

LIBRARY
INTERNATIONAL REFERENCE CENTRE
FOR COMMUNITY WATER SUPPLY AND
SANITATION (IRC)



DIARRHOEAL DISEASES CONTROL PROGRAMME

Case-Control Studies of Childhood Diarrhoea:

II. Sample Size

by

S.N. Cousens, R.G. Feachem, B. Kirkwood
T.E. Mertens, P.G. Smith



WORLD HEALTH ORGANIZATION

This document is not issued to the general public, and all rights are reserved by the World Health Organization (WHO). The document may not be reviewed, abstracted, quoted, reproduced or translated, in part or in whole, without the prior written permission of WHO. No part of this document may be stored in a retrieval system or transmitted in any form or by any means - electronic, mechanical or other without the prior written permission of WHO.

The views expressed in documents by named authors are solely the responsibility of those authors.

Case-Control Studies of Childhood Diarrhoea:

II. Sample Size

by

S.N. Cousens¹, R.G. Feachem¹, B. Kirkwood¹
T.E. Mertens¹, P.G. Smith²

¹ Department of Tropical Hygiene, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

² Tropical Epidemiology Unit, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

ISBN 11511-2
245.11 88CA

Contents

	<u>Page</u>
PREFACE	3
ABSTRACT	3
1. INTRODUCTION	5
2. AN EXAMPLE OF A CASE-CONTROL STUDY	5
3. DETERMINANTS OF SAMPLE SIZE	7
3.1 Risk factor/intervention under study	7
3.2 Distribution of risk factor/intervention in the population	9
3.3 Power and significance of the study	9
3.4 Case:control ratio	11
4. CONFOUNDING AND SAMPLE SIZE	11
5. LOOKING FOR INTERACTIONS	12
6. MATCHING	13
7. CALCULATING A SAMPLE SIZE	13
8. DISCUSSION	15
ACKNOWLEDGEMENTS	16
REFERENCES	17
ANNEX 1. Statistical formulae	18
ANNEX 2. Sample size tables	20
ANNEX 3. Sample size formulae	23

PREFACE

This document is part of a series prepared for the Diarrhoeal Diseases Control Programme of the World Health Organization. The series is a response to an upsurge of interest in the application of the case-control method to the study of childhood diarrhoea. That interest has been stimulated by the realization that, under certain circumstances, the case-control method can be a relatively quick, inexpensive and reliable method for measuring the impact of diarrhoea control measures or for identifying and quantifying risk factors for diarrhoea.

Case-control studies can be complex in their design and analysis and it is not possible to prepare a manual that can be followed exactly in all circumstances. A considerable amount of epidemiological judgement and skill must be exercised. The aim of this series is to provide the investigator with a clear view of the most important problems in the design, analysis, and interpretation of case-control studies of childhood diarrhoea, and to provide practical suggestions for the resolution of those problems. For the trained and experienced epidemiologist, these documents provide specialized guidance on the application of case-control methods. For others, the series provides an awareness of the methodological issues involved and a familiarity with the language and concepts of case-control studies.

Diarrhoeal diseases remain one of the leading causes of morbidity and mortality among children in poor communities in all parts of the world. Epidemiological studies have already contributed to an understanding of the risk factors involved and to the design and evaluation of appropriate interventions. Continued work in diarrhoea epidemiology is essential to further refine these interventions and to maximize their impact on severe illness and death. The Diarrhoeal Diseases Control Programme of WHO supports a range of research projects in this field in many countries. Those seeking financial or technical support for their research, or wishing to contact others undertaking similar investigations, are invited to contact the above Programme.

ABSTRACT

In this paper we discuss the factors which need to be considered to determine the appropriate size for a case-control study. These include: the magnitude of the association which the investigator wishes to study, the proportion of the population exposed to the risk factor/intervention of interest, the probability level at which results will be deemed statistically significant, the power of the study to detect the association. The smaller the size of the association it is desired to detect, the larger the sample size will have to be. For example, to detect an odds ratio of 1.5 will require an approximately threefold greater sample size than to detect an odds ratio of 2.0. Studies in situations where the exposure of interest is either very common or very rare will also require large samples.

We discuss how the number of cases required may be reduced by increasing the number of controls recruited. Little is gained by recruiting more than 4 controls per case.

The effect on sample size requirements of the presence of confounding variables is considered. It is suggested that increasing sample size by 25% will cope with the effect of confounders.

Matching, and its implications for sample size, are discussed briefly. It is suggested that sample sizes for matched studies may be calculated using the formulae for unmatched studies.

Sample size tables are provided and the formulae from which they were derived are also given.

RESUME

Ce document examine les facteurs à prendre en compte lorsqu'il s'agit de décider de la taille appropriée d'une étude cas-témoin. Ces facteurs sont les suivants: ampleur de l'association que le chercheur souhaite étudier, la proportion de la population exposée au facteur de risque ou à l'intervention étudiée, le niveau de probabilité auquel les résultats seront jugés statistiquement significatifs et, enfin, le pouvoir de l'étude à déceler l'association. Plus l'association qu'on veut déceler est petite, plus l'échantillon devra être grand. Par exemple, pour déceler un risque relatif de 1,5, l'échantillon devra être trois fois plus grand environ que s'il s'agissait de déceler un risque relatif de 2,0. Dans les cas où l'exposition étudiée est très courante ou très rare, les échantillons devront également être importants.

Nous examinons comment il est possible de réduire le nombre des cas en augmentant le nombre des témoins recrutés. Il n'y a guère avantage à recruter plus de 4 témoins par cas.

L'effet de la présence de facteurs parasites sur les dimensions de l'échantillon est examiné. Pour neutraliser le risque d'erreur dû à des facteurs parasites, il est proposé d'augmenter de 25% la taille de l'échantillon.

L'appariement, et ses incidences sur la taille de l'échantillon, sont examinés brièvement. Pour calculer la taille des échantillons en vue d'études recourant à l'appariement, il est suggéré d'utiliser les mêmes formules que pour les études ne recourant pas à l'appariement.

Le document contient des tableaux où figurent diverses tailles d'échantillons et les formules utilisées pour les calculer.

1. INTRODUCTION

In 1985 the World Health Organization issued a document entitled "Measuring the impact of water supply and sanitation facilities on diarrhoea morbidity: prospects for case-control methods" (Briscoe et al., 1985). This document was one of the products of two scientific meetings held in Cox's Bazaar, Bangladesh, and Geneva, Switzerland, at which methodologies for measuring the impact of water supply and sanitation projects on health were discussed. In the document, case-control studies were put forward as an alternative to longitudinal studies whose use in this field had been discouraged by a report of an expert panel to the World Bank (IBRD, 1976).

The present series of papers considers the wider application of case-control methods to the study of various risk factors for, and interventions for the control of, childhood diarrhoea caused by enteric infections. The first paper in the series discussed ways in which bias may arise in case-control studies of childhood diarrhoea and proposed strategies for the minimization of bias (Cousens et al., 1988). This paper, the second of the series, focuses on the sample sizes required in case-control studies of childhood diarrhoea.

In this paper we avoid the use of complex algebraic expressions and present instead simple numerical examples wherever possible. Statistical formulae used in the presentation of the examples are provided in Annex 1. We begin by considering a hypothetical case-control study.

2. AN EXAMPLE OF A CASE-CONTROL STUDY

As an example we consider a case-control study designed to assess the association between the presence of domestic animals in the home and the risk of diarrhoea morbidity in children aged less than 5 years. The study is based on patients attending a single health facility. "Cases" are those children reporting to the clinic in whom diarrhoea caused by an enteric infection is diagnosed; "controls" are randomly selected from those children reporting to the clinic who are not suffering from diarrhoea. Information concerning the presence of domestic animals in the households of both cases and controls is collected. In their simplest form the results of the study may be presented in the form of a 2 x 2 table:

	Cases	Controls	
Animals present	10	4	14
Animals not present	30	36	66
Total	40	40	80

The measure of association used in the analysis of case-control studies is the odds ratio (OR). For the above table this is calculated as follows:

$$OR = \frac{10 \times 36}{4 \times 30} = 3.0$$

This result suggests that children who live in houses where domestic animals are present are approximately 3 times more likely to suffer an attack of diarrhoea than children in houses without animals.

To assist in interpreting the results we need to test the statistical significance of the association we have found in our sample. Is there really an underlying association between the presence of animals and risk of diarrhoea or could our result have been obtained

by chance? Even when studying a factor which is not associated with diarrhoea (i.e., true odds ratio = 1.0) it is unlikely that the estimate we obtain from a case-control study will be exactly equal to 1.0 due to sampling variations. How likely is it that our estimate of 3.0 has arisen in this way? A method of testing the significance of an association in a 2 x 2 table is to perform a chi-squared (X^2) test. From the table,

$$X^2 = \frac{80 (10 \times 36 - 4 \times 30) - 0.5 \times 80)^2}{40 \times 40 \times 66 \times 14}$$

$$= 2.16$$

Comparing this value against a table of values for a chi-squared distribution with 1 degree of freedom it may be seen that the probability of obtaining a similar or more extreme result purely by chance in a situation in which the true odds ratio equals 1 is greater than 0.1. Thus our result is not statistically significant at the 10% level of significance. In this particular example we have not found strong evidence of an association. There are two possible reasons for this:

- (1) no association exists between the presence of animals and the risk of diarrhoea,
- (2) an association does exist, but our study was too small to detect it (i.e., to find a statistically significant association).

If we had recruited 200 cases and controls instead of 40, and the distribution of animals in the households of cases and controls had remained the same, we would have obtained the following table of results:

	Cases	Controls	
Animals present	50	20	70
Animals not present	150	180	330
Total	200	200	400

Once again the estimate of the odds ratio is:

$$OR = \frac{50 \times 180}{20 \times 150} = 3.0$$

but the value of the chi-squared statistic is now:

$$X^2 = \frac{400 [| 50 \times 180 - 20 \times 150 | - 0.5 \times 400]^2}{200 \times 200 \times 70 \times 330}$$

$$= 14.56$$

The probability of obtaining a value of X^2 as large as this when the true odds ratio is 1.0 is less than 0.0001, i.e., there is less than one chance in 10 000 that sampling variation explains the association we have observed between risk of diarrhoea and the presence of animals. This would be regarded as very strong evidence that such an association does exist.

3. DETERMINANTS OF SAMPLE SIZE

When deciding on the sample size for a study we are faced with two conflicting interests. For a study to be worthwhile we must recruit sufficient subjects to have a "good" chance of detecting the association if it exists, while at the same time we wish to avoid wasting resources on an unnecessarily large study. A number of factors influence the size of sample required for a case-control study to have that "good" chance. These are discussed below.

3.1 Risk factor/intervention under study

The magnitude of the association in which we are interested will play a major role in determining our sample size requirements. A small association will be harder to detect than a large one and consequently will require a larger sample. To illustrate this point we consider two studies of similar size. The first is a study of the association between water supply and risk of diarrhoea morbidity in an area where 50% of the population are exposed and the true odds ratio is 1.5. The second is a study of the association between domestic hygiene and risk of diarrhoea morbidity in an area where, as in the first study, 50% of the population are exposed. The true odds ratio is 2.0. The results of the two studies are presented below.

Water supply

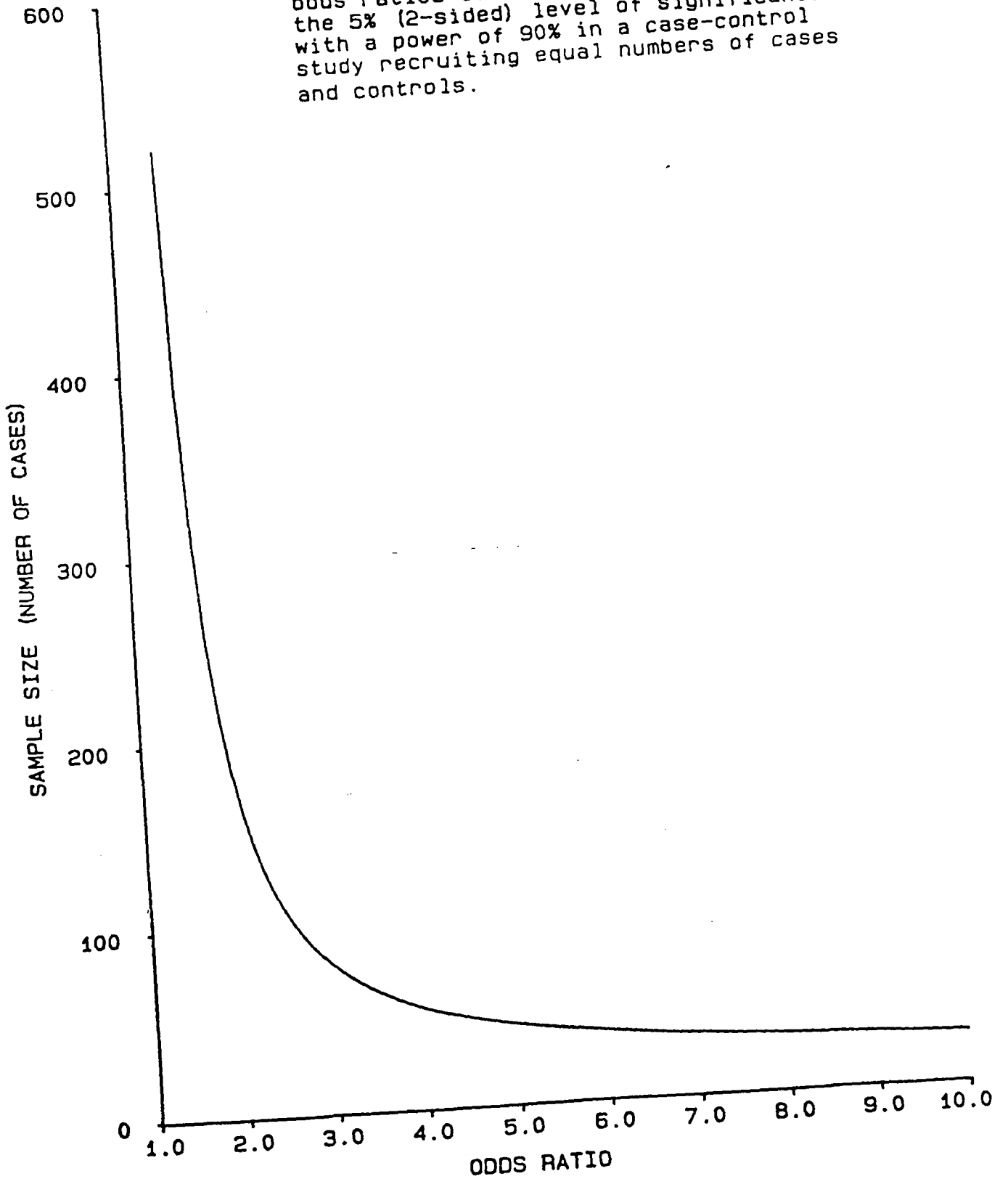
	Cases	Controls	
Unimproved supply	60	50	110
Improved supply	40	50	90
Total	100	100	200
			OR = 1.5
			$\chi^2 = 1.64, p = 0.2$

Domestic hygiene

	Cases	Controls	
Poor	67	50	117
Good	33	50	83
Total	100	100	200
			OR = 2.0
			$\chi^2 = 5.27, p < 0.025$

In the study of water supply the chi-squared statistic tells us that there is quite a high probability (0.2 or 20%) that the result we have obtained could be observed by chance even if there is no real underlying association. Such a result would not be regarded as strong evidence of an association between water supply and risk of diarrhoea. On the other hand, the association between domestic hygiene and diarrhoea morbidity is statistically significant, i.e., the probability that we obtained this result by chance, in the absence of any association, is small (less than 0.025 or 2.5%). These two studies differ only in the magnitude of the association observed, yet the inferences which may be drawn from each concerning the observed associations are different. Figure 1 shows graphically the relationship between the magnitude of the odds ratio and the sample size required to detect it. (The graph has been drawn under the assumptions that 50% of the population [i.e., the controls] are exposed, and that a 5%, 2-sided level of significance and a power of 90% are required [see section 3.3]). From Figure 1 it is evident that the sample size required by a study increases sharply as the value of the odds ratio to be detected falls below 2.5. A

FIGURE 1. The sample size required to detect odds ratios of different sizes at the 5% (2-sided) level of significance with a power of 90% in a case-control study recruiting equal numbers of cases and controls.



study designed to detect an odds ratio of 1.5 will require a sample about three times larger than a similar study designed to detect an odds ratio of 2.0.

3.2 Distribution of risk factor/intervention in the population

In the previous example a study of the association between domestic hygiene and risk of diarrhoea was conducted in an area where 50% of the population was exposed. Now consider a similar study conducted in an area where only 15% of the population are exposed.

Domestic hygiene

	Cases	Controls	
Poor	26	15	41
Good	74	85	159
Total	100	100	200

OR = 2.0
 $\chi^2 = 3.07, p > 0.05$

Once again the odds ratio is 2.0 but the statistical significance of the study's results is different. The probability of obtaining this result by chance if the true odds ratio is one is greater than 0.05 (5%). This would not usually be regarded as strong evidence of an association. The only way in which this example differs from the previous one is in the proportion of the population exposed. Figure 2 shows the relationship between the percentage of the population exposed and the sample size required (assuming an odds ratio of 2.0 is to be detected, at the 5% 2-sided level of significance with a power of 90%). Studies in areas where roughly half the population are exposed will require smaller samples than similar studies in areas where exposure is either very common or very rare. This point is very important and must be remembered when selecting a study site. In areas where the proportion of the population exposed is less than 10% or more than 90%, sample size requirements may become prohibitive. Clearly, when computing sample sizes we shall require an estimate of the proportion of the population exposed to the risk factor of interest.

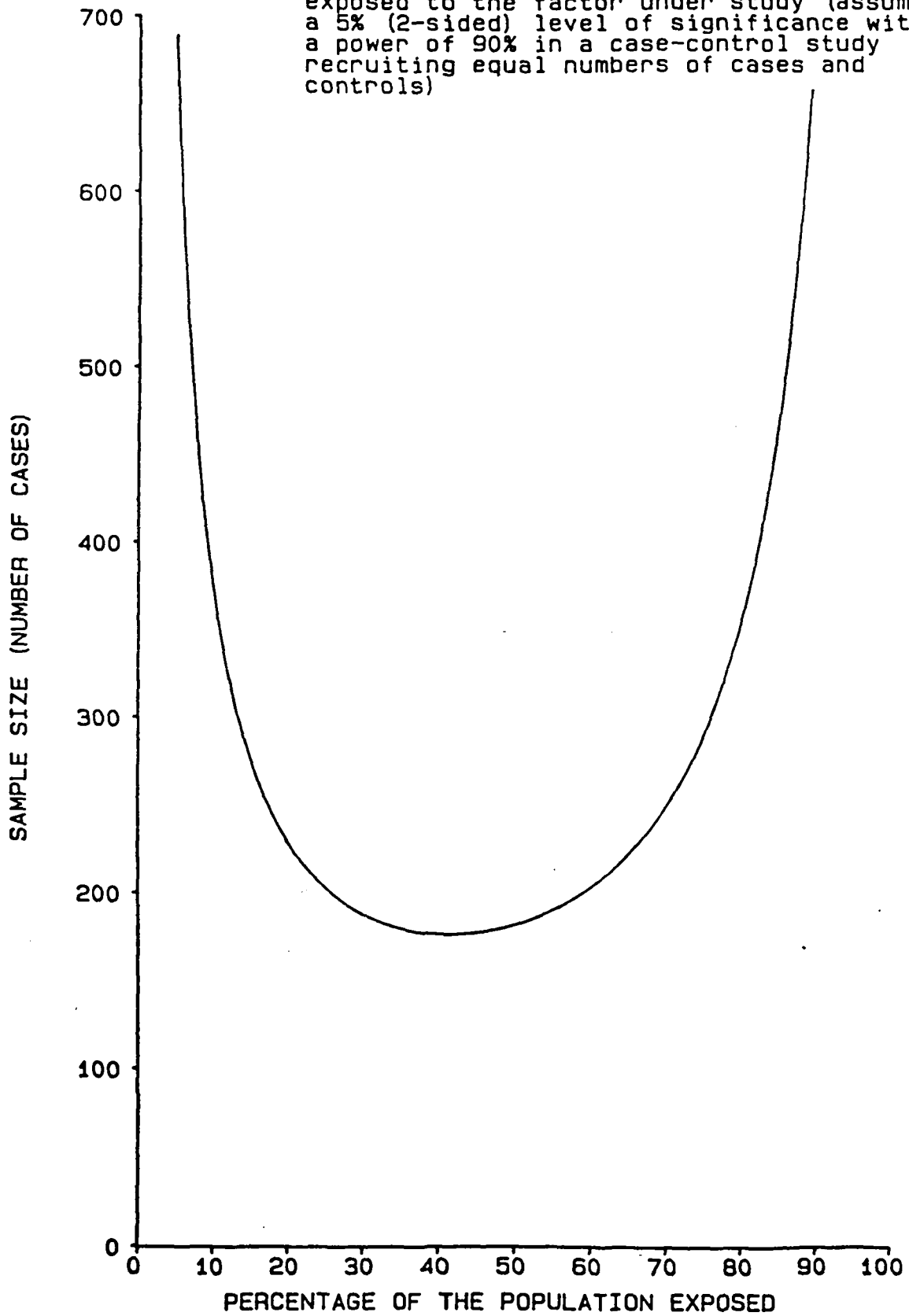
3.3 Power and statistical significance of the study

The terms "power" and "statistical significance" have precise and closely associated meanings when applied in a statistical context.

"Significance" has already been mentioned and refers to the strength of the evidence we require about an association, i.e., how sure we wish to be that an observed association is due to an underlying association rather than chance. An association which is observed to be significant at the 5% level is one which has a probability of less than 0.05 (1 in 20) of arising if there really is no true association. Similarly, an association significant at the 1% level has a probability of less than 0.01 (1 in 100) of arising if there really is no true association. The level of significance indicates, therefore, the probability of a false positive result. Significance at the 1% level is stronger evidence of an association than significance at the 5% level. Results that are significant at the 5% level are usually regarded as providing some evidence of an association, while results significant at the 1% level are regarded as strong evidence, and results significant at the 0.1% level are regarded as very strong evidence.

"Power" measures the probability that a study will detect an association of a given size at a given level of significance, i.e., how "good" a study is at detecting an association. A study with a power of 90% will detect the specified association 90 times out of 100 (when the association exists). The power is thus 1 minus the probability of a false

FIGURE 2. The sample size required to detect an odds ratio of 2.0 in situations where different proportions of the population are exposed to the factor under study (assuming a 5% (2-sided) level of significance with a power of 90% in a case-control study recruiting equal numbers of cases and controls)



negative result. There is little point in conducting a study which has only a small chance of detecting the effect of interest and we would recommend that studies be designed to have a power of at least 80%.

Power and significance are interrelated. We can increase the power of a study simply by accepting a reduced level of significance. Similarly, we can increase the level of significance if we are prepared to accept a reduction in power. However, if we wish to increase the power of a study without reducing the level of significance (or vice versa), we shall have to increase the size of our sample. A common choice is a 5% level of significance with a power of 90%.

3.4 Case:control ratio

Each of the examples considered so far has featured a study in which equal numbers of cases and controls have been recruited. In some situations, for example when the study has to be completed in a short period of time, the investigator may find it difficult to recruit the required number of cases. To overcome this problem the investigator may reduce the number of cases required by recruiting a larger number of controls. By recruiting C times as many controls as cases the investigator may reduce the number of cases required from N to $(C + 1)N/2C$. This strategy may reduce the number of cases needed by up to 50%, but no more. In practice, there is little to be gained by recruiting more than 4 controls per case, since the number of cases required will not be much further reduced, and the number of controls required will become very large. For example, a study with 100 cases and 100 controls has the same "power" as a study with 75 cases and 150 controls, or one with 63 cases and 252 controls, or one with 55 cases and 550 controls. Note that as the number of cases recruited decreases the total sample size (cases + controls) increases.

4. CONFOUNDING AND SAMPLE SIZE

A confounding variable is one which is both a risk factor for the disease under study (paediatric diarrhoea) and is associated with the risk factor of interest. In analysing the data from a case-control study it is essential to control for confounding variables. Failure to do so will result in biased estimates of the odds ratio (see, for example, Schlesselman, 1982; Cousens et al., 1988). Previously we looked at an example of a study of domestic hygiene, recruiting 100 cases and 100 controls in an area where 50% of the population was exposed. The odds ratio was 2.0 and the chi-squared (X^2) statistic took a value of 5.27 ($p < 0.025$). This example was constructed assuming no confounding variables to be present. Now consider a similar study, with 100 cases and 100 controls in an area where 50% of the population are exposed, conducted in the presence of a confounding variable, socioeconomic status (SES). A simple, unstratified analysis of the data produces the following 2 x 2 table:

	Cases	Controls	
Poor hygiene	74	50	124
Good hygiene	26	50	76
Total	100	100	200
			OR = 2.8
			$X^2 = 11.2$

Stratifying on socioeconomic status produces the following 2 tables:

	Low SES			High SES		
	Cases	Controls		Cases	Controls	
Poor hygiene	64	40	104	10	10	20
Good hygiene	4	5	9	22	45	67
Total	68	45	113	32	55	87

Mantel-Haenszel odds ratio = 2.0

Mantel-Haenszel X^2 = 2.25, $p > 0.1$

In this example, households of low socioeconomic status are more likely to have poor domestic hygiene than households of high socioeconomic status (89% of controls compared with 18%), and children from low SES households are more likely to suffer from diarrhoea than children from high SES households (68% of cases come from households with low socioeconomic status compared with 45% of controls). The summary (Mantel-Haenszel) estimate of the odds ratio equals 2.0 (the true value), compared with the biased estimate of 2.8 obtained from the unstratified analysis. The Mantel-Haenszel X^2 statistic equals 2.25 (compared with a value of 5.27 in the unconfounded situation), indicating that the probability of observing such an association when there is, in reality, no underlying association, is greater than 0.1. The results of this study would not, therefore, be regarded as providing any real evidence of an association between poor hygiene practices and risk of diarrhoea. The result of the study conducted in the unconfounded situation ($p < 0.025$) would, on the other hand, be regarded as providing some evidence of an association. This example shows how the presence of confounding variables may reduce a study's ability (power) to detect the underlying association. This effect should be considered when choosing a sample size.

Smith and Day (1984) have looked at the quantitative effect of a single, dichotomous confounder on sample size. They considered various levels of confounding and concluded that a weak confounder (i.e., one not strongly associated with the disease or risk factor of interest) would require an increase of about 15% in the sample size for the power of the study to be maintained. When a strongly confounding variable is present, a much greater increase (of up to 70%) may be needed. In studies of childhood diarrhoea risk factors may be associated with each other. However, the odds ratios typically associated with these factors tend to be small (< 3.0), and therefore these factors do not constitute strong confounders as defined by Smith and Day (1984). We suggest increasing sample sizes by 25% to allow for the effects of confounding variables. This is based on the figure of 15% quoted by Smith and Day (1984), but makes some allowance for the fact that in practice confounding will not be restricted to a single dichotomous confounding variable.

5. LOOKING FOR INTERACTIONS

Consider the following results from a study of water supply and sanitation:

	Sanitation					
	No			Yes		
	Cases	Controls		Cases	Controls	
Unimproved supply	72	42	114	54	55	109
Improved supply	57	53	110	17	50	67
Total	129	95	224	71	105	176

OR = 1.59

OR = 2.89

The estimated odds ratios in the two strata are different. This suggests that there may be an interaction between the effects of water supply and sanitation on the risk of childhood diarrhoea; i.e., the effectiveness of the improved water supply in reducing the risk of diarrhoea morbidity depends on whether or not adequate sanitation facilities are available and used. It is possible to test the statistical significance of the interaction we have observed - i.e., is there an underlying difference between the odds ratios of the two strata, or could the difference we have observed have arisen due to sampling variation? The power of such a test will clearly depend on the size of the study. Smith and Day (1984) have also investigated the sample sizes required in studies designed to test for interactions. Their conclusion, that such studies will, in general, require sample sizes at least 4 times as large as studies of simple effects, suggests that it will rarely be possible to design case-control studies of interaction effects and childhood diarrhoea. This does not mean that such effects should not be looked for and tested during the analysis, but that the power of these studies to demonstrate that interactions are statistically significant will be limited.

6. MATCHING

Matching is a strategy used to reduce the loss in power resulting from the presence of confounding variables (see section 4). It will be discussed in detail in a later paper in this series. Matching is performed by selecting controls who are similar to cases with respect to the factor(s) being matched on. For example, in a study matching on age, recruitment of a 2-year-old case will lead to the recruitment of a 2-year-old control. This ensures that when the data are analysed and age is used to stratify the data, there will be similar numbers of cases and controls in each stratum. In theory, the sample size required by a matched design should be computed using a formula different to that for unmatched designs. Sample size calculations are, however, approximate and should be treated as such. In most circumstances the use of these special formulae for matched designs produces samples of similar, though not identical, size to the calculations described below. Given the approximate nature of all these calculations we consider that it is unnecessary to use a different formula for matched designs and advocate instead the method outlined below. In case-control studies of risk factors/interventions for childhood diarrhoea there are likely to be many potential confounding variables. It will rarely be feasible to match on all these variables and we suggest, therefore, that the 25% allowance for confounding, proposed in section 4, be retained even when some degree of matching takes place.

7. CALCULATING A SAMPLE SIZE

We have seen that a number of considerations influence the required sample size of a study. To calculate this sample size the following must be specified:

- (1) the odds ratio that the study is being designed to detect (R),

- (2) the proportion of the population exposed to the risk factor of interest (P),
- (3) the level at which results will be deemed statistically significant,
- (4) the required power of the study.

We present two methods for calculating sample sizes: using the tables provided in Annex 2, and using the formulae provided in Annex 3. The use of the tables in Annex 2 is illustrated in the three examples below. An example of the use of the formulae is given in Annex 3.

Example 1

A study of the effect of an improved water supply on risk of diarrhoea is being designed. The investigator wishes to be able to detect an odds ratio of 1.5 (if it exists) with a power of 90% at the (2-sided) 5% level of significance and it has been decided to recruit one control for each case. About 30% of the population in the study area are thought to use the improved supply. Then,

$$R = 1.5, \quad P = 0.7, \quad \text{significance} = 5\%, \quad \text{and} \quad \text{power} = 90\%$$

Note that $P = 0.7$ rather than 0.3 because P is the proportion of the population thought to be exposed to increased risk, i.e., the proportion of the population using an unimproved supply. To calculate the sample size for a study with a power of 90% at the 5% level of significance we must use Table A2.4. From this table it can be seen that a sample of 869 cases is required to detect an odds ratio of 1.5 in a population where 70% are exposed. This is the number of cases required assuming that an equal number of controls are to be recruited. It includes an allowance of 25% for confounding. Since it is intended to recruit 1 control per case no further adjustments are necessary and the final sample size is 869 cases and 869 controls. A large sample is needed in this example because we are trying to detect a small association (odds ratio = 1.5).

Example 2

In this example a case-control study of the association between vitamin A deficiency and risk of diarrhoea morbidity is being designed. The investigators are interested in detecting an odds ratio of 3.0 (or more) with a power of 90% at the (2-sided) 10% rather than 5% level of significance. Four controls are to be recruited for each case and it is thought that about 10% of the population suffer from low vitamin A levels.

$$R = 3.0, \quad P = 0.1, \quad \text{significance} = 10\%, \quad \text{and} \quad \text{power} = 90\%$$

For this calculation Table A2.3 must be used. From this table a figure of 145 is obtained. Remember that this is the number of cases required by a study recruiting equal numbers of cases and controls. In this example, however, the investigator is to recruit 4 controls per case. Table A3.3 is used to adjust the sample size accordingly. To calculate the number of cases required, multiply the number obtained from the table in Annex 2 (N) by the appropriate adjustment factor from Table A3.3 (F).

$$\begin{aligned}
 \text{i.e., number of cases required} &= N \times F \\
 &= 145 \times 0.625 \\
 &= 91
 \end{aligned}$$

Four controls are to be recruited per case = $4 \times 91 = 364$. Thus our final sample size is 91 cases and 364 controls, a total of 455 recruits. Note that this is larger than the total sample size for a study recruiting 1 control per case (= $145 + 145 = 290$). The reduction in the number of cases required (from 145 to 91) has been achieved at the cost of an overall increase in sample size (from 290 to 455).

Example 3

As a final example consider a study of the impact of a hygiene education programme on the risk of diarrhoea morbidity. The investigator wishes to detect an odds ratio of 2.0 with a power of 95% at the 5% level of significance. About 40% of the population have been exposed to the education programme. The investigator is constrained by the need to complete the study within 6 months, during which time it is expected that about 200 cases of diarrhoea will be seen at the clinic.

$$R = 2.0, \quad P = 0.6, \quad \text{significance} = 5\%, \quad \text{and power} = 95\%$$

Note $P = 0.6$ since P is the proportion of the population in the higher risk group, i.e., those who have not received any hygiene education.

For a study with power 95% at the 5% level of significance, Table A2.6 must be used. From this table it can be seen that about 334 cases would be required by a study recruiting 1 control per case. The investigator knows that only about 200 cases can be recruited, about 60% (or 0.6) of the number given by the table. Looking at Table A3.3, it can be seen that the use of 4 controls per case will reduce the number of cases needed to about 63% of the original number. The use of 5 controls per case would reduce the number of cases to 60% of the original number. Thus a sample of 200 cases and $5 \times 200 = 1000$ controls will give the study the desired power and significance. The use of 4 controls per case would give almost the same power, but with a considerably reduced sample size (1000 instead of 1200). This small loss of power might be worth accepting in return for the expected reduction in the cost of the study.

8. DISCUSSION

Various elements of a study design need to be considered when deciding on sample size in a case-control study. Some of these elements may be varied by the investigator. The power of the study to detect an association, the probability level at which results will be deemed statistically significant, and the magnitude of association which the study is designed to detect are all set by the investigator. There is, however, little point in conducting a study which has only a small chance of detecting the effect of interest, or accepting a level of significance that will convince no-one apart from the investigator. The magnitude of association that it will be of interest to detect should depend on the feasibility and cost of intervening on the factor under study. If it is cheap and easy to intervene on a factor we shall be interested in a relatively small association between that factor and disease (diarrhoea). The more expensive and difficult an intervention becomes, the larger the association must be to make the intervention worthwhile.

Other factors affecting the choice of sample size are more difficult for the investigator to control. The distribution among the population of the risk factor of interest may greatly affect the sample size needed. In areas where exposure is either very rare, or very common, a case-control study may not be practicable. The distribution of exposure should always be taken into account when selecting the site for a study. Confounding (associations between different risk factors) may also affect sample size requirements. The investigator may reduce the effect of confounding to some extent by matching on confounding variables. Matching will be discussed in a later paper in this series. While it is not possible to give a hard and fast rule concerning confounding, we recommend increasing sample size by about 25% to allow for its effects. It will not, in general, be feasible to design studies to detect interactions.

In situations in which there are constraints (e.g., time) on the number of cases that can be recruited, the number of cases required by a study may be reduced by increasing the number of controls recruited.

Formulae for the calculation of sample sizes for matched designs are not included here. Given the approximate nature of all sample size calculations and the similarity of the results obtained using the different formulae, we consider that it is unnecessary to use special formulae to calculate sample sizes for matched designs.

ACKNOWLEDGEMENTS

The Diarrhoeal Diseases Control Programme of the World Health Organization provided financial support for the preparation of this document. The authors would like to thank the following people for their constructive comments on earlier drafts of this document:
R.E. Black, J.D. Clemens, I. de Zoysa, M.H. Merson, N.F. Pierce, P. Sandiford.

REFERENCES

Briscoe, J., Feachem, R.G. and Rahaman, M.M. (1985) Measuring the impact of water supply and sanitation facilities: prospects for case-control methods. WHO unpublished document WHO/CWS/85.3.

Cousens, S.N., Feachem, R.G., Kirkwood, B.R., Mertens, T.E. and Smith, P.G. (1988) Case-control studies of childhood diarrhoea: I. Minimising bias. WHO unpublished document CDD/EDP/88.2.

IBRD (1976) Measurement of the health benefits of investments in water supply. Report of an Expert Panel, Public Utilities Department Report No. PUN 20. Washington D.C., World Bank.

Schlesselman, J.J. (1982) Case-control studies. Oxford, Oxford University Press.

Smith, P.G. and Day, N.E. (1984) The design of case-control studies: the influence of confounding and interaction effects. International Journal of Epidemiology, 13: 356-365.

ANNEX 1

STATISTICAL FORMULAE

1. Analysis of a single 2 x 2 table

	Case	Control	
Exposed	a	b	r1
Unexposed	c	d	r2
	m1	m2	n

$$\text{Odds ratio} = \frac{a \times d}{b \times c}$$

$$X^2 = \frac{n \times [|axd - bxc| - 0.5xn]^2}{m1 \times m2 \times r1 \times r2}$$

The statistical significance of the observed association is found by comparing the value of X^2 with the percentage points of the chi-squared distribution with 1 degree of freedom. If X^2 is greater than 3.84 then the association is significant at the 5% level; if X^2 is greater than 6.63 then the association is significant at the 1% level.

2. Stratified analysis

The data have been divided into several strata, each of which may be represented in the form of a 2 x 2 table. The (i) indicates that this table represents the ith strata.

	Case	Control	
Exposed	a(i)	b(i)	r1(i)
Unexposed	c(i)	d(i)	r2(i)
	m1(i)	m2(i)	n(i)

$$\text{Mantel-Haenszel OR} = \frac{\frac{a(1)xd(1)}{n(1)} + \frac{a(2)xd(2)}{n(2)} + \dots}{\frac{b(1)xc(1)}{n(1)} + \frac{b(2)xc(2)}{n(2)} + \dots}$$

$$\text{Mantel-Haenszel } X^2 = \frac{N}{D}$$

where

$$N = \left[\frac{a(1)xd(1) - b(1)xc(1)}{n(1)} + \frac{a(2)xd(2) - b(2)xc(2)}{n(2)} + \dots \right]^2$$

and

$$D = \frac{m1(1) \times m2(1) \times r1(1) \times r2(1)}{n(1) \times n(1) \times [n(1)-1]} + \frac{m1(2) \times m2(2) \times r1(2) \times r2(2)}{n(2) \times n(2) \times [n(2)-1]} + \dots$$

Annex 1

The statistical significance of the observed overall association, as estimated by the Mantel-Haenszel odds ratio, is found by comparing the value of the Mantel-Haenszel X^2 statistic with the percentage points of the chi-squared distribution with 1 degree of freedom. If X^2 is greater than 3.84 then the association is significant at the 5% level; if X^2 is greater than 6.63 then the association is significant at the 1% level.

ANNEX 2

SAMPLE SIZE TABLES

The following tables give the number of cases required by studies recruiting 1 control per case. The numbers include an allowance of 25% for confounding.

Table A2.1 Significance = 10% (2-sided), Power = 80%

	Odds ratio						
	1.5	2.0	2.5	3.0	4.0	5.0	10.0
.10	982	306	164	109	64	46	21
.20	577	186	103	70	43	32	16
.30	459	153	87	61	39	29	16
.40	419	144	84	60	39	31	18
.50	419	148	88	64	43	34	22
.60	454	165	100	74	51	41	27
.70	539	201	125	93	66	54	36
.80	736	281	177	134	96	79	55
.90	1358	531	339	259	189	158	112

Table A2.2 Significance = 5% (2-sided), Power = 80%

	Odds Ratio						
	1.5	2.0	2.5	3.0	4.0	5.0	10.0
.10	1222	381	204	135	80	57	25
.20	718	232	128	87	54	40	20
.30	571	190	108	75	48	37	20
.40	521	179	104	74	49	38	23
.50	521	184	110	79	54	43	27
.60	565	205	125	92	63	51	33
.70	671	251	155	115	81	67	45
.80	915	350	220	166	119	99	68
.90	1689	661	422	322	235	197	140

Table A2.3 Significance = 10% (2-sided), Power = 90%

	Odds ratio						
	1.5	2.0	2.5	3.0	4.0	5.0	10.0
.10	1307	407	218	145	85	60	27
.20	768	248	137	93	57	42	22
.30	611	204	116	81	51	39	22
.40	557	192	112	79	52	41	24
.50	557	197	117	85	57	46	28
.60	604	220	133	98	68	55	36
.70	718	268	166	123	87	71	48
.80	979	375	235	178	128	106	73
.90	1807	707	452	345	252	210	149

Table A2.4 Significance = 5% (2-sided), Power = 90%

	Odds ratio						
	1.5	2.0	2.5	3.0	4.0	5.0	10.0
.10	1582	493	264	175	103	73	33
.20	930	300	166	113	69	51	26
.30	739	246	140	97	62	47	26
.40	674	232	135	96	63	49	29
.50	674	239	142	103	69	55	34
.60	731	266	161	118	82	66	43
.70	869	324	200	149	105	86	58
.80	1185	453	285	215	154	128	88
.90	2187	856	547	417	305	255	181

Annex 2

Table A2.5 Significance = 10% (2-sided), Power = 95%

	Odds ratio						
	1.5	2.0	2.5	3.0	4.0	5.0	10.0
.10	1679	523	280	186	109	78	35
.20	987	318	176	120	74	54	28
.30	785	262	148	103	66	50	28
.40	716	246	143	102	67	52	31
.50	716	253	151	109	74	58	36
.60	776	282	171	126	87	70	46
.70	922	344	213	158	112	91	61
.80	1258	481	302	228	164	136	94
.90	2321	908	580	443	323	270	192

Table A2.6 Significance = 5% (2-sided), Power = 95%

	Odds ratio						
	1.5	2.0	2.5	3.0	4.0	5.0	10.0
.10	1988	619	332	220	129	92	41
.20	1168	377	208	142	87	64	33
.30	929		176	122	78	59	33
.40	848		169	120	79	61	36
.50	848	300	178	129	87	69	43
.60	910	304	203	149	103	83	54
.70	109	368	252	188	132	108	73
.80	1688	559	358	270	194	160	111
.90	2321	908	687	524	383	320	227

SAMPLE SIZE FORMULAE

(see, for example, Schlesselman [1982])

Four factors must be specified in order to calculate a sample size:

1. the odds ratio the study is being designed to detect (R),
2. the proportion of the population exposed to the risk factor of interest (P),
3. the level at which results will be deemed significant,
4. the power of the study.

Having specified the power and level of significance required, the following tables are used to ascertain two numbers which will be called S and T.

Table A3.1 Values of S corresponding to different levels of significance

	Level of significance	10%	5%	1%
S		1.7	2.0	2.6

Table A3.2 Values of T corresponding to different powers

	Power	80%	90%	95%
T		0.9	1.3	1.7

These four numbers, R, P, S and T, are used in the following formulae:

$$A = \frac{P \times R}{1 + P(R - 1.0)}$$

$$C = \frac{A + P}{2}$$

$$D = 1 - C$$

$$N = \frac{2 \times C \times D \times (S + T)^2}{(P - A) \times (P - A)}$$

N is the number of cases required in a study recruiting equal numbers of cases and controls. It does not include any allowance for confounding. To allow for confounding, multiply N by 1.25. To adjust for studies recruiting more than one control per case use Table A3.3 below as described in Example 2 in Section 7.

Example

To illustrate the use of the formulae we consider Example 1 from Section 7. We have R = 1.5, P = 0.7, significance = 5%, and power = 90%. From Table A3.1 we see that S = 2.0. From Table A3.2, T = 1.3. Then,

Annex 3

$$\begin{aligned} A &= \frac{0.7 \times 1.5}{1 + 0.7 \times (1.5 - 1.0)} \\ &= \frac{1.05}{1.35} \\ C &= \frac{1}{2} \times \frac{1.05}{1.35} + 0.7 \\ &= \frac{1.995}{2.7} \\ D &= 1 - \frac{1.995}{2.7} = \frac{0.705}{2.7} \\ N &= \frac{2 \times 1.995 \times 0.705 \times (2.0 + 1.3)^2}{2.7 \times 2.7 \times \left(0.7 - \frac{1.05}{1.35}\right) \times \left(0.7 - \frac{1.05}{1.35}\right)} \\ &= \frac{2 \times 1.995 \times 0.705 \times 3.3^2 \times 1.35^3}{2.7 \times 2.7 \times 0.105 \times 0.105} \\ &= 695 \end{aligned}$$

These formulae make no adjustment for confounding. To allow for this we increase the figure of 695 by 25%.

$$\text{Sample size} = 695 \times 1.25 = 869$$

This agrees with the figure we obtained earlier from Table A2.4. Since one control is being recruited for each case, no further adjustments need to be made.

Warning When using these formulae care should be taking to avoid rounding errors. An example of the effect these can have is illustrated below. The above calculations are repeated, but rounding off occurs at each step.

Example (cont'd)

$$\begin{aligned} A &= \frac{0.7 \times 1.5}{1 + 0.7 \times (1.5 - 1.0)} \\ &= \frac{1.05}{1.35} = 0.78 \\ C &= \frac{0.78 + 0.7}{2} = 0.74 \\ D &= 1 - 0.74 = 0.26 \\ N &= \frac{2 \times 0.74 \times 0.26 \times (2.0 + 1.3)^2}{(0.7 - 0.78) \times (0.7 - 0.78)} \\ &= \frac{4.19}{0.0064} \\ &= 655 \end{aligned}$$

Annex 3

Increasing this figure by 25% produces a sample size of 819 compared with the figure of 869 produced by the more careful calculation.

Table A3.3 Adjustment factors for samples recruiting more than one control per case

Number of controls per case (C)	Adjustment factor (F)
1	1 or 1.0
2	3/4 or 0.75
3	2/3 or 0.67
4	5/8 or 0.625
5	3/5 or 0.60
6	7/12 or 0.58
7	4/7 or 0.57
8	9/16 or 0.56
9	5/9 or 0.56
10	11/20 or 0.55

= = =