target population to ascertain the degree to which selective processes may have affected entry into the formal health services and, thus, into the case-control study. This not only allows a possible correction for selection bias but also a description of the characteristics of those choosing not to use health services for antenatal care or delivery and the estimation of attributable risk.

The design of population sample surveys was discussed in section 5.3 and only relevant aspects will be considered here.

Recall information is likely to be more reliable with recent events. Therefore, for studies relating to pregnancy, the study population could comprise those women who have had a live birth or stillbirth within the previous three years, a long enough period to capture a sufficient number of pregnancies.

Information to be collected during interviews with the women would include socioeconomic characteristics, reproductive history, use of health services (especially care during pregnancy and delivery), breastfeeding practices and use of contraception.

The required sample size can be estimated using the method described on page 128 as this survey is aimed at determining the prevalence of certain risk factors in the target population. The procedures for selecting a sample frame and taking a random sample are described in section 5.3.

The data analysis should, in the first place, provide an estimate of the prevalence of the risk factors and other characteristics such as use of antenatal care in the target population. This would enable the estimation of attributable risk for risk factors by using the odds ratio obtained from the case-control study and the prevalence from the population survey.

Cross tabulations could be done to identify the characteristics of women not covered by antenatal care. For example, if 30% of the target population is found not to be receiving antenatal care, a breakdown of the antenatal coverage by age group may help to identify those who do not use the services. The following table provides a hypothetical example. It has been constructed with an extreme age selection bias for illustrative purposes.

Use of antenatal care by maternal age

Maternal age in years	Numbers in survey population	Use of antenatal care		Percentage not covered by antenatal care
		Yes	No	
-19	100	35	65	65
20-	300	217	83	28
25-	300	196	104	35
30+	300	252	48	16
All	1000	700	300	30

It is clear that teenagers are particularly prone to lack of antenatal care.

Finally, the data from the prevalence survey can provide useful information on the extent to which the selection of subjects may have affected the case-control study. For example, the distribution of maternal age in the target population and among controls in the case-control study may be examined in the context of the above hypothetical example.

Distribution of maternal age for controls in case-control study and sample population by use of antenatal care

Maternal age in years	Sample from target population				Controls in case-control study	
	Covered by antenatal care		Not covered by antenatal care		Number	Percentage distribution
	Number	Percentage distribution	Number	Percentage distribution		
-19	35	5.0	65	21.7	26	7.5
20-	217	31.0	83	27.7	107	31.0
25-	196	28.0	104	34.6	89	26.0
30+	252	36.0	48	16.0	123	35.5
All	700	100.0	300	100.0	345	100.0

The distribution of maternal age is similar among antenatal care users in the target population and among controls in the case-control study. This shows that the controls were representative of antenatal care users. A comparison of the distribution of maternal age among users and non-users of antenatal care shows that there is a preponderance of teenagers among non-users, a finding compatible with the earlier conclusion that teenagers are prone to a lack of antenatal care.

In many settings, those not receiving antenatal care are often younger, primiparous and of lower socioeconomic status. This fact has to be taken into account in the analysis of case-control studies. The example of a predominance of mothers of higher socioeconomic status in a hospital setting as compared to the target population will be used to illustrate the two main ways in which selection can affect the results of a case-control study:

a. Cases and controls may be equally affected. This occurs commonly in studies based in hospitals and other health facilities, in view of the differential self-referral for care. It results in the mothers in hospitals and other health facilities being of higher socioeconomic status than the target population. Consequently, the odds ratios from a study conducted in hospitals and health facilities can be applied to the higher socioeconomic group in the target population but one would have less confidence in applying these odds ratios to the lower socioeconomic group in the target population. The extent of this selection can be ascertained by comparing the distribution of the relevant characteristic (socioeconomic status in our example) in those receiving and not receiving antenatal care.

b. Cases and controls may be unequally affected. This occurs when there are additional reasons for the differential use of services. For example, many mothers of lower socioeconomic status may receive antenatal care mainly when severe complications arise. These mothers represent a group with higher risk than other mothers of lower socioeconomic status in the target population. As a result, those of lower socioeconomic status would be overrepresented among the cases with the unwanted outcome and underrepresented among the controls. The odds ratio would therefore be a biased estimate of the actual value.

Appendix

Estimation of the relative risk from the odds ratio for unwanted outcome

As discussed in section 5.10.2, the odds ratio approximates to the relative risk as the incidence of the unwanted outcome becomes more and more rare. When the incidence of the unwanted outcome is comparatively high, the odds ratio overestimates the actual value of the relative risk. It is however possible, as shown in the following table, to correct the odds ratio by taking into account:

- the incidence of the unwanted outcome
- the prevalence of the risk factor.

For example, if the incidence of the unwanted outcome is 50 per 1000 and the prevalence of the risk factor is 25%, an odds ratio of 4.0 would be indicative of a relative risk of 3.7. As the incidence of the unwanted outcome is low, the magnitude of the odds ratio and the relative risk are of the same order.

However, if the incidence of the unwanted outcome is 250 per 1000 instead of 50 per 1000 with the prevalence of the risk factor and the value of the odds ratio being unchanged, the corresponding estimate of the relative risk would be 2.6.

In summary, even with common unwanted outcomes, a case-control study can provide a reasonable estimate of the relative risk.

Estimation of relative risk from odds ratio,
incidence of unwanted outcome and prevalence of risk factor in target population

Prevalence of the risk factor (%)	Incidence of the unwanted outcome per 1000									
	50					100				
	5	15	25	35	45	5	15	25	35	45
Odds ratio										
2	1.9	1.9	1.9	1.9	1.9	1.8	1.8	1.8	1.9	1.9
3	2.8	2.8	2.8	2.8	2.8	2.5	2.6	2.6	2.7	2.7
4	3.5	3.6	3.7	3.7	3.8	3.1	3.3	3.4	3.4	3.5
5	4.3	4.4	4.5	4.6	4.7	3.7	3.9	4.1	4.2	4.3
6	5.0	5.2	5.4	5.5	5.5	4.2	4.5	4.8	5.0	5.1
7	5.6	6.0	6.2	6.3	6.5	4.6	5.1	5.5	5.7	5.9
8	6.3	6.7	7.0	7.2	7.3	5.1	5.7	6.1	6.4	6.7
9	6.9	7.5	7.8	8.1	8.2	5.4	6.2	6.8	7.2	7.5
10	7.5	8.2	8.7	8.9	9.1	5.8	6.8	7.4	7.9	8.3

Prevalence of the risk factor (%)	Incidence of the unwanted outcome per 1000									
	150					200				
	5	15	25	35	45	5	15	25	35	45
Odds ratio										
2	1.7	1.8	1.8	1.8	1.8	1.7	1.7	1.7	1.7	1.7
3	2.3	2.4	2.5	2.5	2.5	2.2	2.2	2.3	2.4	2.4
4	2.8	3.0	3.1	3.2	3.3	2.6	2.7	2.8	2.9	3.0
5	3.2	3.5	3.7	3.8	4.0	2.9	3.1	3.3	3.5	3.7
6	3.6	3.9	4.2	4.5	4.7	3.2	3.5	3.8	4.0	4.3
7	3.9	4.4	4.8	5.1	5.4	3.4	3.8	4.2	4.6	4.9
8	4.2	4.8	5.3	5.7	6.1	3.6	4.1	4.6	5.1	5.5
9	4.4	5.2	5.8	6.3	6.7	3.7	4.4	5.0	5.6	6.0
10	4.6	5.5	6.3	7.0	7.4	3.9	4.6	5.4	6.1	6.6

		Incidence of the unwanted outcome per 1000									
		250					300				
Prevalence of the risk factor (%)		5	15	25	35	45	5	15	25	35	45
Odds ratio											
2		1.6	1.6	1.6	1.7	1.7	1.5	1.6	1.6	1.6	1.6
3		2.0	2.1	2.2	2.2	2.3	1.9	2.0	2.0	2.1	2.1
4		2.3	2.5	2.6	2.7	2.8	2.2	2.3	2.4	2.5	2.6
5		2.6	2.8	3.0	3.2	3.4	2.4	2.5	2.7	2.9	3.1
6		2.8	3.1	3.4	3.6	3.9	2.5	2.7	3.0	3.3	3.5
7		3.0	3.3	3.7	4.1	4.4	2.6	2.9	3.3	3.6	3.9
8		3.1	3.5	4.0	4.5	4.9	2.7	3.1	3.5	3.9	4.4
9		3.2	3.7	4.3	4.9	5.4	2.8	3.2	3.7	4.2	4.8
10		3.3	3.9	4.6	5.3	5.9	2.9	3.3	3.9	4.5	5.2

Exercises 5.10

- 5.10.1 It has been decided to carry out a survey in Mineralia in the Mountain region because risk factors for perinatal mortality in Mineralia could be different from those in the rest of Fictitia.
- Justify the choice of a case-control survey type.
- 5.10.2 Describe how you would choose cases and controls for that case-control study.
- 5.10.3 It was decided to carry out a case-control study in the hospital and health facilities of Mineralia as it was felt that almost all perinatal deaths in the community are reported in these institutions. The next birth in the same institution was chosen as a control.
- Of 300 cases of perinatal death, 115 were to mothers of 19 years old or less, as compared to 60 of the 300 controls.
- Construct a 2 by 2 table to present the results of that case-control study. Estimate the relative risk and attributable risk for maternal age of 19 years old or less for perinatal mortality in Mineralia.
- 5.10.4 It was later decided to choose controls from the community. Of the 300 controls chosen, 75 were mothers of 19 years old or less.
- Estimate the relative risk and attributable risk for maternal age of 19 years old or less for perinatal mortality in Mineralia.
- How and why do the results differ from those in exercise 5.10.3?