Chapter 15
Size of a study

15.1 Introduction

It is important to ensure at the design stage that the proposed number of subjects to be recruited into any study will be appropriate to answer the main objective(s) of the study. A small study may fail to detect important effects on the outcomes of interest, or may estimate them too imprecisely, no matter how good its design may be in other respects. A study larger than necessary, while less common in practice, may waste valuable resources.

Example 15.1. *Suppose that a trial was set up to assess the value of a new treatment for breast cancer. A total of 200 women with newly diagnosed breast cancer were randomly allocated to receive either the new or the standard treatment. All patients were followed up for one year after their entry into the trial (or until death if it occurred earlier). The outcome of interest was the proportion of women still alive by the end of the trial. The results are shown in Table 15.1.*

		New treatment	Standard treatment	Total
Alive one year after entry into the trial	Yes	80 (a)	70 (b)	150 (n_1)
	No	20 (c)	30 (d)	50 (n_0)
	Total	**100 (m_1)**	**100 (m_0)**	**200 (N)**

$p_1 = 80/100 = 80\%$
$p_0 = 70/100 = 70\%$
Risk difference = $p_1 - p_0 = 10\%$
95% confidence interval = – 2% to + 22%
$\chi^2 = 2.65$; $P > 0.10$

Table 15.1.
Number of patients with breast cancer still alive one year after entry into the trial by type of treatment administered: hypothetical data.

In Example 15.1, the χ^2 test of the difference between these two proportions gives a value of 2.65, which corresponds to $P > 0.10$. Thus the difference between the two proportions could easily have arisen by chance. However, we cannot conclude from this that there is no true difference between the treatments, since the 95% confidence interval for the difference between the proportions of patients still alive one year after entry into the trial is – 2% to +22%. Therefore the data from this trial are con-

sistent with a proportion of surviving patients on the new treatment up to 22% higher or 2% lower than the proportion of those on the standard treatment.

Thus, although the trial has shown that the new treatment does not perform appreciably worse than the standard treatment, it is unclear whether the two treatments have similar effects or whether the new treatment increases survival substantially. This is because the sample size of this trial was far too small to provide an appropriate answer to the question being addressed.

In the rest of this chapter, we will show how sample size estimates can be obtained in the simplest situation where two groups are to be compared. The calculations are based on the statistical methods presented in Chapter 6 and its appendix and readers should be familiar with their content before proceeding.

There are two main approaches to sample size calculations. One is based on the concept of *power* of a study, i.e., its ability to detect a statistically significant result if the true magnitude of the effect is as anticipated. Thus, this approach to sample size calculations focuses on the *significance test* that will be performed at the end of the study. In Example 15.1, we may estimate the sample size necessary to ensure that the study will have a certain probability ('power') of yielding a *P*-value less than 0.05 if the true difference in survival between the two treatments is 10%. The second approach focuses on the *precision* of the estimate, i.e., on the level of sampling error we regard as acceptable. As we saw in Chapter 6, the confidence interval provides an indication of how precise our sample estimate is in relation to the true population value. Thus, this approach focuses on the *width of the confidence interval* that will be obtained when the results of the study are analysed. In the breast cancer trial, we may estimate the sample size necessary to ensure that the trial will be able to estimate the true difference in survival within ± 2.5% of its value (i.e., the confidence interval will extend 2.5% to either side of the sample estimate).

In this chapter, we start by considering sample size calculations based on power and then move to calculations based on precision. The chapter ends with a discussion of how to apply such calculations to more complex study designs and other practical issues that need to be taken into account in estimating sample sizes. It is, however, important to emphasize at this stage that any sample size calculations involve some guesswork, since we have to start by anticipating the results of the proposed study and, therefore, these calculations should be regarded as *providing only a rough estimate of the required study size*. Moreover, as we shall see later in this chapter, other aspects (e.g., costs, availability of eligible subjects, logistic problems), independent of statistical considerations, also have to be taken into account in any practical situation.

15.2 Power of a study

The power of a study is the probability of obtaining a 'statistically significant' result, that is, a *P*-value below a certain pre-established 'signifi-

cance' level (usually 0.05) if the true magnitude of the effect is as anticipated. However, as discussed in Section 6.3, there is a close link between *P*-values and confidence intervals. Therefore, power can also be interpreted as the *probability of obtaining an estimate whose confidence interval does not include the value stipulated by the null hypothesis*. The null hypothesis states that the exposure has no effect on the outcome of interest corresponding to a value of zero, if the exposure effect is measured on an absolute scale (e.g., risk or rate difference), or one, if measured on a ratio scale (e.g., risk or rate ratio).

Figure 15.1(*a*) illustrates the relationship between the null hypothesis value, the anticipated effect estimate and its confidence interval when the exposure is associated with an increase in the occurrence of the outcome of interest. For the study to have appropriate power to detect such an effect, the lower limit of the confidence interval of the anticipated effect estimate has to be above the value stipulated by the null hypothesis. Similarly, when the exposure is associated with a decrease in incidence (i.e., the exposure is protective), the upper limit of the confidence interval has to be below the null hypothesis value (Figure 15.1(*b*)). Thus, the power of the study to detect a 'statistically significant' effect, if the true effect is as anticipated, is the probability that the lower limit of the confidence interval falls above (or, if the exposure is protective, the upper limit falls below) the null hypothesis value.

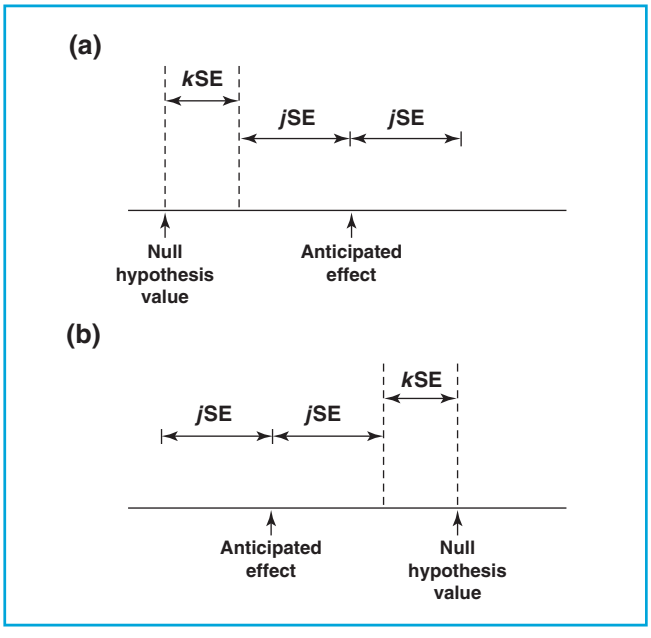

Figure 15.1.
Diagram illustrating the relationship of the null hypothesis value, the anticipated effect and its confidence interval when the exposure is associated with (*a*) an increase or (*b*) a decrease in the occurrence of the outcome of interest (adapted from Clayton & Hills, 1993).

The position of the lower limit (or the upper limit, if the exposure is protective) of the confidence interval of the anticipated effect estimate is determined by the *width of the confidence interval (± jSE)* (Figure 15.1), which in turn depends upon the *study size* (the bigger the study, the smaller the standard error (SE) and, therefore, the narrower the confidence interval) and upon the *significance (confidence) level chosen (j)*. For a 95% confidence interval, *j* would be equal to 1.96 (Table 15.2); that is, the confidence interval will extend 1.96SE to each side of the sample estimate. For a 99% confidence interval, *j* will be 2.576 and, therefore, the confidence interval will be wider. The wider the confidence interval, the lower the power of a study of a given size.

Suppose that the study were repeated several times. The effect estimates obtained each time and their confidence intervals would differ because of sampling variation. If the effect estimates obtained each time were plotted, we would obtain a Normal sampling distribution with a standard error of SE. Similarly, if the lower limits (or upper limits, if the exposure is protective) of each of the confidence intervals were plotted, we would obtain a Normal distribution, with the same standard error SE, centred around the anticipated value of the lower (or upper) limit. The power of a study is

Significance level	*j*
0.10	1.645
0.05	1.960
0.01	2.576

Power	*k*
0.95	1.645
0.90	1.282
0.75	0.674
0.50	0.0
<0.50	<0

Table 15.2.
Values of *k* and *j* for different significance levels and powers.

the probability that the lower limit of the confidence interval would fall above the null hypothesis value (or the upper limit would fall below it, if the exposure is protective). This probability depends upon the *number of standard errors (k)* between the null hypothesis and the anticipated position of the lower limit (or upper limit, if the exposure is protective) of the confidence interval of the anticipated effect estimate (Figure 15.1). It can be shown mathematically that if k is equal to 1.645, the study will have 95% power (Table 15.2). In other words, if the study were to be conducted repeatedly, we would expect only 5 out of 100 resulting 95% confidence intervals to include the null hypothesis value, if the true magnitude of the effect is as anticipated. When the anticipated location of the lower (or upper) confidence limit is exactly at the null hypothesis, so that $k = 0$, the power is 0.50 and there is an even chance of obtaining a significant result. If $k < 0$, the power will be less than 50%. In general, a power of less than 80% is regarded as unacceptable.

Thus, to summarize, the power of a study depends upon:

1. The *magnitude of the anticipated effect* (i.e., the distance between the null hypothesis value and the anticipated effect). The greater the effect, the higher the power to detect it as 'statistically significant' for a study of a given size.

2. The *width of the confidence interval (jSE)*, which determines the position of the lower limit (or the upper limit, if the exposure is protective). The wider the confidence interval, the lower the power of a study of a given size. This in turn depends on:

   (*a*) The study size. The bigger the study, the smaller the standard error (SE) and, therefore, the narrower the confidence interval.

   (*b*) The significance (confidence) level chosen (*j*). For instance, a 95% confidence interval ($j = 1.96$) will be narrower than a 99% confidence interval ($j = 2.576$) for a sample of a given size.

It is useful to construct power curves to show how the power varies with the study size for different significance levels and different magnitudes of the anticipated effect. Figure 15.2 shows some examples of such curves for the breast cancer trial described in Example 15.1.

In most real situations, researchers have a very good idea of the number of eligible subjects they will be able to recruit into their study. This number is usually determined by availability of eligible subjects, logistics of recruitment, costs, etc. In these circumstances the relevant question is not '*How large should the study be?*' but '*Is the available number of subjects enough to provide a clear answer to the study objectives?*'. To answer this last question, we need to estimate the power of the study with the proposed number of

subjects. If these calculations reveal that the power will be too low, it will be necessary to estimate by how much our sample size needs to be increased to ensure that the study will achieve the desired power. This is the approach suggested by Clayton & Hills (1993), which we will follow in the rest of this section.

15.2.1 Comparing two proportions (prevalences or risks)

To calculate the power of our breast cancer trial (Example 15.1) to detect a 10% difference in survival between the two treatments, we need to calculate first the SE of the difference between the two proportions of women still alive by the end of the first year. As we saw in Section A6.1.2, the SE of the difference between two proportions can be estimated, approximately, as

$$\mathrm{SE}(p_1 - p_0) = \sqrt{\frac{p_1^2(1-p_1)}{a} + \frac{p_0^2(1-p_0)}{b}}$$

Thus, in our example,

$$\mathrm{SE} = \sqrt{\frac{0.80^2(1-0.80)}{80} + \frac{0.70^2(1-0.70)}{70}} = 0.061$$

Figure 15.1(a) shows that the distance between the anticipated effect and the null hypothesis value (in our example, 0.10–0) is the sum of the two components, one deriving from the width of the confidence interval (jSE) and the other from the distance between the anticipated position of the lower limit of the confidence interval and the null hypothesis (kSE). Hence,

$$0.10 = j\,\mathrm{SE} + k\,\mathrm{SE}$$

If the significance level is set to 0.05 (j = 1.96) and SE = 0.061,

$$0.10 = 1.96 \times 0.061 + k \times 0.061$$

$$k = -0.32$$

This value of k corresponds to a power of less than 50% (Table 15.2). Thus, the probability of the trial being able to detect a statistically significant 10% difference in survival between the two treatments, even if such a difference truly exists, is below 50% because the confidence interval is too wide, due to the large SE. This is illustrated in Figure 15.3.

If we assume that the true difference in survival is 10% and we fix the significance level to 0.05 (j=1.96), we can calculate the value of the standard error that will be required to ensure that the power of the trial is 0.95 (k = 1.645):

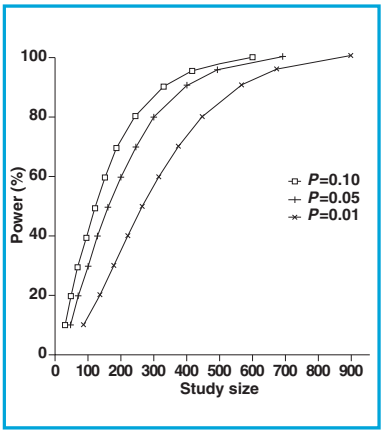

Figure 15.2.
Power for detecting a one-year survival difference of 10% when baseline survival is 70% (equivalent to 14% increase in survival) for various sample sizes and significance levels.

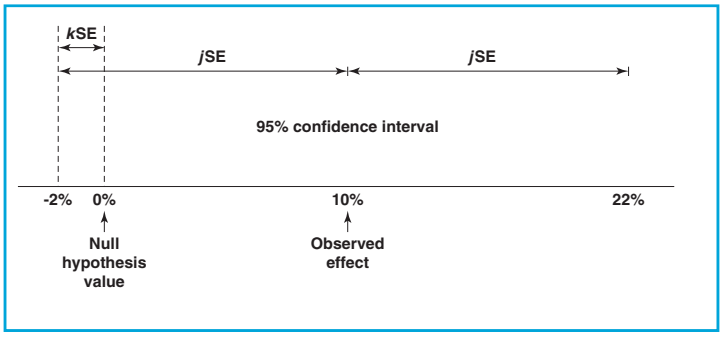

Figure 15.3.
Power calculation for the breast cancer clinical trial illustrated in Example 15.1: observed effect = 10% (= 80%–70%) difference in survival; sample size = 100 in each group; significance level = 0.05; power < 0.50.

$$1.96 \times SE + 1.645 \times SE = 0.10$$

$$\text{Required SE} = 0.10/(1.960 + 1.645) = 0.028$$

This value of the SE is much smaller than the value of 0.061 obtained with 100 subjects in each group. To reduce the SE to the required value, it is necessary to increase the size of the study. Clayton & Hills (1993) have provided the following formula to calculate a factor by which the study size must be increased in order to achieve the specified power:

$$(\text{Current value of SE}/\text{Required value of SE})^2$$

Thus, in our trial,

$$\text{Current value of SE} = 0.061$$
$$\text{Required value of SE} = 0.028$$
$$\text{Scale factor} = 4.8$$

The initial sample size was 100 in each group. To ensure that the study will have 95% power to detect a 10% difference in survival at the 5% significance level, we need to multiply this original sample size by 4.8. Thus, we need to enrol *480 subjects in each treatment group*.

As we saw in Section A6.1.3, the significance test is the same regardless of the type of measure of effect (ratio or difference) used to compare the two groups. Power calculations are based on the significance test that will be performed at the end of the study and, as a logical consequence, *similar sample size estimates would be obtained if the calculations were based on the ratio of the two proportions (80%/70%) rather than on their difference (80%–70%)*. The approach would be similar to the one followed above, except that the calculations would be based on the SE of the ratio of the two proportions rather than on the SE of their difference. Since the confidence interval around a ratio of proportions is asymmetric, taking only values from zero to infinity (Section A6.1.2), we first convert the estimated survival risk ratio into its natural logarithm (denoted by ln):

$$\ln(0.80/0.70) = \ln(1.14) = 0.134$$

We can now calculate an 'approximate' standard error of the logarithm of the ratio of two proportions (R) by using the formula given in Section A6.1.2:

$$\text{SE}(\ln R) = \sqrt{(1/a + 1/b - 1/m_1 - 1/m_0)}$$
$$= \sqrt{(1/80 + 1/70 - 1/100 - 1/100)} = 0.082$$

The distance between the logarithm of the anticipated effect and the logarithm of the null hypothesis value of 1 is equal to

$$\ln (1.14) - \ln (1) = 0.134 - 0 = 0.134$$

Thus,

$$0.134 = j \times \text{SE} + k \times \text{SE}$$

The value of the required SE to ensure that the study will have 95% power (k = 1.645) to detect a risk ratio of 1.14 at the 5% significance level (j = 1.96) will be

$$0.134 = 1.96 \times \text{SE} + 1.645 \times \text{SE}$$

$$\text{Required SE} = 0.134/(1.960 + 1.645) = 0.037$$

Therefore,

$$\text{Scale factor} = (0.082/0.037)^2 = 4.8$$

This value is exactly the one we obtained before when the calculation was based on the difference between the two proportions rather than on their ratio. Thus, the sample size required to detect a 10% increase in survival from 70% to 80% is equivalent to the sample size required to detect a risk ratio of 80%/70% = 1.14.

15.2.2 Comparing two rates

Similar sample size calculations can be performed for intervention trials and cohort studies in which the rate ratio (or rate difference) is the appropriate measure of effect.

In Example 15.2, we can predict the values of a and b by using the total person-time of observation in the proposed cohort study (y) and the estimated lung cancer incidence rate in the unexposed (r_0). Since lung cancer is a rare condition, we can estimate the total person-time of observation as

$$y = 40\,000 \times 5 \text{ years} = 200\,000 \text{ pyrs}$$

assuming there were no losses to follow-up and no other competing causes of death.

If 40% of the cohort is exposed to the risk factor under study, 0.60 is unexposed. Thus,

$$y_1 = 200\,000 \times 0.40 = 80\,000 \text{ pyrs}$$
$$y_0 = 200\,000 \times 0.60 = 120\,000 \text{ pyrs}$$

Since the anticipated rate ratio (RR) is equal to 2.0, the expected numbers of lung cancer cases among exposed and unexposed workers are

a = 80 000 pyrs × (2 × 50 per 100 000 pyrs) = 80
b = 120 000 pyrs × 50 per 100 000 pyrs = 60

Example 15.2. *Suppose we plan to conduct a cohort study to assess the effect of an occupational exposure on the incidence of lung cancer. We intend to recruit 40 000 middle-aged men into the study and follow them up for five years. 40% of the workers are known to be exposed to the hazard and we expect the lung cancer rate in those exposed to be twice that of workers unexposed (i.e., anticipated rate ratio = 2.0). It is estimated that the incidence rate of lung cancer in the unexposed group is 50 per 100 000 pyrs. The results to be obtained from this cohort study will be presented as in Table 15.3.*

Table 15.3.
Results from a hypothetical cohort study.

	Exposure		Total
	Yes	No	
No. of cases	a	b	n
Person-years at risk	y_1	y_0	y
Rate per 100 000 pyrs	r_1	r_0	r

We can now complete Table 15.3 with the results we expect to obtain from this cohort study if our assumptions are correct (Table 15.4).

Table 15.4.
Anticipated results from the proposed cohort study illustrated in Example 15.2.

	Exposure		Total
	Yes	No	
No. of cases	80 (a)	60 (b)	140 (n)
Person-years at risk	80 000 (y_1)	120 000(y_0)	200 000 (y)
Rate per 100 000 pyrs	100 (r_1)	50 (r_0)	70 (r)

As shown in Section A6.1.2, an 'approximate' SE of the logarithm of a rate ratio can be calculated as

$$\text{SE (ln RR)} = \sqrt{(1/a + 1/b)}$$

Thus, in our example,

$$\text{SE (ln RR)} = \sqrt{(1/80 + 1/60)} = 0.171$$

The number of SEs between the logarithm of the anticipated rate ratio (ln (2.0) = 0.693) and the logarithm of the null hypothesis value (ln (1) = 0) is

$$0.693 - 0 = j \times 0.\ 171 + k \times 0.171$$

$$j + k = 0.693/0.171 = 4.05$$

For a significance level of 0.05, j would be equal to 1.96 and, hence,

$$k = 4.05 - 1.96 = 2.09$$

This value of k corresponds to a power greater than 0.95 (Table 15.2). Thus, the probability that this cohort study will detect a true rate ratio of 2.0 (at the 0.05 significance level) is greater than 95%. Similar sample size estimates would be obtained if the calculations were based on the anticipated rate difference of 50 per 100 000 pyrs (100 per 100 000 pyrs – 50 per 100 000 pyrs).

If the power of the study with the proposed number of subjects were too low, we could have calculated by how much the sample size would have to be increased in order to achieve the required level by using the procedure described above when comparing two proportions.

15.2.3 Comparing two odds

A similar approach can be used in case–control studies. To estimate the power of the case–control study in Example 15.3, we need to guess the values of a, b, c and d. Since the controls are supposed to be representative of the population from which the cases will arise, we would expect 33.5% of them to be exposed and the rest to be unexposed to the factor under investigation. Thus

$$c = 200 \times 0.335 = 67$$
$$d = 200 - 67 = 133$$

Example 15.3. *Suppose we wish to conduct a study of 200 cases and 200 controls to detect an odds ratio of 0.5 for a particular exposure. The prevalence of this exposure in the population from which the cases arise is known to be 33.5%. The results to be obtained from this case–control study will be presented as in Table 15.5.*

	Exposure		Total
	Yes	No	
Cases	a	b	n_1
Controls	c	d	n_0
Total	$\boldsymbol{m_1}$	$\boldsymbol{m_0}$	$\boldsymbol{N}$

Table 15.5. Results from a hypothetical case-control study.

We can now calculate the odds of exposure among the controls as

$$c/d = 67/133 = 0.5$$

Since we anticipate an odds ratio equal to 0.5, we can calculate the odds of exposure among the cases as

$$a/b = 0.5 \times 0.5 = 0.25$$

Thus,

$$a = 0.25 \times b$$

Since $a + b = 200$, it follows that

$$b = 160$$
$$a = 200 - 160 = 40$$

We can now complete Table 15.5 with the values of a, b, c and d we expect to observe in the proposed case–control study (Table 15.6).

Table 15.6. Anticipated results from the hypothetical case–control study described in Example 15.3.

	Exposure		Total
	Yes	**No**	
Cases	40 (a)	160 (b)	200 (n_1)
Controls	67 (c)	133 (d)	200 (n_0)
Total	**107 (m_1)**	**293 (m_0)**	**400 (N)**

With these data, we can calculate an 'approximate' SE of the logarithm of the anticipated odds ratio by using the formula given in Section A6.1.2:

$$SE = \sqrt{(1/a + 1/b + 1/c + 1/d)}$$

Thus, in our case–control study,

$$SE = \sqrt{(1/40 + 1/160 + 1/67 + 1/133)}$$
$$= 0.232$$

To calculate the power of the study for a significance level of 0.05 (j = 1.960), we need first to calculate ln (OR) = ln (0.5) = –0.693. This is a negative value, but since we are interested in the absolute distance between the anticipated value and the null hypothesis we can ignore the minus sign and proceed as usual:

$$0.693 = 1.96 \times SE + k \times SE$$
$$0.693 - 1.96 \times 0.232 = k \times 0.232$$

$$0.238 = k \times 0.232$$
$$k = 1.03$$

A k value of 1.03 corresponds to a power around 0.85 (Table 15.2). If we wish to increase the power to 0.95, the study size has to be increased:

$$0.693 = 1.96 \times \text{SE} + 1.645 \times \text{SE}$$

$$\text{Required SE} = 0.693/(1.96 + 1.645) = 0.192$$

$$\text{Factor} = (0.232/0.192)^2 = 1.46$$

Thus, the proposed sample size must be increased by a factor of 1.46, that is from 200 to 292, if we wish to increase the power of the study from around 0.85 to 0.95.

Example 15.4. *Consider again the breast cancer trial (Example 15.1). For the trial to have 95% power to detect a 10% difference in survival at the 5% significance level, the sample size has to be increased from 100 to 480 subjects in each of the two treatment groups (Section 15.2.1). With this new sample size, we anticipate the following results (Table 15.7 and Figure 15.4).*

		New treatment	Standard treatment	Total
Alive one year after entry into the trial	Yes	384 (a)	336 (b)	720 (n_1)
	No	96 (c)	144 (d)	240 (n_0)
	Total	**480 (m_1)**	**480 (m_0)**	**960 (N)**

$p_1 = 384/480 = 80\%$

$p_0 = 336/480 = 70\%$

Risk difference = $p_1 - p_0 = 10\%$

95% confidence interval = +4.6% to + 15.4%

$\chi^2 = 12.79$; $P < 0.001$

Table 15.7.
Anticipated results of a hypothetical trial to assess the value of a new treatment on the one-year survival from breast cancer. Anticipated effect = 10% (= 80%–70%) difference in survival; significance level = 0.05; power = 0.95; sample size = 480 women in each treatment group.

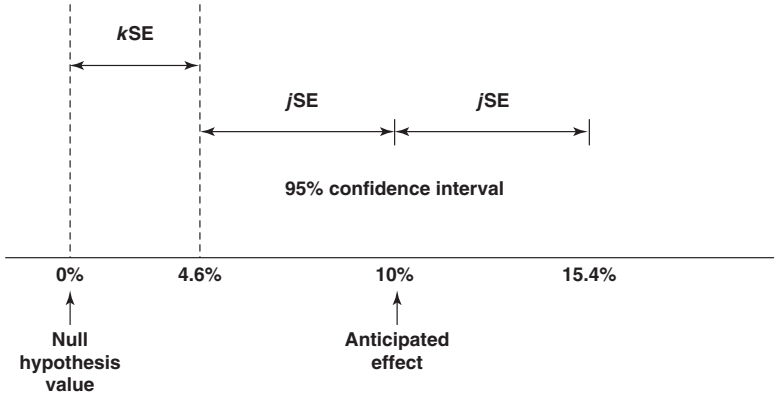

Figure 15.4.
Anticipated effect and its 95% confidence interval for the hypothetical breast cancer trial described in Example 15.4: anticipated effect = 10% (= 80%–70%) difference in survival; significance level = 0.05; power = 0.95; sample size = 480 women in each treatment group.

15.3 Sample size calculations based on precision

The approach to sample size calculations discussed in the previous section focused on the statistical significance test to be conducted at the end of the study. The main limitation of this approach is that it may produce very imprecise estimates of the effect on the outcome(s) of interest; thus, although the confidence interval will not include the null hypothesis value, it may still be too wide to be informative.

In Example 15.4, although the anticipated confidence interval does not include the null hypothesis value of no difference between the two treatments, its width is compatible with an improvement in one-year survival which ranges from 4.6% to 15.4% (Table 15.7; Figure 15.4). This is not a very precise estimate. Indeed, the range is wider than 10%, the difference we anticipate. We may consider more acceptable a width of ±2.5% either side of the sample estimate of 10%, so that the confidence interval of the difference in the proportion of women still alive by the end of the first year will range from 7.5% to 12.5%.

Sample size calculations based on power may be appropriate for new exposures, when it is not known whether there will be any impact at all on the outcomes of interest. If, however, other studies have already shown that the exposure is associated with either an increase or a decrease in incidence, there is not much point in testing the null hypothesis, and the objective should be to estimate the magnitude of the effect as precisely as possible. In these situations, it is more appropriate to choose a sample size that will yield a confidence interval of a predefined width.

Example 15.5. *Suppose we wish to conduct a cross-sectional survey in a certain area to estimate the prevalence of current oral contraceptive use among women aged 20–44 years. We plan to take a random sample from the population of all women aged 20–44 years living in the study area. We would like to calculate the sample size required to ensure that the study will be able to estimate the true prevalence of current oral contraceptive users in the study area within 5% of its value (i.e. the confidence interval that we will obtain when the results of the study are analysed will extend 5% to either side of the sample estimate).*

15.3.1 Estimating a single crude proportion (prevalence or risk)

We can estimate the sample size necessary to ensure that the confidence interval for a single proportion (prevalence or risk) is of a predetermined width.

In Example 15.5, we plan to take a random sample of n women aged 20–44 years. If among them, a are current users, we estimate the prevalence of oral contraceptive use as

$$p = a/n$$

As we saw in Section 6.1, this estimate is subject to sampling error, but the 95% confidence interval will give a range of values within which the true pop-

ulation prevalence will lie with 95% confidence.

$$SE(p) = \sqrt{\frac{p^2(1-p)}{a}}$$

Suppose we wish our confidence interval around the sample estimate to be of a certain width (± w). The value of w depends upon the standard error SE and the significance level (j) chosen:

$$w = j \times SE$$

For a 95% confidence interval, j = 1.96, that is the interval extends 1.96 standard errors either side of the estimate (w = 1.96 × SE).

Hence, we can estimate the prevalence of oral contraceptive use (p) with a pre-defined degree of precision by choosing an appropriate sample size (n). We must first guess the value of p. Suppose that statistics on oral contraceptive sales indicate that the prevalence of use is about 50% and we want to estimate it to within ±5%. Thus

$$p = 0.50$$
$$w = j \times SE = 0.05$$

Choosing a 95% confidence level,

$$1.96 \times SE = 0.05$$
$$SE(p) = 0.05/1.96 = 0.0255$$

Since p = 0.5, we can estimate a from the formula for the SE:

$$SE(p) = \sqrt{\frac{p^2(1-p)}{a}} = 0.0255$$

$$\sqrt{(0.5^2 \times 0.5)/a} = 0.0255$$
$$0.125/a = 0.0255^2$$
$$a = 192$$

Finally, we can calculate the sample size (n) required as

$$p = a/n$$
$$0.5 = 192/n$$
$$n = 384$$

Thus, we need to enrol 384 women into the study. When planning a study, it is a good idea to find out what sample size will be required for various levels of precision (Table 15.8 and Figure 15.5).

Table 15.8.
Sample sizes required to estimate a true prevalence of 0.50 with 95% confidence intervals of different widths (±*w*).

Width (±*w*)	Sample size (*n*)
0.01	9612
0.02	2403
0.03	1068
0.04	600
0.05	384
0.06	266
0.07	196
0.08	150
0.09	118
0.10	96
0.15	43

Thus, to estimate a true prevalence of oral contraceptive use of 50% within ±1% (i.e., from 49% to 51%), we would need to recruit 9612 women.

It is also important to calculate the required sample size for different values of the anticipated prevalence *p*. As we can see in Table 15.9, the sample size required does not vary much for values of *p* between 0.3 and 0.7, being greatest when $p = 0.50$. Thus, to be on the safe side, we can set $p = 0.50$ and obtain the maximum sample size (*n*) required.

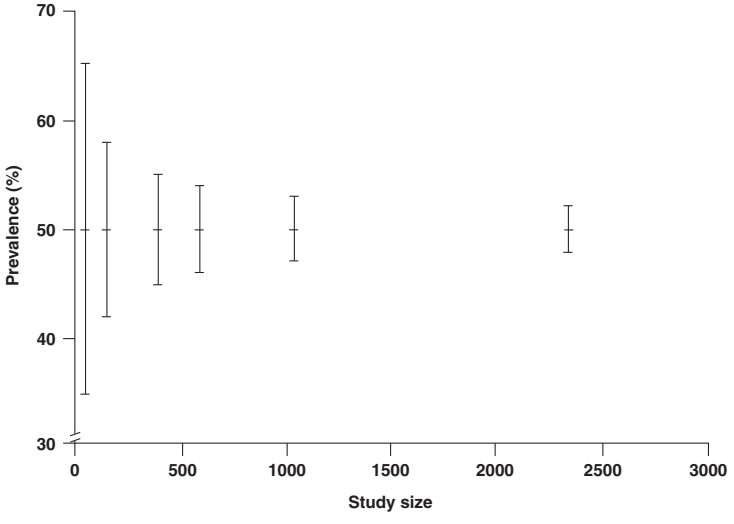

Figure 15.5.
Anticipated 95% confidence intervals for a true prevalence of 0.50 for various sample sizes.

Prevalence (*p*)	Sample size (*n*)
0.10	138
0.20	245
0.30	322
0.40	367
0.50	384
0.60	367
0.70	322
0.80	245
0.90	138

Table 15.9.
Sample size required to ensure that the 95% confidence interval for different levels of prevalence (*p*) will extend $w = 0.05$ to each side of the sample estimate.

15.3.2 Estimating a single crude rate

A similar approach can be used to estimate the sample size necessary to ensure that the confidence interval for a single rate is of a predetermined width.

***Example 15.6**. Suppose we wish to determine the incidence rate of a particular condition in a certain population. Based on data from previously conducted studies, we expect the rate to be about 50 per 10 000 pyrs. We want to determine the size of the sample that will be required to estimate the incidence rate in that population within ±5 per 10 000 pyrs.*

In Example 15.6, we plan to take a random sample of individuals from the study population. Thus, for a 95% confidence level ($j = 1.96$)

$$w = 1.96 \times \text{SE}$$
$$5 \text{ per } 10\,000 \text{ pyrs} = 1.96 \times \text{SE}(r)$$
$$\text{SE}(r) = 2.55 \text{ per } 10\,000 \text{ pyrs}$$

An 'approximate' standard error of a rate can be calculated as indicated in Section A.6.1.1:

$$\mathrm{SE}(r) = \frac{r}{\sqrt{a}}$$

where r is the estimated rate and a is the number of cases that occurred during the observation period. Thus, in our example,

$$\frac{50 \text{ per } 10\,000 \text{ pyrs}}{\sqrt{a}} = 2.55 \text{ per } 10\,000 \text{ pyrs}$$

$$a = 384$$

We can now calculate the person-time at risk (y) required to originate 384 cases:

$r = a/y$
50 per 10 000 pyrs = $384/y$
y = 76 800 pyrs

This level of total person-years at risk can be achieved by following 76 800 individuals for one year or 38 400 for two years, etc.

15.3.3 Estimating a difference or ratio of two proportions (risks or prevalences)

Let us consider again the breast cancer trial (Examples 15.1 and 15.4). Suppose we want to ensure that the width of the confidence interval for the difference in proportions will be equal to $\pm w = 2 \times 2.5\%$. Thus

$$w = 1.96 \times \mathrm{SE}\,(p_1 - p_0) = 0.025$$

The required SE should be

$$0.025/1.96 = 0.0128$$

and

$$\mathrm{SE}(p_1 - p_0) = \sqrt{\frac{p_1^{\,2}(1-p_1)}{a} + \frac{p_0^{\,2}(1-p_0)}{b}} = 0.0128$$

Since $a = 0.80m_1$, $b = 0.70m_0$ and $m_0 = m_1$, it follows that

$$m_0 = m_1 = a/0.80$$
$$b = 0.70 \times (a/0.80) = 0.875\,a$$

The above formula for the standard error can then be re-written as

$$SE = \sqrt{\frac{0.80^2(1-0.80)}{a} + \frac{0.70^2(1-0.70)}{0.875a}} = 0.0128$$

$$(0.0128)^2 = 0.128/a + 0.147/0.875a$$
$$0.000164 = ((0.128 \times 0.875) + 0.147)/0.875a$$
$$0.875a = 0.259/0.000164 = 1579.3$$
$$a = 1805$$
$$m_1 = m_0 = 1805/0.80 = 2256$$
$$b = 2256 \times 0.70 = 1579$$

Thus, to obtain a 95% confidence interval for the survival difference between the two treatments with a width of 2.5% either side of the anticipated effect of 10% (i.e., from 7.5% to 12.5%), we need to enrol 2256 subjects in each treatment group.

Similarly, we can calculate the sample size required to estimate a ratio of two proportions with a pre-defined level of precision. The approach would be similar to the one just illustrated, except that the calculations would be based on the formula for the SE of a ratio of proportions (see Section A6.1.2).

15.3.4 Estimating a rate difference or a rate ratio

A similar approach can be used in studies where the appropriate measure of effect is the rate ratio.

***Example 15.7**. The incidence of stomach cancer among men aged 50–59 years in a particular population is 65 per 100 000 person-years. Suppose that we are planning to conduct a trial in that population to assess whether a particular intervention reduces the rate of stomach cancer in men of that age-group. Eligible subjects will be randomized to receive either the intervention or the placebo in a ratio of 1:1. You expect the rate among those who receive the intervention to be 60% of the rate in those administered placebo (i.e., a 40% reduction in risk). We wish the confidence interval for the rate ratio estimate to have a width (on a logarithmic scale) of 1.30 either side of the sample estimate (i.e., from 0.46 (= 0.60/1.30) to 0.78 (= 0.60 × 1.30)). The results from this trial will be presented as in Table 15.3.*

An 'approximate' 95% confidence interval of the logarithm of a rate ratio (r_1/r_0) can be estimated as

$$95\%\ \text{confidence interval} = \ln RR \pm 1.96 \times SE(\ln RR)$$

In Example 15.7,

$$w = \ln(1.30) = 1.96 \times \text{SE}(\ln \text{RR})$$

$$\text{SE}(\ln \text{RR}) = 0.262/1.96 = 0.134$$

An 'approximate' standard error of the logarithm of an estimated rate ratio (RR) can then be obtained as follows:

$$\text{SE}(\ln \text{RR}) = \sqrt{(1/a + 1/b)}$$

In our example,

$$0.134 = \sqrt{(1/a + 1/b)}$$

Since the same number of subjects are to be allocated to each arm of the trial and we anticipate a rate ratio of 0.6, then

$$a = 0.6 \times b$$
$$0.134 = \sqrt{(1/0.6b + 1/b)} = \sqrt{(1.6/0.6b)}$$
$$(0.134)^2 = 1.6/0.6b$$
$$0.6b = 1.6/0.018 = 88.89$$
$$b = 88.89/0.6 = 148$$
$$a = 0.6 \times 148 = 89$$

The stomach cancer incidence rate for those given placebo is believed to be 65 per 100 000 pyrs. Thus, we need 227 692 pyrs (= 148/0.000065 pyrs) of follow-up in each arm of the trial in order to accumulate 148 stomach cancer cases among those receiving the intervention and 89 among those given placebo. This will ensure that the result of the trial will have the desired level of precision. This can be achieved by following 227 692 men for one year, 113 846 for two years and so on.

A similar approach can be used to calculate sample size estimates based on precision in situations where the appropriate measure of effect is the rate difference rather than the rate ratio, except that the calculations would be based on the formula of the SE of a difference in rates (see Section A6.1.2).

15.3.5 Estimating an odds ratio

Consider Example 15.8.

Example 15.8*. Suppose we wish to conduct a study to detect an odds ratio of 2.0 for an exposure present in 33.5% of the population from which the cases will arise. A similar number of cases and controls will be recruited into the study. We want to ensure that the 95% confidence interval for the odds ratio will have a width (on a logarithmic scale) of 1.25 either side of the sample estimate (i.e., from 1.6 (= 2.0/1.25) to 2.5 (= 2.0 × 1.25)). The results from this case–control study will be presented as in Table 15.5.*

A 95% confidence interval of the logarithm of an odds ratio can be estimated as

$$95\%\ \text{confidence interval} = \ln \text{OR} \pm 1.96 \times \text{SE}(\ln \text{OR})$$

In our example,

$$w = \ln(1.25) = 1.96 \times \text{SE}(\ln \text{OR})$$
$$\text{SE}(\ln \text{OR}) = 0.223/1.96 = 0.114$$

An 'approximate' standard error of the logarithm of the estimated odds ratio can be obtained as

$$\text{SE}(\ln \text{OR}) = \sqrt{(1/a + 1/b + 1/c + 1/d)}$$

The odds of exposure among the cases is expected to be twice the odds of exposure among the controls:

$$a/b = 2 \times (c/d)$$

and the prevalence of exposure among the controls is expected to be 33.5%:

$$c = n_0 \times 0.335$$
$$d = n_0 \times 0.665$$

Hence,

$$a/b = 2 \times \frac{n_0 \times 0.335}{n_0 \times 0.665} = 1.008$$

$$a = 1.008 \times b$$
$$n_0 = n_1 = a + b = 1.008b + b = 2.008b$$
$$0.114 = \sqrt{\frac{1}{1.008b} + \frac{1}{b} + \frac{1}{2.008b \times 0.335} + \frac{1}{2.008b \times 0.665}}$$
$$0.114^2 = 1/b \times (1/1.008 + 1 + 1/0.673 + 1/1.335)$$
$$0.013 = 4.227/b$$
$$b = 325$$
$$a = 328$$
$$c = (325 + 328) \times 0.335 = 219$$
$$d = (325 + 328) \times 0.665 = 434$$
$$n_0 = a + b = 325 + 328 = 653$$
$$n_1 = c + d = 219 + 434 = 653$$

Thus, we need to recruit 653 cases and 653 controls into our case-control study.

15.4 Other considerations concerning study size

The methods described in this chapter should be regarded only as providing a rough estimate of the required study size, as they are based on guesses or approximate estimates of the parameters, on subjective decisions about the size of an effect that we would wish to detect, and on the use of approximate formulae. They only give an idea of the sort of size needed.

In practice there will be constraints on resources, which may limit the maximum possible sample size. Resources in terms of staff, vehicles, laboratory capacity, time and money are all likely to be limited, and it is usually necessary to achieve a balance between the results of the study size calculations and what can reasonably be managed given the resources. Trying to do a study that is beyond the capacity of the available resources is likely to be unfruitful, as data quality will suffer, and the study may collapse before completion, wasting the investment that had already been put into it. On the other hand, if the calculations show that a study of a manageable size will have a power and/or yield a precision that is unacceptably low, it may be worth considering involving other centres.

Study size calculation should always be carried out for several different scenarios, not just one (e.g., different levels of power/precision and of estimates of the effect measure), in order to give a clear picture of the scope of the study. A useful approach in deciding on the trade-off between cost and power is to construct power curves for one or two key outcome variables, to show how the power or precision varies with the study size for different values of the effect measure (as shown in Figure 15.2).

15.4.1 Studies designed to look at multiple exposure–outcome relationships

Many epidemiological studies are designed to look at several exposure-outcome relationships. We can use the methods described in this chapter to calculate the study size necessary to allow us to detect the most important exposure–outcome relationships. Ideally, we would then select the largest of these as our actual study size. We may find that the study size required to detect one or more of the exposure–outcome relationships in which we are interested is clearly beyond the available resources. For instance, we may be interested in examining the relationship between birthweight and the occurrence of benign breast diseases and breast cancer. The study size calculations may reveal that our sample will be sufficient to ensure that the study will have enough power or precision to examine the relationship between birthweight and benign breast diseases but will be far too small to examine the relationship between birthweight and breast cancer incidence. If we are unable to increase the sample size to the desired level, we will have to restrict the objectives of the study and consider benign breast disorders as the only outcome of interest.

15.4.2 Refusals and losses to follow-up

Refusals and/or losses to follow-up are likely to occur in epidemiological studies because subjects move away from the study area, die from some cause unrelated to the outcome of interest, refuse to continue with the study, etc. These losses reduce the size of the study available for analysis and, therefore, decrease the power or precision of the study. We can compensate for these losses by increasing the initial study size. For example, if study size calculations suggest that 320 subjects are required, and a 20% loss to follow-up is expected, the study size should be increased to about 400 ($400 - 400 \times 0.20 = 320$). Note, however, that although this strategy will ensure that the study will have the required power or precision, it will not avoid the possibility of *selection bias*, as the individuals who refuse to participate or are lost to follow-up may differ in important respects from those who remain in the study.

15.4.3 Unequal sized groups

There are situations where we may wish to study groups of different sizes. For example, when the number of available cases and controls is large and the cost of obtaining information from both groups is comparable, the optimal control-to-case ratio is 1:1. However, if the available number of cases for the study is limited, or when the cost of obtaining information is greater for cases than controls, the number of controls per case can be increased to achieve the necessary power or precision. 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. It is usually not recommended to increase the control:case ratio beyond 4:1, because there is only a small increase in statistical power with each additional control beyond this point. This small increase in power is generally not worth the increase in logistics and costs of recruiting a much larger total number of subjects unless the data for the controls are available at very little extra cost.

15.4.4 Confounding and interaction

The sample size estimations presented here refer to *crude* measures. In most studies, however, it is essential to control for confounding variables by calculating *adjusted* measures. These adjustments usually lead to a loss of power or precision of the study. However, this loss is substantial only when a very strong confounding variable is present (Smith & Day, 1984).

In contrast, and as discussed in the previous chapter, studies designed with the aim of detecting interactions require much larger sample sizes than those designed to look for simple effects (Smith & Day, 1984). In practice, this restricts the number of studies that can be carried out explicitly to examine interactions.

15.4.5 Special study designs

The sample size calculations for case–control studies presented in this chapter referred to *unmatched* studies. Methods for calculation of adequate sample sizes for matched case–control studies are given in Breslow & Day (1980).

The methods presented in this chapter assume that *individuals* rather than *groups* are the units of study. The calculations would be different if, for instance, the unit of study were communities (see Smith & Morrow, 1996).

Box 15.1. Key issues

- Sample size calculations can be made to ensure that:

  1. the study has enough *power*, i.e., ability to detect a statistically significant result if the true magnitude of the effect is as anticipated. Thus, this approach focuses on the significance test that will be performed at the end of the study.

  2. the sample estimates are *precise*, i.e., the level of sampling error is low. This approach focuses on the width of the confidence interval.

- Sample size calculations based on power may be appropriate to identify new exposures when it is not known whether they will have any effect at all on the outcome(s) of interest. If, however, it is already known that the exposure is associated with the outcome, the objective of the study should be to quantify the magnitude of the effect as precisely as possible rather than just testing the null hypothesis of no effect. In these circumstances, sample size calculations should be based on precision.

- Sample size calculations should be taken as providing only a rough idea of the number of subjects that need to be recruited. Other considerations such as availability of subjects, resources, costs, etc. should be considered carefully.

Further reading

* The power calculations presented in Section 15.2 follow the approach suggested by Clayton & Hills (1993).

* Sample size calculations for more complex designs of cohort, case–control and intervention studies are given respectively in Breslow & Day (1987, 1980) and Smith & Morrow (1996).