# 1

# Basic Design Considerations

**SUMMARY**

This chapter reviews the reasons why sample size considerations are important when planning a clinical study of any type. The basic elements underlying this process include the null and alternative study hypotheses, effect size, statistical significance level and power, each of which are described. We introduce the notation to distinguish the population parameters we are trying to estimate with the study, from their anticipated value at the planning stages and also from their estimated value once the study has been completed. We emphasise for comparative studies that, whenever feasible, it is important to randomise the allocation of subjects to respective groups.

The basic properties of the standardised Normal distribution are described. Also discussed is how, once the effect size, statistical significance level and power for a comparative study using a continuous outcome are specified, the Fundamental Equation (which essentially plays a role in most sample size calculations for comparative studies) is derived.

The Student's *t*-distribution and the Non-central *t*-distribution are also described. In addition the Binomial, Poisson, Negative-Binomial, Beta and Exponential statistical distributions are defined. In particular, the circumstances (essentially large study sizes) in which the Binomial and Poisson distributions have an approximately Normal shape are described. Methods for calculating confidence intervals for a population mean are indicated together with (suitably modified) how they can be used for a proportion or a rate in larger studies. For the Binomial situation, formulae are also provided where the sample size is not large. Finally, a note concerning numerical accuracy of the calculations in the illustrative examples of later chapters is included.



## 1.1 Why Sample Size Calculations?

To motivate the statistical issues relevant to sample size calculations, we will assume that we are planning a two-group clinical trial in which subjects are allocated at random to one of two alternative treatments for a particular medical condition and that a single endpoint measure has been specified in advance. However, it should be emphasised that the basic principles described, the formulae, sample size tables and associated software included in this book are equally relevant to a wide range of design types covering all areas of medical research ranging from the epidemiological to clinical and laboratory-based studies.

Whatever the field of inquiry the investigators associated with a well-designed study will have considered the research questions posed carefully, formally estimated the required sample size (the particular focus for us in this book), and recorded the supporting reasons for their choice. Awareness of the importance of these has led to the major medical and related journals demanding that a detailed justification of the study size be included in any submitted article as it is a key component for peer

*Sample Sizes for Clinical, Laboratory and Epidemiology Studies*, Fourth Edition. David Machin, Michael J. Campbell, Say Beng Tan and Sze Huey Tan.

reviewers to consider when assessing the scientific credibility of the work undertaken. For example, the *General Statistical Checklist* of the *British Medical Journal* asks statistical reviewers of their submitted papers 'Was a pre-study calculation of study size reported?' Similarly, many research grant funding agencies such as the Singapore National Medical Research Council now also have such requirements in place.

In any event, at a more mundane level, investigators, grant-awarding bodies and medical product development companies will all wish to know how much a study is likely to 'cost' both in terms of time and resources consumed as well as monetary terms. The projected study size will be a key component in this 'cost'. They would also like to be reassured that the allocated resource will be well spent by assessing the likelihood that the study will give unequivocal results. In particular for clinical trials, the regulatory authorities, including the Committee for Proprietary Medicinal Products (CPMP, 1995) in the European Union and the Food and Drug Administration (FDA, 1988 and 1996) in the USA, require information on planned study size. These are encapsulated in the guidelines of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (1998) ICH Topic E9.

If too few subjects are involved, the study is potentially a misuse of time because realistic differences of scientific or clinical importance are unlikely to be distinguished from chance variation. Too large a study can be a waste of important resources. Further, it may be argued that ethical considerations also enter into sample size calculations. Thus a small clinical trial with no chance of detecting a clinically useful difference between treatments is unfair to all the patients put to the (possible) risk and discomfort of the trial processes. A trial that is too large may be unfair if one treatment could have been 'proven' to be more effective with fewer patients as a larger than necessary number of them has received the (now known) inferior treatment.

Providing a sample size for a study is not simply a matter of providing a single number from a set of statistical tables. It is, and should be, a several-stage process. At the preliminary stages, what is required are 'ball-park' figures that enable the investigators to judge whether or not to start the detailed planning of the study. If a decision is made to proceed, then the later stages are used to refine the supporting evidence for the preliminary calculations until they make a persuasive case for the final patient numbers chosen. Once decided this is then included (and justified) in the final study protocol.

After the final sample size is determined and the protocol is prepared and approved by the relevant bodies, it is incumbent on the research team to expedite the recruitment processes as much as possible, ensure the study is conducted to the highest of standards possible, and ensure that it is eventually reported comprehensively.

## 1.2 Statistical Significance

### Notation

In very brief terms the (statistical) objective of any study is to estimate from a sample the value of a population parameter. For example, if we were interested in the mean birth weight of babies born in a certain locality, then we may record the weight of a selected sample of $N$ babies and their mean weight $\overline{w}$ is taken as our estimate of the population mean birth weight denoted $\omega_{Pop}$. The Greek $\omega$ distinguishes the population value from its estimate, the Roman $\overline{w}$. When planning a study, we are clearly ignorant of $\omega_{Pop}$ and neither do we have the data to calculate $\overline{w}$. As we shall see later, when

planning a study the investigators will usually need to provide some value for what $\omega_{Pop}$ may turn out to be. This anticipated value is denoted $\omega_{Plan}$. This value then forms (part of) the basis for subsequent sample size calculations.

### Outcomes

In any study, it is necessary to define an outcome (endpoint) which may be, for example, the birth weight of the babies concerned, as determined by the objectives of the investigation. In other situations this outcome may be a measure of blood pressure, wound healing time, degree of palliation, a patient reported outcome (PRO) that indicates the level of some aspect of their Quality of Life (QoL) or any other relevant and measureable outcome of interest.

### The Effect Size

Consider, as an example, a proposed randomised trial of a placebo (control, $C$) against acupuncture ($A$) for the relief of pain in patients with a particular diagnosis. The patients are randomised to receive either $A$ or $C$ (how placebo acupuncture can be administered is clearly an important consideration). In addition, we assume that pain relief is assessed at a fixed time after randomisation and is defined in such a way as to be unambiguously evaluable for each patient as either 'success' or 'failure'. We assume the aim of the trial is to estimate the true difference $\delta_{Pop}$ between the true success rate $\pi_{PopA}$ of $A$ and the true success rate $\pi_{PopC}$ of $C$. Thus the key (population) parameter of interest is $\delta_{Pop}$ which is a composite of the two (population) parameters $\pi_{PopA}$ and $\pi_{PopC}$.

At the completion of the trial the $A$ patients yield a treatment success rate $p_A$ which is an estimate of $\pi_{PopA}$ and for $C$ the corresponding items are $p_C$ and $\pi_{PopC}$. Thus, the observed difference, $d = p_A - p_C$, provides an estimate of the true difference (the effect size) $\delta_{Pop} = \pi_{PopA} - \pi_{PopC}$.

### Significance Tests

In a clinical trial, two or more forms of therapy or intervention may be compared. However, patients themselves vary both in their baseline characteristics at diagnosis and in their response to subsequent therapy. Hence in a clinical trial, an apparent difference in treatments may be observed due to chance alone, that is, we may observe a difference but it may be explained by the intrinsic characteristics of the patients themselves rather than 'caused' by the different treatments given. As a consequence, it is customary to use a 'significance test' to assess the weight of evidence and to estimate the probability that the observed data could in fact have arisen purely by chance.

### The Null Hypothesis and Test Size

In our example, the null hypothesis, termed $H_{Null}$, implies that $A$ and $C$ are equally effective or that $\delta_{Pop} = \pi_{PopA} - \pi_{PopC} = 0$. Even when that null hypothesis is true, at the end of the study an observed difference, $d = p_A - p_C$ other than zero, may occur. The probability of obtaining the observed difference $d$ or a more extreme one, on the *assumption* that $\delta_{Pop} = 0$, can be calculated using a statistical test. If, under this null hypothesis, the resulting probability or $p$-value is very small, then we reject this null hypothesis of no difference and conclude that the two treatments do indeed differ in efficacy.

The critical value taken for the $p$-value is arbitrary and is denoted by $\alpha$. If, once calculated following the statistical test, the $p$-value $\leq \alpha$ then the null hypothesis is rejected. Conversely, if the $p$-value $> \alpha$, one does not reject the null hypothesis. Even when the null hypothesis is in fact true there is a risk of rejecting it. To reject the null hypothesis when it is true is to make a Type I error and the associated probability of this is $\alpha$. The quantity $\alpha$ can be referred to either as the test size, significance level, probability of a Type I error or, sometimes, the false-positive error.

### The Alternative Hypothesis and Power

Usually in statistical significance testing, by rejecting the null hypothesis, we do not specifically accept any alternative hypothesis, and it is usual to report the range of plausible population values with a confidence interval (*CI*) as we describe in **Section 1.6**. However, sample size calculations are usually posed in a hypothesis test framework, and this requires us to specify an alternative hypothesis, termed $H_{Alt}$, that the *true* effect size is $\delta_{Pop} = \pi_{PopA} - \pi_{PopC} \neq 0$.

The clinical trial could yield an observed difference $d$ that would lead to a $p$-value $> \alpha$ even though the null hypothesis is really *not* true, that is, $\pi_{PopA}$ truly differs from $\pi_{PopC}$ and so $\delta_{Pop} \neq 0$. In such a situation, we then *fail* to reject the null hypothesis although it is indeed false. This is called a Type II or false-negative error and the probability of this is denoted by $\beta$.

As the probability of a Type II error is based on the assumption that the null hypothesis is *not* true, that is, $\delta_{Pop} \neq 0$, then there are many possible values for $\delta_{Pop}$ in this instance. Since there are countless potential values then each would give a different value for $\beta$.

The *power* is defined as one minus the probability of a Type II error, $1 - \beta$. Thus 'power' is the probability of what 'you want', which is obtaining a 'significant' $p$-value when the null hypothesis is *truly* false and so a difference between two interventions may be claimed.

## 1.3 Planning Issues

### The Effect Size

Of the parameters that have to be pre-specified before the sample size can be determined, the true effect size is the most critical. Thus, in order to estimate sample size, one must first identify the magnitude of the difference between the interventions $A$ and $C$ that one wishes to detect (strictly the minimum size of scientific or clinical interest) and quantify this as the (anticipated) effect size denoted $\delta_{Plan}$. Although what follows is couched in terms of planning a randomised control trial, analogous considerations apply to all comparative study types.

Sometimes there is prior knowledge that enables an investigator to anticipate what size of benefit the test intervention is likely to bring, and the role of the trial is to confirm that expectation. In other circumstances, it may be possible to say that, for example, only the prospect of doubling of their median survival would be worthwhile for patients with a fatal disease who are rapidly deteriorating. This is because the test treatment is known to be toxic and likely to be a severe burden for the patient as compared to the standard approach.

One additional problem is that investigators are often optimistic about the effect of test interventions; it can take considerable effort to initiate a trial and so, in many cases, the trial would only be launched if the investigating team is enthusiastic about the new treatment $A$ and is sufficiently

convinced about its potential efficacy over *C*. Experience suggests that as trials progress there is often a growing realism that, even at best, the initial expectations were optimistic. There is also ample historical evidence to suggest that trials which set out to detect large effects nearly always result in 'no significant difference was detected'. In such cases there may have been a true and clinically worthwhile, but smaller, benefit that has been missed, since the level of detectable difference set by the design was unrealistically high and hence the sample size too small to detect this important difference.

It is usual for most clinical trials that there is considerable uncertainty about the relative merits of the alternative interventions so that even when the new treatment or intervention under test is thought for scientific reasons to be an improvement over the current standard, the possibility that this is not the case is allowed for. For example, in the clinical trial conducted by Chow, Tai, Tan, *et al* (2002) it was thought, at the planning stage, that high dose tamoxifen would not compromise survival in patients with inoperable hepatocellular carcinoma. This turned out not to be the case and, if anything, tamoxifen was detrimental to their ultimate survival time. This is not an isolated example.

In practice, when determining an appropriate effect size, a form of iteration is often used. The clinical team might offer a variety of opinions as to what clinically useful difference will transpire — ranging perhaps from an unduly pessimistic small effect to the optimistic (and unlikely in many situations) large effect. Sample sizes may then be calculated under this range of scenarios with corresponding patient numbers ranging perhaps from extremely large to relatively small. The importance of the clinical question and/or the impossibility of recruiting large patient numbers may rule out a very large trial but conducting a small trial may leave important clinical effects not firmly established. As a consequence, the team may next define a revised aim maybe using a summary derived from their individual opinions, and the calculations are repeated. Perhaps the sample size now becomes attainable and forms the basis for the definitive protocol.

There are a number of ways of eliciting useful effect sizes using clinical opinion: a Bayesian perspective has been advocated by Spiegelhalter, Freedman and Parmar (1994), an economic approach by Drummond and O'Brien (1993) and one based on patients' perceptions rather than clinicians' perceptions of benefit by Naylor and Llewellyn-Thomas (1994). Gandhi, Tan, Chung and Machin (2015) give a specific case study describing the synthesis of prior clinical beliefs, with information from non-randomised and randomised trials concerning the treatment of patients following curative resection for hepatocellular carcinoma. Cook, Hislop, Altman et al (2015) also give useful guidelines for selection of an appropriate effect size.

### One- or Two-Sided Significance Tests

It is plausible to assume in the acupuncture trial referred to earlier that the placebo is in some sense 'inactive' and that any 'active' treatment will have to perform better than the 'inactive' treatment if it is to be adopted into clinical practice. Thus rather than set the alternative hypothesis as $H_{Alt}$: $\pi_{PopA} \neq \pi_{PopC}$, it may be replaced by $H_{Alt}$: $\pi_{PopA} > \pi_{PopC}$. This formulation leads to a 1-sided statistical significance test.

On the other hand, if we cannot make this type of assumption about the new treatment at the design stage, then the alternative hypothesis is $H_{Alt}$: $\pi_{PopA} \neq \pi_{PopC}$. This leads to a 2-sided statistical significance test.

For a given sample size, a 1-sided test is more powerful than the corresponding 2-sided test. However, a decision to use a 1-sided test should never be made after looking at the data and observing

the direction of the departure. Such decisions should be made at the design stage, and a 1-sided test should *only* be used if it is *certain* that departures in the particular direction *not anticipated* will always be ascribed to chance and therefore regarded as non-significant, however large they turn out to be.

It is more usual to carry out 2-sided tests of significance *but*, if a 1-sided test is to be used, this should be indicated and justified clearly for the problem in hand. **Chapter 6**, which refers to post-marketing studies, and **Chapter 11**, which discusses non-inferiority trials, give some examples of studies where the use of a 1-sided test size can be clearly justified.

### Choosing $\alpha$ and $\beta$

It is customary to start by specifying the effect size required to be detected and then to estimate the number of patients necessary to enable the trial to detect this difference if it truly exists. Thus, for example, it might be anticipated that acupuncture could improve the response rate from 20% with *C* to 30% with *A* and, since this is deemed a plausible and medically important improvement, it is desired to be reasonably certain of detecting such a difference if it really exists. 'Detecting a difference' is usually taken to mean 'obtaining a statistically significant difference with the $p$-value $< 0.05$'; and similarly the phrase 'to be reasonably certain' is usually interpreted to mean something like 'to have a chance of at least 90% of obtaining such a $p$-value' if there really is an improvement from 20 to 30%. This latter statement corresponds, in statistical terms, to saying that the power of the trial should be 0.9 or 90%.

The choice for $\alpha$ is essentially an arbitrary one, the choice being made by the study investigating team. However, practice, accumulated over a long period of time, has established $\alpha = 0.05$ as something of a convention. Thus in the majority of cases, investigators, editors of journals and their readers have become accustomed to anticipate this value. If a different value is chosen then investigators would be advised to explain why.

Convention is not so well established with respect to the size of $\beta$, although in the context of a randomised control trial, to set $\beta > 0.2$, implying a power of less than 80%, would be regarded with some scepticism. Indeed, the use of 90% has become more of the norm (however, see **Chapter 16**, concerned with feasibility studies where the same considerations will not apply). In some circumstances, it may be the type of study to be conducted that determines this choice. Nevertheless, it is the investigating team which has to consider the possibilities and make the final choice.

### Sample Size and Interpretation of Significance

The results of the significance test, calculated on the assumption that the null hypothesis is true, will be expressed as a '$p$-value'. For example, at the end of the trial if the difference between treatments is tested, then a $p$-value $< 0.05$ would indicate that so extreme or greater an observed difference could be expected to have arisen by chance alone less than 5% of the time, and so it is quite likely that a treatment difference really is present.

However, if only a few patients were entered into the trial then, even if there really was a true treatment difference, the results are likely to be less convincing than if a much larger number of patients had been assessed. Thus, the weight of evidence in favour of concluding that there is a treatment effect will be much less in a small trial than in a large one. In statistical terms, we would say that the 'sample size' is too small and that the 'power of the test' is very low.

Suppose the results of an *observed* treatment difference in a clinical trial are declared 'not statistically significant'. Such a statement only indicates that there was insufficient weight of evidence to be able to declare that 'the observed difference is *unlikely* to have arisen by chance'. It does *not* imply that there is 'no clinically important difference between the treatments' as, for example, if the sample size was too small the trial might be very unlikely to obtain a significant $p$-value even when a clinically relevant difference is truly present. Hence, it is of crucial importance to consider sample size and power when interpreting statements about 'non-significant' results. In particular, if the power of the statistical test was very low, all one can conclude from a non-significant result is that the question of treatment differences remains unresolved.

## 1.4 The Normal Distribution

The Normal distribution plays a central role in statistical theory and frequency distributions resembling the Normal distribution form are often observed in practice. Of particular importance is the standardised Normal distribution, which is the Normal distribution that has a mean equal to 0 and a standard deviation (*SD*) equal to 1. The probability density function of such a Normally distributed random variable $z$ is given by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}z^2\right), \tag{1.1}$$

where $\pi$ represents the irrational number 3.14159…. The curve described by equation (1.1) is shown in **Figure 1.1**

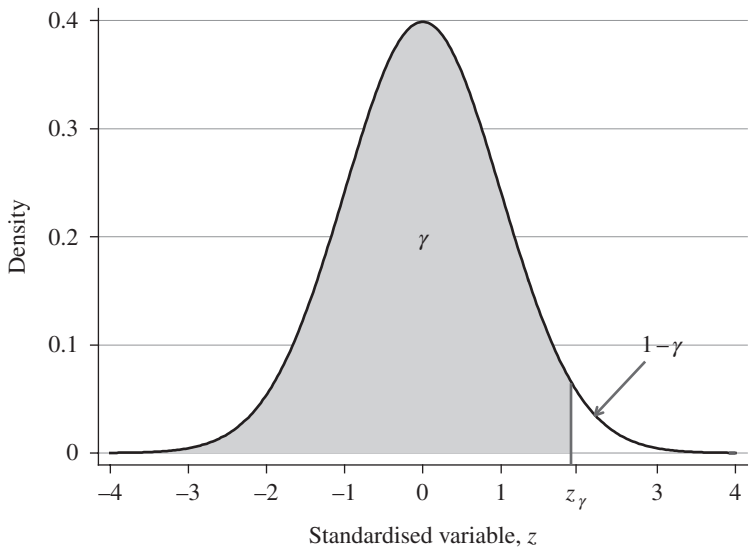


**Figure 1.1** The probability density function of a standardised Normal distribution. (*See insert for color representation of the figure.*)

For sample size purposes, we shall need to calculate the area under some part of this Normal curve. To do this, use is made of the symmetrical nature of the distribution about the mean of 0 and the fact that the total area under a probability density function is unity.

Any shaded area similar to that in **Figure 1.1** which has area γ (here $\gamma \geq 0.5$) has a corresponding value of $z_\gamma$ along the horizontal axis that can be calculated. This may be described in mathematical terms by the following integral:

$$\gamma = \int_{-\infty}^{z_\gamma} \phi(z)dz = \Phi(z_\gamma). \tag{1.2}$$

For areas with $\gamma < 0.5$ we can use the symmetry of the distribution to calculate, in this case, the values for the unshaded area. For example if $\gamma = 0.5$, then one can see from **Figure 1.1** that $z_\gamma = z_{0.5} = 0$. It is also useful to be able to find the value of γ for a given value of $z_\gamma$ and this is tabulated in **Table 1.1**. For example if $z_\gamma = 1.96$ then **Table 1.1** gives $\gamma = 0.97500$. In this case, the shaded area of **Figure 1.1** is then 0.975 and the unshaded area is $1 - 0.975 = 0.025$.

For purposes of sample size estimation, it is the area in the tail, $1 - \gamma$, that is often needed and so we most often need the value of $z$ for a specified area. In relation to test size, we denote the area by $\alpha$ and **Table 1.2** gives the value of $z$ for differing values of $\alpha$. Thus for *1-sided* $\alpha = 0.025$ we have $z = 1.9600$. As a consequence of the symmetry of **Figure 1.1**, if $z = -1.9600$ then $\alpha = 0.025$ is also in the lower tail of the distribution. Hence, the tabular value of $z = 1.9600$ also corresponds to *2-sided* $\alpha = 0.05$. Similarly, **Table 1.2** gives the value of $z$ corresponding to the appropriate area under the curve for one- and two-tailed values of $1 - \beta$.

### The 'Fundamental Equation'

When the outcome variable of a study is continuous and Normally distributed, the mean, $\bar{x}$, and standard deviation, $s$, calculated from the data obtained on $n$ subjects provide estimates of the population mean $\mu_{Pop}$ and standard deviation $\sigma_{Pop}$ respectively. The corresponding standard error of the mean is then estimated by $SE(\bar{x}) = \dfrac{s}{\sqrt{n}}$.

In a parallel group trial to compare two treatments, with $n$ patients in each group, the true relative efficacy of the two treatments is $\delta_{Pop} = \mu_{Pop1} - \mu_{Pop2}$, and this is estimated by $d = \bar{x}_1 - \bar{x}_2$, with standard error $SE(d) = \sqrt{\dfrac{s_1^2}{n} + \dfrac{s_2^2}{n}}$. It is usual to assume that the standard deviations are the same in both groups, so $\sigma_{Pop1} = \sigma_{Pop2} = \sigma_{Pop}$ (say). In which case a pooled estimate obtained from the data of both groups is $s = \sqrt{\dfrac{s_1^2 + s_2^2}{2}}$, so that $SE(d) = \sqrt{\dfrac{s^2}{n} + \dfrac{s^2}{n}} = s\sqrt{\dfrac{2}{n}}$.

The null hypothesis of no difference between groups is expressed as $H_0$: $\delta = \mu_{Pop1} - \mu_{Pop2} = 0$. This corresponds to the left hand Normal distribution of **Figure 1.2** centred on 0. Provided the groups are sufficiently large, then a test of the null hypothesis, $H_0$: $\delta = 0$, of equal means calculates

$$z = \frac{d - 0}{SE(d)} = \frac{d}{s\sqrt{\dfrac{2}{n}}} \tag{1.3}$$

and, for example, if this is sufficiently large, it indicates evidence against the null hypothesis.

Now if this significance test, utilising the data we have collected, is to be *just* significant at some level $\alpha$, then the corresponding value of $z$ is $z_{1-\alpha}$ and that of $d$ is denoted $d_\alpha$. That is, if the observed value $d$ equals or exceeds the critical value $d_\alpha$, then the result is declared statistically significant at significance level $\alpha$.

At the planning stage of the study, when we have no data, we would express the conceptual result of equation (1.3) by

$$z_{1-\alpha} = \frac{d_\alpha}{\sigma\sqrt{\frac{2}{n}}} \quad \text{or} \quad d_\alpha = z_{1-\alpha}\sigma\sqrt{\frac{2}{n}}. \tag{1.4}$$

The alternative hypothesis, $H_{Alt}$: $\delta \neq 0$, where we assume $\delta > 0$ for convenience, corresponds to the right hand Normal distribution of **Figure 1.2** centred on $\delta$. If this were the case then we would expect $d$ to be close to $\delta$, so that $d - \delta$ will be close to zero. To just *reject* the hypothesis that $\delta = \mu_1 - \mu_2 \neq 0$, we require our observed data to provide

$$z = \frac{d-\delta}{SE(d)} = \frac{d-\delta}{s\sqrt{\frac{2}{n}}.} = -z_{1-\beta}\,. \tag{1.5}$$

At the planning stage of the study, when we have no data, we would express this conceptual result by

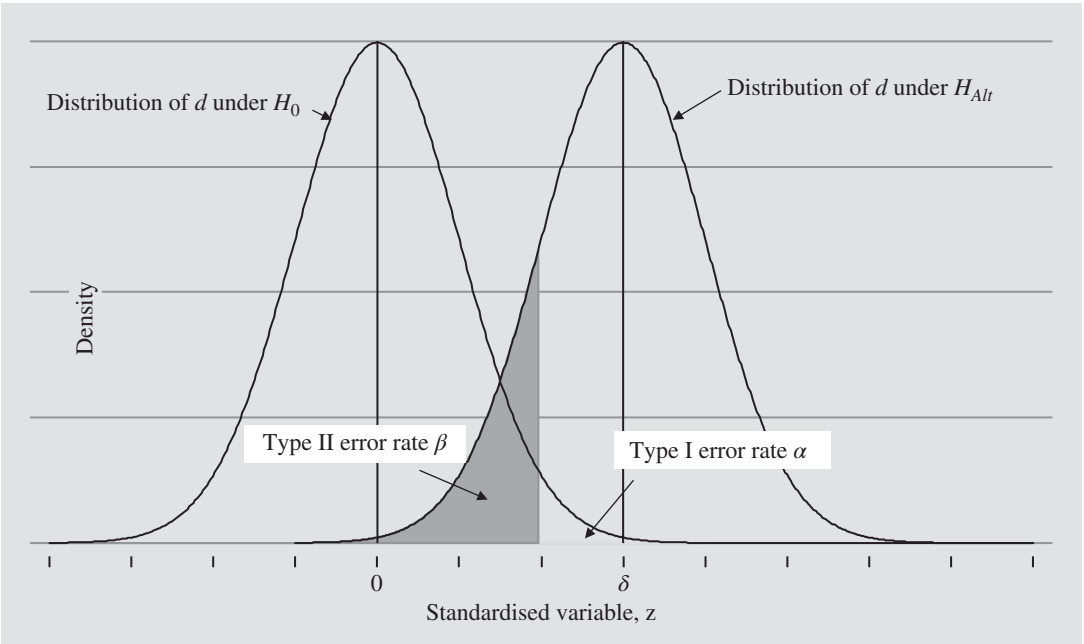


**Figure 1.2** Distribution of $d$ under the null ($\delta = 0$) and alternative hypotheses ($\delta > 0$).

$$d_\alpha = \delta - z_{1-\beta}\sigma\sqrt{\frac{2}{n}}. \tag{1.6}$$

Equating (1.4) and (1.6) for $d_\alpha$, and rearranging, we obtain the total sample size for the trial as

$$N = 2n = \frac{4\left(z_{1-\alpha} + z_{1-\beta}\right)^2}{\left(\delta/\sigma\right)^2} = \frac{4\left(z_{1-\alpha} + z_{1-\beta}\right)^2}{\Delta^2}. \tag{1.7}$$

Here $\Delta = \delta/\sigma$ is termed the standardised effect size. The essential structure of equation (1.7) occurs in many calculations of sample sizes and this is why it is termed the 'Fundamental Equation'.

The use of (1.7) for the case of a *two-tailed* test, rather than the *one-tailed* test discussed previously, involves a slight approximation since $d$ is also statistically significant if it is less than $-d_\alpha$. However, with $d$ positive the associated probability is negligible. Thus, for the more usual situation of a 2-sided test, we simply replace $z_{1-\alpha}$ in (1.7) by $z_{1-\alpha/2}$.

In applications discussed in this book, 2-*sided* $\alpha$ and 1-*sided* $\beta$ correspond to the most frequent application. A 1-sided $\alpha$ and/or 2-sided $\beta$ are used less often (see **Chapter 11** concerned with non-inferiority designs, however).

### Choice of Allocation Ratio

Even though the Fundamental Equation (1.7) has been derived for comparing two groups of equal size, it will be adapted in subsequent chapters to allow for unequal subject numbers in the comparator groups. Thus, for example, although the majority of clinical trials allocate subjects to the two competing interventions on a 1:1 basis, in many other situations there may be different numbers available for each group so that allocation is planned in the ratio 1: $\varphi$ with $\varphi \neq 1$.

If equal allocation is used, then $\varphi = 1$, and so equation (1.7) yields $N_{Equal}$ and hence $n_{Equal} = N_{Equal}/2$ per group. However if $\varphi \neq 1$, then '$2n$' is replaced by '$n + \varphi n$' and the '4' by '$(1+\varphi)^2/\varphi$'. This in turn implies $N_{Unequal} = n_{Equal}(1+\varphi)^2/2\varphi$. The minimum value of the ratio $(1+\varphi)^2/2\varphi$ is 2 when $\varphi = 1$. Hence, $N_{Unequal} > N_{Equal}$ and therefore a study using unequal allocation will require a larger number of subjects to be studied.

*In order to design a study comparing two groups the design team supplies*

- The allocation ratio, $\varphi$
- The *anticipated standardised effect size*, $\Delta$, which is the size of the anticipated difference between the two groups expressed in relation to the *SD*.
- The probability of a Type I error, $\alpha$, of the statistical test to be used in the analysis.
- The probability of a Type II error, $\beta$, equivalently expressed as the power $1 - \beta$.

### Notation

Throughout this book, we denote a 2-*sided* (or two-tailed) value for $z$ corresponding to a 2-sided significance level, $\alpha$, by $z_{1-\alpha/2}$ and for a 1-*sided* significance level by $z_{1-\alpha}$. The same notation is used in respect to the Type II error $\beta$.

### Use of Tables 1.1 and 1.2

#### Table 1.1

**Example 1.1** In retrospectively calculating the power of the test from a completed trial comparing two treatments, an investigator has obtained $z_{1-\beta} = 1.05$ and would like to know the corresponding power, $1 - \beta$.

In the terminology of **Table 1.1**, the investigator needs to find $\gamma$ for $z_\gamma = 1.05$. Direct entry into the table with $z_\gamma = 1.05$ gives the corresponding $\gamma = 0.85314$. Thus, the power of the test would be approximately $1 - \beta = 0.85$ or 85%.

#### Table 1.2

**Example 1.2** At the planning stage of a randomised trial, an investigator is considering using a one-sided or one-tailed test size $\alpha$ of 0.05 and a power of 0.8. What are the values of $z_{1-\alpha}$ and $z_{1-\beta}$ that are needed for the calculations?

For a one-tailed test one requires a probability of $\alpha$ in one tail of the corresponding standardized Normal distribution. The investigator thus needs to find $z_\gamma = z_{1-\alpha}$ or $z_{0.95}$. A value of $\gamma = 0.95$ could be found by searching in the body of **Table 1.1**. Such a search gives $z$ as being between 1.64 and 1.65. However, direct entry into the second column of **Table 1.2** with $\alpha = 0.05$ gives the corresponding $z = 1.6449$. To find $z_{1-\beta}$ for $1 - \beta = 0.80$, enter the second column to obtain $z_{0.80} = 0.8416$.

At a later stage in the planning, the investigator is led to believe that a 2-sided test would be more appropriate; how does this affect the calculations?

For a two-tailed test with $\alpha = 0.05$, direct entry into the second column of **Table 1.2** gives the corresponding $z_{0.975} = 1.9600$.

## 1.5 Distributions

### Central and Non-Central *T*-Distributions

Suppose we had $n$ Normally distributed observations with mean $\bar{x}$ and SD $s$. Then, under the null hypothesis, $H_0$, that the true mean value $\mu = 0$, the function

$$t = \frac{\bar{x} - 0}{s/\sqrt{n}} \tag{1.8}$$

has a Student's $t$-distribution with degrees of freedom ($df$) equal to $n - 1$.

**Figure 1.3** shows how the central $t$-distribution is less peaked, with fatter tails, than the corresponding Normal distribution. However, once the $df$ attains 30, it becomes virtually identical to the Normal distribution in shape.

Values of $t_{df,1-\alpha/2}$ are given in **Table 1.3**. For example if $df = 9$ and 2-sided $\alpha = 0.05$ then $t_{9,0.975} = 2.2622$. As the $df$ increase, the corresponding tabular values decrease until, when $df = \infty$, $t_{9,0.975} = 1.9600$. This is now the same as $z_{0.975} = 1.9600$ found in **Tables 1.1** and **1.2** for the Normal distribution.

Under the alternative hypothesis, $H_{Alt}$, that $\mu \neq 0$, the function

$$t_{Non-Central} = \frac{\bar{x} - \mu}{s/\sqrt{n}} \tag{1.9}$$

has a Non-Central-$t$ (NCT) distribution, with $df = n - 1$ and non-centrality parameter $\psi = \dfrac{\mu\sqrt{n}}{\sigma}$. Thus if $\mu$ and $\sigma$ are fixed, the $\psi$ depends only on the square root of the sample size, $n$.

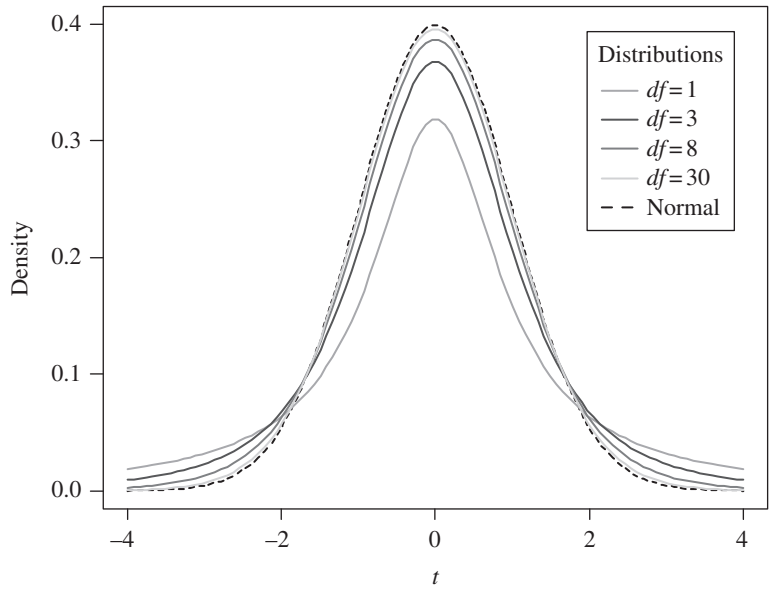


**Figure 1.3** Central *t*-distributions with different degrees of freedom (*df*) and the corresponding Normal distribution. (*See insert for color representation of the figure.*)

**Figure 1.4** shows the distribution of various NCT distributions with $\mu = \sigma = 1$; $df = 1$, 3, 8 and 30, and hence non-centrality parameter $\psi = \sqrt{2}$, $\sqrt{4}$, $\sqrt{9}$ and $\sqrt{31}$ respectively. In general as $\psi$ increases, the mean of the NCT distribution moves away from zero, the SD decreases and so the distribution becomes less skewed. However, as shown in **Figure 1.4**, even with $n = 31$, the NCT distribution is slightly positively skewed relative to the Normal distribution with the same mean and SD.

The cumulative NCT distribution represents the area under the corresponding distribution to the left of the ordinate $x$ and is denoted by $T_{df}(t|\psi)$. However, in contrast to the value of $z_{1-\alpha/2}$ in **Table 1.1**, which depends only on $\alpha$, and $t_{df,1-\alpha/2}$ of **Table 1.3**, which depends on $\alpha$ and $df$, the corresponding $NCT_{1-\alpha/2,\, df,\, \psi}$ varies according to the three components $\alpha$, $df$ and $\psi$ and so the associated tables of values would need to be very extensive. As a consequence, specific computer-based algorithms, rather than tabulations, are used to provide the specific ordinates needed.

### Binomial

In many studies the outcome is a response and the results are expressed as the proportion of subjects who achieve this response. As a consequence, the Binomial distribution plays an important role in the design and analysis of the corresponding trials.

For a specified probability of response $\pi$, the Binomial distribution is the probability of observing exactly $r$ (ranging from 0 to $n$) responses in $n$ patients or

$$b\left(r;\pi,n\right)=\frac{n!}{r!\left(n-r\right)!}\pi^{r}\left(1-\pi\right)^{n-r}. \tag{1.10}$$

Here, for example, $n! = n \times (n-1) \times (n-2) \times \ldots \times 2 \times 1$ and $0! = 1$.

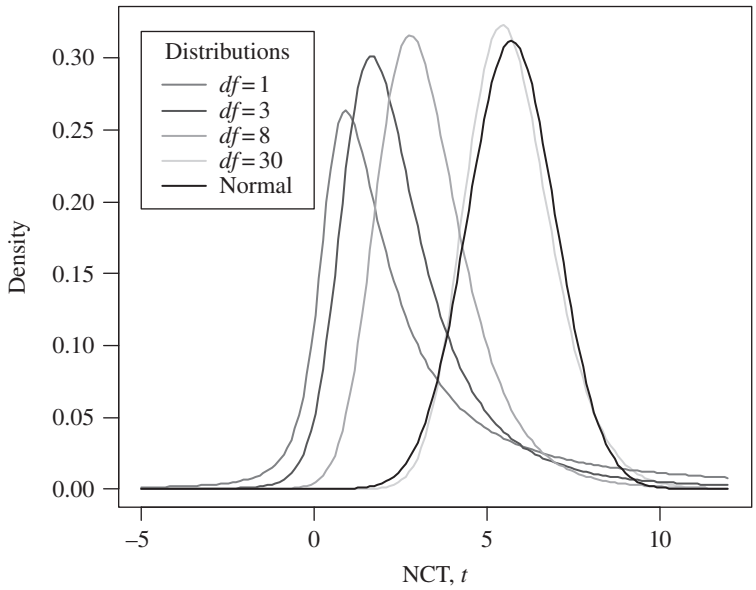


**Figure 1.4** Non-central *t*-distributions with $\mu = \sigma = 1$, hence non-centrality parameters $\psi = \sqrt{n}$, with increasing $df = n - 1$ with $n$ equal to 2, 4, 9 and 31. For $n = 31$ the corresponding Normal distribution with mean $\sqrt{31} = 5.57$ is added. (*See insert for color representation of the figure.*)

For a fixed sample size $n$, the shape of the Binomial distribution depends only on $\pi$. Suppose $n = 5$ patients are to be treated, and it is known that on average 0.25 will respond to this particular treatment. The number of responses actually observed can only take integer values between 0 (no responses) and 5 (all respond). The Binomial distribution for this case is illustrated in **Figure 1.5**. The distribution is not symmetric, it has a maximum at one response, and the height of the blocks corresponds to the probability of obtaining the particular number of responses from the five patients yet to be treated. It should be noted that the mean or expected value for $r$, the number of successes yet to be observed if we treated $n$ patients, is $n\pi$. The potential variation of this expectation is expressed by the corresponding $SD(r) = \sqrt{n\pi(1-\pi)}$.

**Figure 1.5** illustrates the shape of the Binomial distribution for $\pi = 0.25$ and various $n$ values. When $n$ is small (here 5 and 10), the distribution is 'skewed to the right' as the longer tail is on the right side of the peak value. The distribution becomes more symmetrical as the sample size increases (here 20 and 50). We also note that the width of the bars decreases as $n$ increases since the total probability of unity is divided amongst more and more possibilities.

If $\pi$ were set equal to 0.5, then all the distributions corresponding to those of **Figure 1.5** would be symmetrical whatever the size of $n$. On the other hand if $\pi = 0.75$, then all the distributions would be skewed to the left.

The cumulative Binomial distribution is the sum of the probabilities of equation (1.10) from $r = 0$ to a specific value of $r = R$, that is

$$B(R;\pi,n) = \sum_{r=0}^{r=R} \frac{n!}{r!(n-r)!}\pi^r(1-\pi)^{n-r}. \tag{1.11}$$

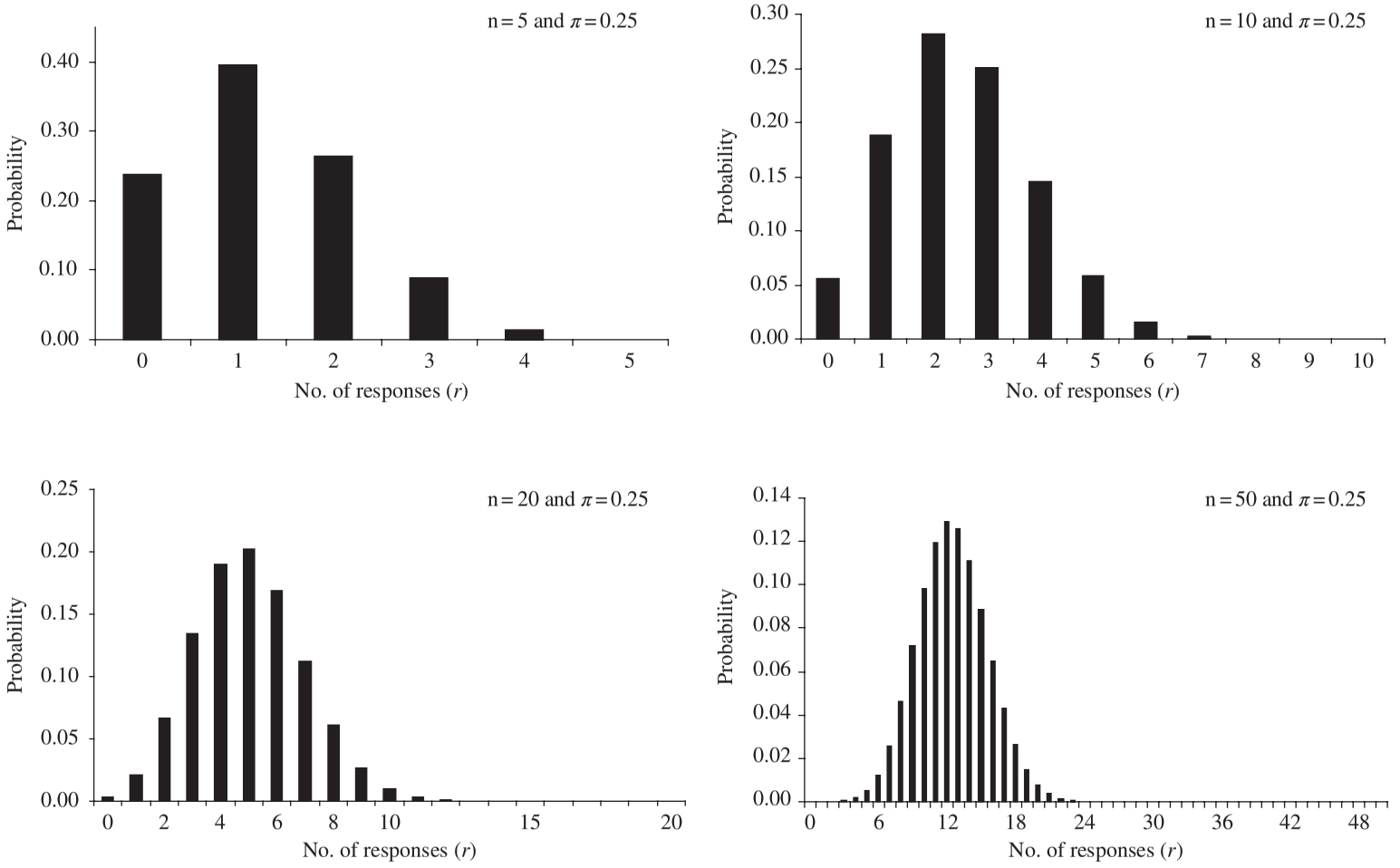


**Figure 1.5** Binomial distribution for $\pi = 0.25$ and various values of $n$ (adapted from Campbell, Machin and Walters, 2007).

The values given to $r$, $R$, $\pi$ and $n$ in expressions (1.10) and (1.11) will depend on the context. This expression corresponds to equation (1.2), and the unshaded area in **Figure 1.1**, of the standardised Normal distribution.

### Poisson

The Poisson distribution is used to describe discrete quantitative data such as counts that occur independently and randomly in time at some average rate. For example, the number of deaths in a town from a particular disease per day and the number of admissions to a particular hospital casualty department typically follow a Poisson distribution.

Suppose events happen randomly and independently in time at a constant rate. If the events happen with a rate of $\lambda$ events per unit time, the probability of $r$ events happening in unit time is

$$Poisson(r) = \frac{exp(-\lambda)\lambda^r}{r!}, \tag{1.12}$$

where $exp(-\lambda)$ is a convenient way of writing the exponential constant $e$ raised to the power $-\lambda$. The constant $e$ is the base of natural logarithms which is 2.718281 ….

The mean of the Poisson distribution for the number of events per unit time is simply the rate, $\lambda$. The variance of the Poisson distribution is also equal to $\lambda$, and so the $SD = \sqrt{\lambda}$.

**Figure 1.6** shows the Poisson distribution for four different means $\lambda = 1$, 4, 10 and 15. For $\lambda = 1$ the distribution is very right skewed, for $\lambda = 4$ the skewness is much less, and as the mean increases to $\lambda = 10$ or 15, the distribution is more symmetrical. These look more like the Binomial distribution of **Figure 1.5** and ultimately the Normal distribution shape of **Figure 1.1**.

### Negative-Binomial

A key property of the Poisson distribution is that the mean and the variance are both equal to $\lambda$. However, there are situations where the mean and variance may be expected to differ. In which case, the Negative-Binomial (NB) distribution which we consider when comparing rates in **Chapter 6** may provide an appropriate description for the resulting data. The distribution is defined by

$$NB(r) = \frac{\Gamma(r + 1/\kappa)}{\Gamma(r+1)\Gamma(1/\kappa)} \frac{(\kappa\lambda)^r}{(1+\kappa\lambda)^{r+1/\kappa}}. \tag{1.13}$$

Here the underlying mean rate is $\lambda$ and the over-dispersion (a variance greater than $\lambda$) is accounted for by the parameter $\kappa$ and implies a variance of $\lambda(1 + \kappa\lambda)$. In equation (1.13) the quantity $\Gamma(r + 1)$ represents the gamma function which, when $r$ is a non-negative integer, equals $r!$. Thus if $r = 3$, $\Gamma(4) = 3 \times 2 \times 1 = 6$ while if $r = 4$, $\Gamma(5) = 4 \times 3 \times 2 \times 1 = 24$. However if, for example, $\kappa = 2$ then $\Gamma(r + ½)$ cannot be expressed as a simple product of successive non-negative integers. In fact if $r = 4$, then $\Gamma(4.5) = \int_0^\infty e^{-u} u^{3.5} du \approx 11.63$. As might be expected, this is somewhere between $3! = 6$ and $4! = 24$.

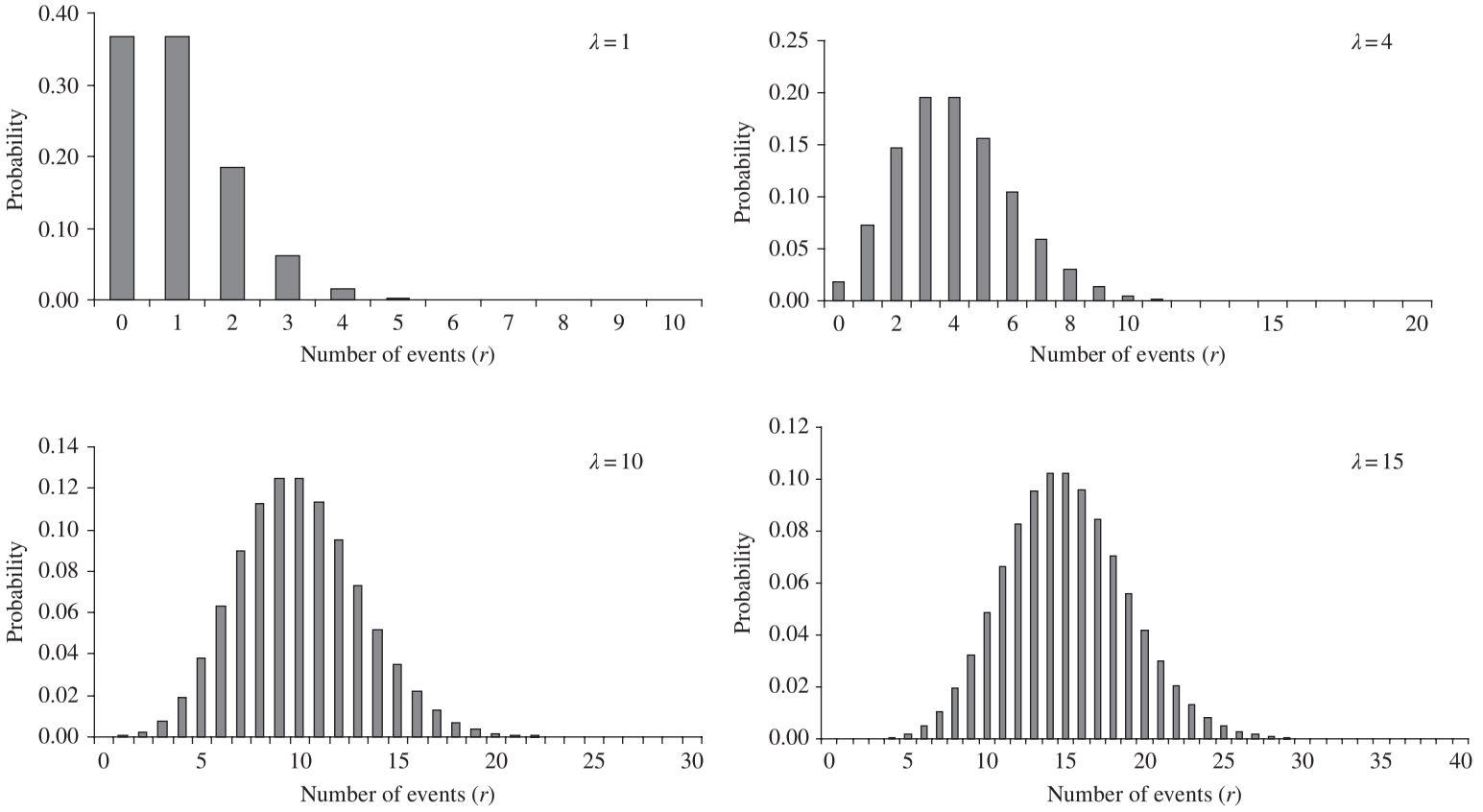


**Figure 1.6** Poisson distribution for various values of $\lambda$ (adapted from Campbell, Machin and Walters, 2007).

### Beta

Another distribution that we will utilise when discussing therapeutic exploratory studies including dose-finding studies and phase II trials in **Chapters 17 and 18** is the Beta distribution. This distribution is similar to the Binomial distribution of equation (1.10) but allows non-integer powers of the terms $\pi$ and $(1 - \pi)$. It takes the form

$$beta(\pi, v, w) = \frac{\pi^{v-1}(1-\pi)^{w-1}}{Beta(v, w)}, \tag{1.14}$$

where $v$ and $w$ are usually > 1 for our purpose and $Beta(v, w) = \int_0^1 u^{v-1}(1-u)^{w-1}\,du$. This integral can be solved numerically for a given $v$ and $w$ and its value ensures that the sum (strictly the integral) of all the terms of (1.14) is unity. In contrast to (1.10) the Beta distribution is that of the continuous variable $\pi$ rather than of the integer $r$ of the Binomial distribution.

In general when planning a study where the outcome of interest is measured as a proportion, the Beta distribution may be used to encapsulate, given our *prior* knowledge about $\pi$, the parameter we are trying to estimate with the trial. This prior knowledge may include relevant information from other sources such as the scientific literature or merely reflect the investigator's belief in the ultimate activity of the therapy under test.

Once trial recruitment is complete and $r$ responses from the $n$ subjects concerned are observed, the prior knowledge is then combined with the study data to obtain a *posterior* distribution for $\pi$. This is formed from the product of parts of equations (1.14) and (1.10), that is, $\pi^{v-1}(1 - \pi)^{1-w} \times \pi^r(1 - \pi)^{n-r} = \pi^{r+v-1}(1 - \pi)^{n-r+1-w}$. The Beta distribution is chosen as it combines easily with the Binomial distribution in this way. The posterior distribution forms the basis of Bayesian methods and represents our overall belief at the close of the trial about the distribution of the population parameter, $\pi$.

Once we have obtained the posterior distribution, we can compute the probability that $\pi$ falls within any pre-specified region of interest. For example, the investigator might wish to know the probability that the true response proportion exceeds a pre-specified target value. This contrasts with the confidence interval approach of **Section** 1.6, which does not answer this question but provides an estimate of the true response proportion, along with the associated 95% confidence interval (termed Frequentist as opposed to Bayesian). Arguably, in the context of early stage trials discussed in **Chapter 17**, since their main goal is not to obtain a precise estimate of the response rate of the new drug but rather to accept or reject the drug for further testing in a randomised controlled trial, a Bayesian approach seems best. However, the majority of studies are not designed using a Bayesian framework.

### Exponential

In survival time studies, such as those describing the subsequent survival experience of a group of patients diagnosed with cancer, if the death rate is constant then the pattern of their deaths follows an Exponential distribution.

If the death rate is $\theta$ per unit time, then the proportion of subjects alive at time $t$ is

$$S(t) = e^{-\theta t}. \tag{1.15}$$

This is often written $S(t) = exp(-\theta t)$ and is termed the survival function of the Exponential distribution. More generally the death rate is replaced by the hazard rate as the event of concern may not be

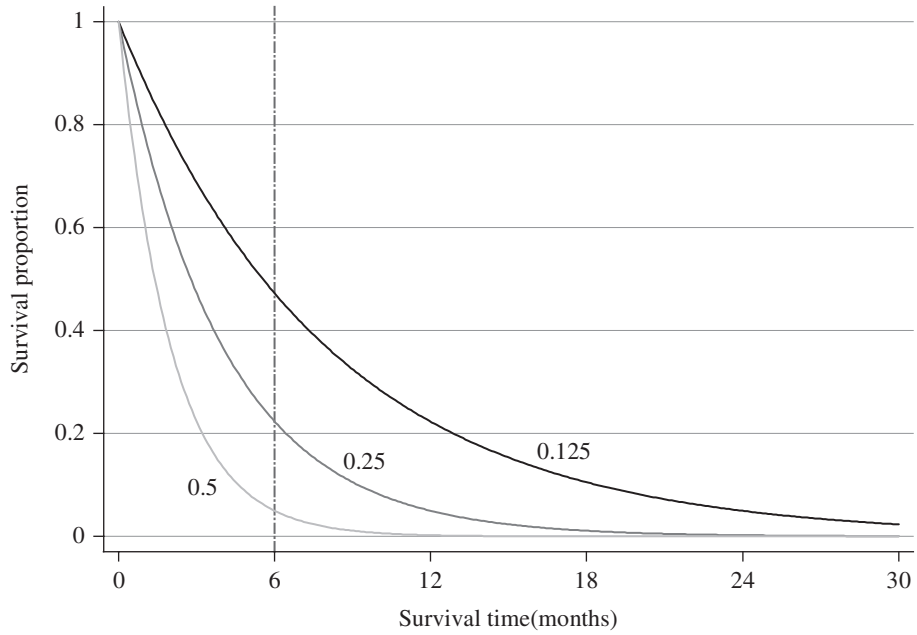


**Figure 1.7** The Exponential survival function with constant hazards of $\theta = 0.125$, 0.25 and 0.5. (*See insert for color representation of the figure.*)

death but (say) time to relapse of a disease or the healing time of an ulcer. The constant hazard rate is a unique property of the Exponential distribution. Sample sizes for survival time studies are discussed in **Chapter 7**.

The shape of the Exponential survival distribution of equation (1.15) is shown in **Figure 1.7** for a hazard rate $\theta = 0.25$ per month. It is clear from this graph that only about 0.2 (20%) of the population remains alive at 6 months, less than 10% at 12 months, and very few survivors beyond 18 months. This is not very surprising since the hazard rate tells us that one-quarter of those alive at a given time will die in the following month.

As **Figure 1.7** also shows, with a hazard rate $\theta = 0.125$ the Exponential survival function will lie above that of $\theta = 0.25$ since the death rate is lower, while for $\theta = 0.5$ it falls below since, in this case, the death rate is higher.

A constant value of the hazard rate implies that the probability of death remains constant as successive days go by. This idea extends to saying that the probability of death in any time interval depends only on the width of the interval. Thus the wider the time interval, the greater the probability of death in that interval, but where the interval begins (and ends) has no influence on the death rate.

## 1.6 Confidence Intervals

When describing the Fundamental Equation (1.7), we have presumed that the study involves two treatment groups and that, once the data are all collated, a statistical significance test of the null hypothesis will be conducted from which a $p$-value will be determined. Whether this is statistically

significant or not, a *p*-value alone gives the reader, who wishes to make use of the published results of a particular study, little practical information. As a consequence, it is therefore incumbent on the investigating team to quote the estimated effect (the observed difference between the treatments) together with an indication of the uncertainty attached to this value by the corresponding (usually 95%) confidence interval (*CI*). Together these enable an interested reader of the final report of the study to better judge the relative impact of the alternative interventions.

Even in situations when no comparison is to be made, for example in estimating the prevalence of a particular disease, it remains important to provide the relevant confidence interval.

In general, for the purposes of this book and at the planning stage of a study, discussion is easier in terms of statistical significance but nevertheless we emphasise that key *CI*s should always be quoted in the final report of any study of whatever design.

In the following sections, we give the expressions for standard errors (*SEs*) and *CI*s for some key summary statistics including the mean, proportion, rate and hazard rate corresponding to data obtained from the Normal, Binomial, Poisson and Exponential distributions. Although not detailed here, there are corresponding *CI*s for the appropriate measures of the difference between groups. *CI*s for some of these latter situations are included in **Chapter 9**.

## Normal

### Confidence interval for a mean

#### *Large samples*

The sample mean, proportion or rate is the best estimate we have of the true population mean, proportion or rate. We know that the distribution of these parameter estimates from many samples of the same size will be more or less Normal. As a consequence, we can construct a *CI*—a range of values in which we are confident the true population value of the parameter is likely to lie. Such an interval for the population mean $\mu_{Pop}$ is defined by

$$\bar{x} - \left[ z_{1-\alpha/2} \times SE\left(\bar{x}\right) \right] \text{ to } \bar{x} + \left[ z_{1-\alpha/2} \times SE\left(\bar{x}\right) \right], \tag{1.16}$$

where $\bar{x}$ is the mean from a sample of $n$ subjects and $SE\left(\bar{x}\right) = \dfrac{\sigma_{Pop}}{\sqrt{n}}$. To calculate the *CI* an estimate, $s$, of the true *SD* $\sigma_{Pop}$ has to be obtained from the data. Values of $z_{1-\alpha/2}$ are found from **Table 1.2**, so that for a 95% *CI*, $\alpha = 0.05$ and we have $z_{0.975} = 1.9600$.

**Example 1.3** Regional brain volumes in extremely preterm infants

Parikh, Kennedy, Lasky, *et al* (2013) report the mean regional brain volume in 21 high-risk ventilator-dependent infants randomised to receive placebo (*P*) as $\bar{x} = 277.8\,\text{cm}^3$ with $SD = 59.1$. Thus $SE\left(\bar{x}\right) = 59.1/\sqrt{21} = 12.90\,\text{cm}^3$. From these the 95% *CI* for the population mean is $277.8 - (1.96 \times 12.90)$ to $277.8 + (1.96 \times 12.90)$ or 252.5 to 303.1 $\text{cm}^3$.

Hence, loosely speaking, we are 95% confident that the true population mean regional brain volume for such preterm infants lies between 253 and 303 $\text{cm}^3$. Our best estimate is provided by the sample mean of 278 $\text{cm}^3$.

Strictly speaking, it is incorrect to say that there is a probability of 0.95 that the population mean birth weight lies between 253 and 303 $\text{cm}^3$ as the population mean is a fixed number and not a random variable and therefore has no probability attached to it. Nevertheless, many statisticians often

describe *CI*s in that way. The value of 0.95 is really the probability that the *CI* calculated from a random sample will include the population value. Thus for 95% of the *CI*s it will be true to say that the population mean, $\mu_{Pop}$, lies within this interval. However we only ever have one CI and we cannot know for certain whether it includes the population value or not.

#### *Small samples*

Equation (1.16) for the 100(1 – $\alpha$)% *CI* for a mean strictly only applies when the sample size is relatively large—a guide is if $n$, the number of subjects contributing to the mean, exceeds 25. When sample sizes are smaller, the following expression should be used instead

$$\bar{x} - \left[ t_{df,1-\alpha/2} \times SE(\bar{x}) \right] \text{ to } \bar{x} + \left[ t_{df,1-\alpha/2} \times SE(\bar{x}) \right]. \tag{1.17}$$

Here $t_{df,1\text{-}\alpha/2}$ replaces $z_{1\text{-}\alpha/2}$ of equation (1.16).

**Degrees of Freedom (df)** Besides depending on $\alpha$, $t_{df,1\text{-}\alpha}$ of equation (1.17) also depends on the degrees of freedom, $df$, utilised to estimate the true standard deviation, $\sigma$, in the final analysis of the study. For a single mean, the $df = n - 1$. Values of $t_{df,1\text{-}\alpha/2}$ are found from **Table 1.3**. For example, for a sample mean based on $n = 10$ observations, $df = 10 - 1 = 9$. The corresponding 95% *CI* has $\alpha = 0.05$ and so $t_{df,1\text{-}\alpha/2} = t_{9,0.975} = 2.2622$, whereas the corresponding $z_{0.975}$ (see the last row of **Table 1.3**) is 1.9600. Thus the small sample leads, for a given $\alpha$, to a wider *CI*.

### Use of Table 1.3

**Example 1.4** Regional brain volumes in extremely preterm infants
In the randomised trial of Parikh, Kennedy, Lasky, *et al* (2013) of *Example 1.3*, the reported mean regional brain volume in those receiving *P* was based on $n = 21$ infants, which is not a very large sample size. Thus it is more appropriate to estimate the *CI* of the mean using the *t*-distribution with $df = 21 - 1 = 20$. For a 95% *CI*, **Table 1.3** gives $t_{df,1\text{-}\alpha/2} = t_{20,0.975} = 2.0860$ so that equation (1.17) leads to $277.8 - (2.0860 \times 12.90)$ to $277.8 + (2.0860 \times 12.90)$ or 250.9 to 304.7 $cm^3$. This *CI* is a little wider than that calculated using the Normal distribution of **Table 1.1**.

### Binomial

#### Confidence interval for a proportion

If $r$ is the number of patients who respond out of $n$ recruited for a trial, then the response proportion $p = r/n$ is the estimate of the true response rate $\pi_{Pop}$. The *SE* of $p$ is $SE(p) = \sqrt{\dfrac{p(1-p)}{n}}$ and the corresponding approximate 100(1 – $\alpha$)% *CI* for $\pi_{Pop}$ is calculated using the '*traditional*' method by analogy with equation (1.16) as

$$p - \left[ z_{1-\alpha/2} \times SE(p) \right] \text{ to } p + \left[ z_{1-\alpha/2} \times SE(p) \right]. \tag{1.18}$$

The reason we can do this is provided by the distributions shown in **Figure 1.5** where, as $n$ gets larger, the shape of the Binomial distribution comes closer and closer to that of the Normal distribution until they are almost indistinguishable. However, this '*traditional*' approximation of equation

(1.16) should not be used if the proportion responding is either very low or very high or if the numbers of patients involved is small. In these cases we advocate the use of the '*recommended*' method described by Newcombe and Altman (2000), see also Julious (2005), and which is computed as follows.

Calculate $p = \frac{r}{n}$, $A = 2r + z^2_{1-\alpha/2}$, $B = z_{1-\alpha/2}\sqrt{z^2_{1-\alpha/2} + 4r\left(1-p\right)}$ and $C = 2\left(n + z^2_{1-\alpha/2}\right)$.

The corresponding 2-sided (1 – α)% *CI* is then given by

$$\frac{\left(A-B\right)}{C} \quad \text{to} \quad \frac{\left(A+B\right)}{C}. \tag{1.19}$$

This method can be used even when no responses occur, that is when $r = 0$, and hence $p = 0$. In which case the *CI* is

$$0 \quad \text{to} \quad \frac{z^2_{1-\alpha/2}}{\left(n + z^2_{1-\alpha/2}\right)}. \tag{1.20}$$

Furthermore, if all patients respond, $r = n$ so that $p = 1$, and the *CI* then becomes

$$\frac{n}{\left(n + z^2_{1-\alpha/2}\right)} \quad \text{to} \quad 1. \tag{1.21}$$

**Example 1.5** Carboplatin for metastatic rhabdomyosarcoma

Chisholm, Machin, McDowell, *et al* (2007, Table 2) reported 1 complete and 4 partial responses among 17 children or adolescents with newly diagnosed metastatic rhabdomyosarcoma who had received carboplatin. The corresponding overall response rate was $p = 5/17 = 0.2941$ with $SE\left(p\right) = \sqrt{\frac{0.2941\left(1-0.2941\right)}{17}} = 0.1105$.

Using the '*traditional*' method of equation (1.16) gives a 95% *CI* for $\pi_{Pop}$ of 0.0775 to 0.5107, whereas using the '*recommended*' method of equation (1.19) results in 0.1328 to 0.5313. These are quite different but only the latter is correct and should be quoted.

As it is usual to quote response rates in percentages, the corresponding trial report would quote for these data: '… the response rate observed was 29% (95% *CI* 13 to 53%).'

## Poisson

### Confidence interval for a rate

If $r$ events are observed in a very large number of $n$ subjects, then the rate is $R = r/n$ as with the Binomial proportion. However, for the Poisson distribution $r$ is small relative to $n$, so the standard error of $R$, is

$$SE\left(R\right) = \sqrt{\frac{R\left(1-R\right)}{n}} \approx \sqrt{\frac{R}{n}}. \tag{1.22}$$

In this case, the approximate 100(1 – $\alpha$)% *CI* for the population value of $\lambda_{Pop}$ is calculated using the '*traditional*' method, by

$$R-\left[z_{1-\alpha/2}\times SE(R)\right] \text{ to } R+\left[z_{1-\alpha/2}\times SE(R)\right]. \tag{1.23}$$

However, although we refer to the number of events as *r*, we also add that this is the number of events observed in a '*unit of time*' and it is therefore essentially a rate. Thus the estimated rate for $\lambda$ is often expressed as $R = r/Y$, where $Y$ is the unit of time concerned. Thus, in *Example 1.6* below, we refer to $R$ as the number of organ donations per day, which is 1.82 as calculated from 1,330 donations over a two year (731 day) period. In such a case $n$, of equation (1.22), is replaced by $Y$.

The reason we can use equation (1.23) is provided by the distributions shown in **Figure 1.6** where, as $\lambda$ gets larger, the shape of the Poisson distribution comes closer and closer to that of the Normal distribution until they are almost indistinguishable. However, this '*traditional*' approximation of equation (1.16) should not be used if $R$ (before division by $n$ or $Y$ as appropriate) is very low or if the numbers of subjects involved small.

**Example 1.6** Cadaveric heart donors

The study of Wight, Jakubovic, Walters, *et al* (2004) gave the number of organ donations calculated over a two-year period as $R = 1{,}330/731 = 1.82$ per day. This is a rate with $SE(R) = \sqrt{\frac{1.82}{731}} = 0.05$.

Therefore, using equation (1.23) the 95% *CI* for $\lambda_{Pop}$ is $1.82 - 1.96\times 0.05$ to $1.82 + 1.96\times 0.05$ or 1.72 to 1.92 organ donations per day. This *CI* is quite narrow, suggesting that the true value of (more strictly the range for) $\lambda_{Pop}$ is well established.

## Exponential

### Confidence interval for a hazard rate

The hazard rate is estimated by $\theta = D/T$ where $D$ is the number of deaths (or events) while $T$ is the total survival experience in (say) years of the $n$ subjects in the study. When $D$ and/or $n$ is large, an approximate 95% *CI* can be obtained from

$$log\ \theta-\left[1.96\times SE\left(log\ \theta\right)\right] \text{ to } log\ \theta+\left[1.96\times SE\left(log\ \theta\right)\right], \tag{1.24}$$

since $log\ \theta$ often follows more closely a Normal distribution than does $\theta$ itself. In this case, $SE(log\theta) = \frac{1}{\sqrt{D}}$.

**Example 1.7** Glioblastoma

Sridhar, Gore, Boiangiu, *et al* (2009) treated 23 patients with non-extensive glioblastoma with concomitant temozolomide and radiation of whom 18 died in a total of 33.32 years of follow-up while 5 patients remain alive and so provided censored observations. The corresponding hazard rate $\theta = 18/33.32 = 0.5402$ per year.

Substituting $\theta = 0.5402$ in equation (1.24) gives $log\ \theta = -0.6158$, $SE(log\theta) = 1/\sqrt{18} = 0.2357$ and the 95% *CI* for $log\ \theta$ as $-0.6158 - (1.96\times 0.2357)$ to $-0.6158 + (1.96\times 0.2357)$ or $-1.0778$ to $-0.1538$. If we exponentiate (anti-log) both limits of this interval, we obtain $exp(-1.0778) = 0.3403$ to $exp(-0.1538) = 0.8574$ per year or 34 to 84% for the 95% *CI* for $\theta$.

## 1.7 Use of Sample Size Tables

### Number of Subjects

Before conducting a clinical trial to test the value of acupuncture, a researcher believes that the placebo group will yield a response rate of 30%. How many subjects are required to demonstrate an anticipated response rate for acupuncture of 50% at a given significance level and power?

### Power of a Study

A common situation is one where the number of patients that can be recruited for a study is governed by forces such as time, money, human resources, disease prevalence or incidence rather than by purely scientific criteria. The researcher may then wish to know 'What is the probability (the power) of detecting the perceived clinically relevant difference in treatment efficacy if a trial of this given size is conducted?'

### Size of Effect

A reasonable power, say 80%, may be fixed and the investigators wish to explore with a particular sample size in mind, what size of effect could be established within this constraint.

## 1.8 Numerical Accuracy

This book contains formulae for sample size determination for many different situations. If these formulae are evaluated with the necessary input values provided, they will give sample sizes to a mathematical accuracy of a single subject. However, the user should be aware that when planning a study of whatever type, the investigators are planning in the presence of considerable uncertainty with respect to the eventual outcome so it is important not to be misled by this apparent precision.

When calculating sample sizes from the formulae given, as well as in the examples and in the statistical software provided, there may be some numerical differences between what is published and what an investigator may obtain in repeating the calculations.

Such divergences may arise in a number of ways. For example, when a particular calculation is performed there is often a choice of the number of significant figures to be used in the calculation process. Although the final sample size, $N$, must be integer, such a choice will in general provide non-integer values which are then rounded (usually upwards) to the nearest integer value. To give an extreme example, the use of two significant figure accuracy for the individual components within a sample size calculation may result in $N = 123.99$, whereas four figure accuracy may lead to $N = 124.0001$. Rounding upwards then introduces a discrepancy between 124 and 125. Further, if a 1:1 allocation to the two groups to be compared is required, then the former gives $n = 62$ per group but the latter gives $n = 62.5$, which would be rounded to 63 and hence an upward revised $N = 126$. Depending on the numerical values derived, further discrepancies can occur in circumstances if these calculations use, for example, 123.99/2 or 124.0001/2 rather than 124/2 and 125/2.

However, since investigators are usually planning in situations of considerable uncertainty, these differences will usually have little practical consequence. Also in view of this, it would seem in general

rather wise to take, for example, a final $N$ of 100 rather than 98, and certainly 1,000 rather than 998. Indeed, although this will require some judgement, perhaps a calculated $N$ of 90 or 950 may become 100 and 1,000 respectively.

This suggests that, in the majority of applications, the number obtained should be rounded upwards to the nearest 5, 10 or more to establish the required sample size. We advise to round upwards as this gives rise to narrower confidence intervals and hence more 'convincing' evidence.

The above comments clearly make sense when the final sample sizes under discussion are relatively large, but more care will be needed in small and particularly very small sized studies. Also, when discussing the cluster designs of **Chapters 12** and **13**, care is needed. In this context, the final sample size is the product of the total number of clusters, $K$, and the number of subjects within each cluster, $m$, so that $N = Km$. Suppose the investigator plans for $m = 45$ per cluster, and the sample size calculations lead to $K = 7.99$ or 8.0001 depending on the number of significant figures used. Then rounding to 8 and 9 respectively and requiring a 1:1 allocation of the intervention to clusters gives $K = 8$ or 10. In which case, automatically rounding 8.0001 upwards to 10 results in 2 extra clusters and hence a study requiring a further 90 subjects. Again, some judgement is now necessary by the investigating team to decide the final choice of study size.

In some cases, statistical research may improve the numerical accuracy of some of the sample size formulae reproduced here which depend on algebraic approximations. However, these improvements are likely to have less effect on the subsequent subject numbers obtained than changes in the planning values substituted into the corresponding formulae.

## 1.9 Software for Sample Size Calculations

Since sample size determination is such a critical part of the design process, we recommend that all calculations are carefully checked before the final decisions are made. This is particularly important for large and/or resource intensive studies. In-house checking by colleagues is important as well.

Sample size calculations for a number of situations are available in various statistical packages such as SAS, SPSS and Stata. They are also available in a number of propriety packages as listed below.

Borenstein M, Rothstein H and Cohen J (2005). *Power & Precision* (*Power Analysis*)*: Version 4.* Biostat, Englewood, New Jersey, USA.

Lenth RV (2006-9). *Java Applets for Power and Sample Size.* http://www.stat.uiowa.edu/~rlenth/Power.

NCSS, LC (2017). *Pass 15 Power Analysis and Sample Size Software* (*PASS 15*): Kaysville, Utah, USA.

SAS Institute (2004). *Getting Started with the SAS Power and Sample Size Application: Version 9.1*, SAS Institute, Cary, North Carolina.

StataCorp (2014). *Stata Statistical Software: Release 14.* College Station, Texas, USA.

Statistical Solutions (2015). *nQuery Adviser + nTerim: Users Guide.* Cork, Ireland.

## References

***Technical***

Campbell MJ, Machin D and Walters SJ (2007). *Medical Statistics: A Textbook for the Health Sciences*, 4th edn. Wiley, Chichester.

Cook JA, Hislop J, Altman DG, Fayers P, Briggs AH, Ramsay CR, Norrie JD, Harvey IM, Buckley B, Fergusson D, Ford I.(2015). Specifying the target difference in the primary outcome for a randomised controlled trial: guidance for researchers. Trials, **16**, 12.

Drummond M and O'Brien B (1993). Clinical importance, statistical significance and the assessment of economic and quality-of-life outcomes. *Health Economics*, **2**, 205–212.

Gandhi M, Tan S-B, Chung AY-F and Machin D (2015). On developing a pragmatic strategy for clinical trials: A case study of hepatocellular carcinoma. *Contemporary Clinical Trials*, **43**, 252–259.

Julious SA (2005). Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in Medicine*, **24**, 3383–3384.

Naylor CD and Llewellyn-Thomas HA (1994). Can there be a more patients-centred approach to determining clinically important effect size for randomized treatments? *Journal of Clinical Epidemiology*, **47**, 787–795.

Newcombe RG and Altman DG (2000). Proportions and their differences. In Altman DG, Machin D, Bryant TN and Gardner MJ (eds). *Statistics with Confidence*. 2nd edn. British Medical Journal Books, London, pp 45–56.

***Regulatory***

CPMP Working Party on Efficacy of Medicinal Products (1995). Biostatistical methodology in clinical trials in applications for marketing authorizations for medical products. *Statistics in Medicine*, **14**, 1659–1682.

FDA (1988). *Guidelines for the Format and Content of the Clinical and Statistics Section of New Drug Applications*. US Department of Health and Human Services, Public Health Service, Food and Drug Administration.

FDA (1996). *Statistical Guidance for Clinical Trials of Non Diagnostic Medical Devices (Appendix VIII)*. US Department of Health and Human Services, Public Health Service, Food and Drug Administration. (http://www.fda.gov/RegulatoryInformation/Guidances/ucm106757.htm)

International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (1998). *ICH Harmonised Tripartite Guideline—Statistical Principles for Clinical Trials E9*. (http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/statistical-principles-for-clinical-trials.html).

***Examples***

Chow PK-H, Tai B-C, Tan C-K, Machin D, Johnson PJ, Khin M-W and Soo K-C (2002). No role for high-dose tamoxifen in the treatment of inoperable hepatocellular carcinoma: An Asia-Pacific double-blind randomised controlled trial. *Hepatology*, **36**, 1221–1226.

Chisholm JC, Machin D, McDowell H, McHugh K, Ellershaw C, Jenney M, Foot ABM (2007). Efficacy of carboplatin in a phase II window study to children and adolescents with newly diagnosed metastatic soft tissue sarcoma. *European Journal of Cancer*, **43**, 2537–2544.

Parikh NA, Kennedy KA, Lasky RE, McDavid GE and Tyson JE (2013). Pilot randomized trial of hydrocortisone in ventilator-dependent extremely preterm infants: effects on regional brain volumes. *The Journal of Pediatrics*, **162**, 685–690.

Sridhar T, Gore A, Boiangiu I, Machin D and Symonds RP (2009). Concomitant (without adjuvant) temozolomide and radiation to treat glioblastoma: A retrospective study. *Journal of Oncology*, **21**, 19–22.

Wight J, Jakubovic M, Walters S, Maheswaran R, White P and Lennon V (2004). Variation in cadaveric organ donor rates in the UK. *Nephrology Dialysis Transplantation*, **19**, 963–968.

**Table 1.1** The cumulative Normal distribution function, Φ(z): The probability that a Normally distributed variable is less than *z* [Equation (1.2)].

| *z* | **0.00** | **0.01** | **0.02** | **0.03** | **0.04** | **0.05** | **0.06** | **0.07** | **0.08** | **0.09** |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.50000 | 0.50399 | 0.50798 | 0.51197 | 0.51595 | 0.51994 | 0.52392 | 0.52790 | 0.53188 | 0.53586 |
| 0.1 | 0.53983 | 0.54380 | 0.54776 | 0.55172 | 0.55567 | 0.55962 | 0.56356 | 0.56749 | 0.57142 | 0.57535 |
| 0.2 | 0.57926 | 0.58317 | 0.58706 | 0.59095 | 0.59483 | 0.59871 | 0.60257 | 0.60642 | 0.61026 | 0.61409 |
| 0.3 | 0.61791 | 0.62172 | 0.62552 | 0.62930 | 0.63307 | 0.63683 | 0.64058 | 0.64431 | 0.64803 | 0.65173 |
| 0.4 | 0.65542 | 0.65910 | 0.66276 | 0.66640 | 0.67003 | 0.67364 | 0.67724 | 0.68082 | 0.68439 | 0.68793 |
| 0.5 | 0.69146 | 0.69497 | 0.69847 | 0.70194 | 0.70540 | 0.70884 | 0.71226 | 0.71566 | 0.71904 | 0.72240 |
| 0.6 | 0.72575 | 0.72907 | 0.73237 | 0.73565 | 0.73891 | 0.74215 | 0.74537 | 0.74857 | 0.75175 | 0.75490 |
| 0.7 | 0.75804 | 0.76115 | 0.76424 | 0.76730 | 0.77035 | 0.77337 | 0.77637 | 0.77935 | 0.78230 | 0.78524 |
| 0.8 | 0.78814 | 0.79103 | 0.79389 | 0.79673 | 0.79955 | 0.80234 | 0.80511 | 0.80785 | 0.81057 | 0.81327 |
| 0.9 | 0.81594 | 0.81859 | 0.82121 | 0.82381 | 0.82639 | 0.82894 | 0.83147 | 0.83398 | 0.83646 | 0.83891 |
| 1.0 | 0.84134 | 0.84375 | 0.84614 | 0.84849 | 0.85083 | 0.85314 | 0.85543 | 0.85769 | 0.85993 | 0.86214 |
| 1.1 | 0.86433 | 0.86650 | 0.86864 | 0.87076 | 0.87286 | 0.87493 | 0.87698 | 0.87900 | 0.88100 | 0.88298 |
| 1.2 | 0.88493 | 0.88686 | 0.88877 | 0.89065 | 0.89251 | 0.89435 | 0.89617 | 0.89796 | 0.89973 | 0.90147 |
| 1.3 | 0.90320 | 0.90490 | 0.90658 | 0.90824 | 0.90988 | 0.91149 | 0.91308 | 0.91466 | 0.91621 | 0.91774 |
| 1.4 | 0.91924 | 0.92073 | 0.92220 | 0.92364 | 0.92507 | 0.92647 | 0.92785 | 0.92922 | 0.93056 | 0.93189 |
| 1.5 | 0.93319 | 0.93448 | 0.93574 | 0.93699 | 0.93822 | 0.93943 | 0.94062 | 0.94179 | 0.94295 | 0.94408 |
| 1.6 | 0.94520 | 0.94630 | 0.94738 | 0.94845 | 0.94950 | 0.95053 | 0.95154 | 0.95254 | 0.95352 | 0.95449 |
| 1.7 | 0.95543 | 0.95637 | 0.95728 | 0.95818 | 0.95907 | 0.95994 | 0.96080 | 0.96164 | 0.96246 | 0.96327 |
| 1.8 | 0.96407 | 0.96485 | 0.96562 | 0.96638 | 0.96712 | 0.96784 | 0.96856 | 0.96926 | 0.96995 | 0.97062 |
| 1.9 | 0.97128 | 0.97193 | 0.97257 | 0.97320 | 0.97381 | 0.97441 | 0.97500 | 0.97558 | 0.97615 | 0.97670 |
| 2.0 | 0.97725 | 0.97778 | 0.97831 | 0.97882 | 0.97932 | 0.97982 | 0.98030 | 0.98077 | 0.98124 | 0.98169 |
| 2.1 | 0.98214 | 0.98257 | 0.98300 | 0.98341 | 0.98382 | 0.98422 | 0.98461 | 0.98500 | 0.98537 | 0.98574 |
| 2.2 | 0.98610 | 0.98645 | 0.98679 | 0.98713 | 0.98745 | 0.98778 | 0.98809 | 0.98840 | 0.98870 | 0.98899 |
| 2.3 | 0.98928 | 0.98956 | 0.98983 | 0.99010 | 0.99036 | 0.99061 | 0.99086 | 0.99111 | 0.99134 | 0.99158 |
| 2.4 | 0.99180 | 0.99202 | 0.99224 | 0.99245 | 0.99266 | 0.99286 | 0.99305 | 0.99324 | 0.99343 | 0.99361 |
| 2.5 | 0.99379 | 0.99396 | 0.99413 | 0.99430 | 0.99446 | 0.99461 | 0.99477 | 0.99492 | 0.99506 | 0.99520 |
| 2.6 | 0.99534 | 0.99547 | 0.99560 | 0.99573 | 0.99585 | 0.99598 | 0.99609 | 0.99621 | 0.99632 | 0.99643 |
| 2.7 | 0.99653 | 0.99664 | 0.99674 | 0.99683 | 0.99693 | 0.99702 | 0.99711 | 0.99720 | 0.99728 | 0.99736 |
| 2.8 | 0.99744 | 0.99752 | 0.99760 | 0.99767 | 0.99774 | 0.99781 | 0.99788 | 0.99795 | 0.99801 | 0.99807 |
| 2.9 | 0.99813 | 0.99819 | 0.99825 | 0.99831 | 0.99836 | 0.99841 | 0.99846 | 0.99851 | 0.99856 | 0.99861 |
| 3.0 | 0.99865 | 0.99869 | 0.99874 | 0.99878 | 0.99882 | 0.99886 | 0.99889 | 0.99893 | 0.99896 | 0.99900 |
| 3.1 | 0.99903 | 0.99906 | 0.99910 | 0.99913 | 0.99916 | 0.99918 | 0.99921 | 0.99924 | 0.99926 | 0.99929 |
| 3.2 | 0.99931 | 0.99934 | 0.99936 | 0.99938 | 0.99940 | 0.99942 | 0.99944 | 0.99946 | 0.99948 | 0.99950 |
| 3.3 | 0.99952 | 0.99953 | 0.99955 | 0.99957 | 0.99958 | 0.99960 | 0.99961 | 0.99962 | 0.99964 | 0.99965 |
| 3.4 | 0.99966 | 0.99968 | 0.99969 | 0.99970 | 0.99971 | 0.99972 | 0.99973 | 0.99974 | 0.99975 | 0.99976 |
| 3.5 | 0.99977 | 0.99978 | 0.99978 | 0.99979 | 0.99980 | 0.99981 | 0.99981 | 0.99982 | 0.99983 | 0.99983 |
| 3.6 | 0.99984 | 0.99985 | 0.99985 | 0.99986 | 0.99986 | 0.99987 | 0.99987 | 0.99988 | 0.99988 | 0.99989 |
| 3.7 | 0.99989 | 0.99990 | 0.99990 | 0.99990 | 0.99991 | 0.99991 | 0.99992 | 0.99992 | 0.99992 | 0.99992 |
| 3.8 | 0.99993 | 0.99993 | 0.99993 | 0.99994 | 0.99994 | 0.99994 | 0.99994 | 0.99995 | 0.99995 | 0.99995 |
| 3.9 | 0.99995 | 0.99995 | 0.99996 | 0.99996 | 0.99996 | 0.99996 | 0.99996 | 0.99996 | 0.99997 | 0.99997 |
| z | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |

**Table 1.2** Percentage points of the Normal distribution for differing α and 1 − β.

| $\alpha$ | | $1-\beta$ | | |
|---|---|---|---|---|
| **1-sided** | **2-sided** | **1-sided** | **2-sided** | **z** |
| 0.0005 | 0.001 | 0.9995 | 0.999 | 3.2905 |
| 0.0025 | 0.005 | 0.9975 | 0.995 | 2.8070 |
| 0.005 | 0.01 | 0.995 | 0.99 | 2.5758 |
| 0.01 | 0.02 | 0.99 | 0.98 | 2.3263 |
| 0.0125 | 0.025 | 0.9875 | 0.975 | 2.2414 |
| 0.025 | 0.05 | 0.975 | 0.95 | 1.9600 |
| 0.05 | 0.1 | 0.95 | 0.9 | 1.6449 |
| 0.1 | 0.2 | 0.9 | 0.8 | 1.2816 |
| 0.15 | 0.3 | 0.85 | 0.7 | 1.0364 |
| 0.2 | 0.4 | 0.8 | 0.6 | 0.8416 |
| 0.25 | 0.5 | 0.75 | 0.5 | 0.6745 |
| 0.3 | 0.6 | 0.7 | 0.4 | 0.5244 |
| 0.35 | 0.7 | 0.65 | 0.3 | 0.3853 |
| 0.4 | 0.8 | 0.6 | 0.2 | 0.2533 |
| 0.45 | 0.9 | 0.55 | 0.1 | 0.1257 |

**Table 1.3** Student's *t*-distribution, $t_{df,1-\alpha/2}$.

| | **2-sided $\alpha$** | | | |
|---|---|---|---|---|
| ***df*** | **0.20** | **0.10** | **0.05** | **0.01** |
| 1 | 3.0777 | 6.3138 | 12.7062 | 63.6567 |
| 2 | 1.8856 | 2.9200 | 4.3027 | 9.9248 |
| 3 | 1.6377 | 2.3534 | 3.1824 | 5.8409 |
| 4 | 1.5332 | 2.1318 | 2.7764 | 4.6041 |
| 5 | 1.4759 | 2.0150 | 2.5706 | 4.0321 |
| 6 | 1.4398 | 1.9432 | 2.4469 | 3.7074 |
| 7 | 1.4149 | 1.8946 | 2.3646 | 3.4995 |
| 8 | 1.3968 | 1.8595 | 2.3060 | 3.3554 |
| 9 | 1.3830 | 1.8331 | 2.2622 | 3.2498 |
| 10 | 1.3722 | 1.8125 | 2.2281 | 3.1693 |
| 11 | 1.3634 | 1.7959 | 2.2010 | 3.1058 |
| 12 | 1.3562 | 1.7823 | 2.1788 | 3.0545 |
| 13 | 1.3502 | 1.7709 | 2.1604 | 3.0123 |
| 14 | 1.3450 | 1.7613 | 2.1448 | 2.9768 |
| 15 | 1.3406 | 1.7531 | 2.1314 | 2.9467 |
| 16 | 1.3368 | 1.7459 | 2.1199 | 2.9208 |
| 17 | 1.3334 | 1.7396 | 2.1098 | 2.8982 |
| 18 | 1.3304 | 1.7341 | 2.1009 | 2.8784 |
| 19 | 1.3277 | 1.7291 | 2.0930 | 2.8609 |

**Table 1.3** (Continued)

| | 2-sided $\alpha$ | | | |
|---|---|---|---|---|
| ***df*** | **0.20** | **0.10** | **0.05** | **0.01** |
| 20 | 1.3253 | 1.7247 | 2.0860 | 2.8453 |
| 21 | 1.3232 | 1.7207 | 2.0796 | 2.8314 |
| 22 | 1.3212 | 1.7171 | 2.0739 | 2.8188 |
| 23 | 1.3195 | 1.7139 | 2.0687 | 2.8073 |
| 24 | 1.3178 | 1.7109 | 2.0639 | 2.7969 |
| 25 | 1.3163 | 1.7081 | 2.0595 | 2.7874 |
| 26 | 1.3150 | 1.7056 | 2.0555 | 2.7787 |
| 27 | 1.3137 | 1.7033 | 2.0518 | 2.7707 |
| 28 | 1.3125 | 1.7011 | 2.0484 | 2.7633 |
| 29 | 1.3114 | 1.6991 | 2.0452 | 2.7564 |
| 30 | 1.3104 | 1.6973 | 2.0423 | 2.7500 |
| 31 | 1.3095 | 1.6955 | 2.0395 | 2.7440 |
| 32 | 1.3086 | 1.6939 | 2.0369 | 2.7385 |
| 33 | 1.3077 | 1.6924 | 2.0345 | 2.7333 |
| 34 | 1.3070 | 1.6909 | 2.0322 | 2.7284 |
| 35 | 1.3062 | 1.6896 | 2.0301 | 2.7238 |
| 36 | 1.3055 | 1.6883 | 2.0281 | 2.7195 |
| 37 | 1.3049 | 1.6871 | 2.0262 | 2.7154 |
| 38 | 1.3042 | 1.6860 | 2.0244 | 2.7116 |
| 39 | 1.3036 | 1.6849 | 2.0227 | 2.7079 |
| 40 | 1.3031 | 1.6839 | 2.0211 | 2.7045 |
| 41 | 1.3025 | 1.6829 | 2.0195 | 2.7012 |
| 42 | 1.3020 | 1.6820 | 2.0181 | 2.6981 |
| 43 | 1.3016 | 1.6811 | 2.0167 | 2.6951 |
| 44 | 1.3011 | 1.6802 | 2.0154 | 2.6923 |
| 45 | 1.3006 | 1.6794 | 2.0141 | 2.6896 |
| ∞ | 1.2816 | 1.6449 | 1.9600 | 2.5759 |