4. Sample Size And Power In Parallel Group Clinical Trials

4.1 Sample size and power

Sample calculation is important for two reasons

- If too few patients are recruited, the trial may lack statistical power, so the study is likely to fail to answer the question it is attempting to address.
- If more patients than the minimum required to answer the question are recruited, some patients may be unnecessarily exposed to an inferior treatment.

As patient recruitment is often difficult, the first reason is generally more important than the second.

Two approaches to sample size in clinical trials

(i) Predetermined trial size

The number of patients to be recruited is fixed before the trial starts.

(ii) Trial size determined by outcome

Statistical analyses, called *interim analyses*, are carried out intermittently as the trial progresses. The trial is stopped if benefit or harm is demonstrated. This design is called an *group-sequential* trial.

Whilst a *group-sequential* trial design is attractive, for several reasons predetermine sample size is much more often used as they are easier to organise and run. In *group-sequential* trials the outcome needs to be determined shortly after recruitment so the interim analysis can be completed. Statistical analysis is also much more complex as it needs to account for multiple statistical testing.

To maintain an overall significance level of α the test size for each test is made smaller, but this is complex as sequential statistical tests are not independent.

4.2 Statistical Power

Suppose the random variable Y_i represents a continuous outcome measure for the i^{th} patients in either the new treatment group (T) or the control group (C). To test for a treatment effect ($\tau \neq 0$) the hypothesis are:

Null hypothesis $\qquad H_0$: $\tau = 0$

Alternate hypothesis $\qquad H_1$: $\tau \neq 0$

If H_0 is rejected, when H_0 is true, a *Type I* or *false positive* error has occurred an

$$\Pr[\textit{Type I error}] = \Pr[\text{ Reject } H_0 \mid H_0] = \alpha = \text{test size}$$

If H_0 not rejected, when H_0 is false, a *Type II* or false negative error has occurred. Define β as the probability of a *Type II* error. This depends on test size α and the magnitude of the effect we wish to detect.

$$\Pr[\textit{Type II error}] = \Pr[\textit{Not reject } H_0 \mid H_1] = \beta(\alpha, \tau)$$

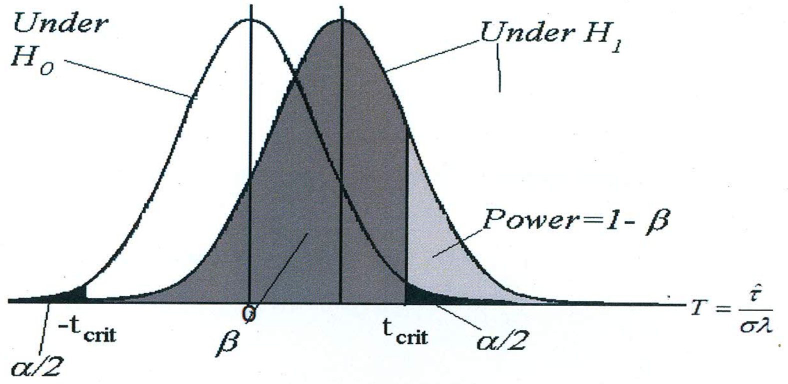
> *Statistical Power* is the probability that a test will detect a difference τ with a significance level α. *Power* $= 1 - \beta(\alpha, \tau)$

Calculation of Power

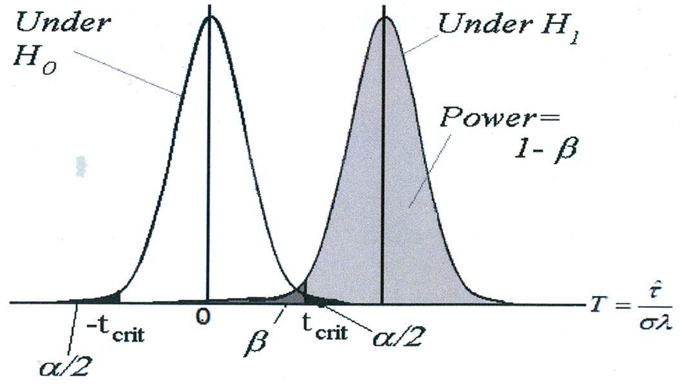
As previously defined the test statistic of a two sample t-test is $T = \dfrac{\hat{\tau}}{\sigma\lambda}$ where $\lambda = \sqrt{1/n_T + 1/n_C}$.

Fig 4.1 Illustration of power calculation for a normally distributed outcome for a two-sided two-sample t-test.

(i) Smaller Sample Size – Low Power



(ii) Larger Sample Size - Increased Power



Under H_1 the test statistic $T = \frac{\hat{\tau}}{\sigma\lambda}$ has the non-central t-distribution.

If F is the cumulative distribution of the non-central t-distribution with $n_T + n_C - 2$ degrees of freedom and non-centrality parameter $\frac{\hat{\tau}}{\sigma\lambda}$, then

RHS Crit. Region LHS Crit Region

$$Power = 1-\beta(\alpha,\tau) = \left(1 - F\left(t_{\alpha/2}(n_T+n_C-2) - \frac{\tau}{\sigma\lambda}\right)\right) + F\left(-t_{\alpha/2}(n_T+n_C-2) - \frac{\tau}{\sigma\lambda}\right)$$

4.3 Sample Size Calculation for Continuous Outcome Measures

Because the central and non-central t-distributions have degrees of freedom determined by sample size there is no closed form formula for sample size assuming this distribution. Instead we shall use the normal distribution as an approximation for the central and non-central t-distributions to get an approximate formula.

For a normally distributed outcome variable the approximate number of subjects required in each of two equal sized groups to have power 1-β to detect a treatment effect τ using a two group t-test with an α two-sided significance level is

$$n = \frac{2\sigma^2}{\tau^2}\left(z_{\alpha/2} + z_\beta\right)^2,$$

where σ is the within group standard deviation.

Assuming n is sufficiently large such that a normal approximation to the central and non-central t-distribution is adequate, the test

statistic T has the standard normal distribution $N[0,1]$ under H_o and $N\left[\frac{\tau}{\sigma\lambda},1\right]$ under H_1. Therefore

Right Tail Left Tail

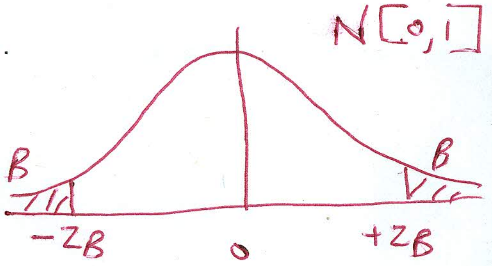
$$Power = 1-\beta = \left(1-\Phi\left(z_{\alpha/2} - \frac{\tau}{\sigma\lambda}\right)\right) + \Phi\left(-z_{\alpha/2} - \frac{\tau}{\sigma\lambda}\right) \quad \textbf{[1]}$$

where $\lambda = \sqrt{1/n_T + 1/n_C}$ and Φ is the cumulative distributionfor $N[0,1]$. The second term on the RHS of equation **[1]** is negligible, for $\tau > 0$ therefore

$$Power = 1-\beta \cong 1-\Phi\left(z_{\alpha/2} - \frac{\tau}{\sigma\lambda}\right).$$

Hence

$$\beta \cong \Phi\left(z_{\alpha/2} - \frac{\tau}{\sigma\lambda}\right).$$



Since $\Phi^{-1}(\beta) = -z_\beta$, it follows that $-z_\beta = z_{\alpha/2} - \frac{\tau}{\sigma\lambda}$

giving $\frac{\tau}{\sigma\lambda} = z_{\alpha/2} + z_\beta$. **[2]**

If equal sized groups are assumed $(n_T = n_C = n)$, then $\lambda = \sqrt{2/n}$.

Substitution into **[2]** gives $\frac{\tau}{\sigma}\sqrt{\frac{n}{2}} = z_{\alpha/2} + z_\beta$.

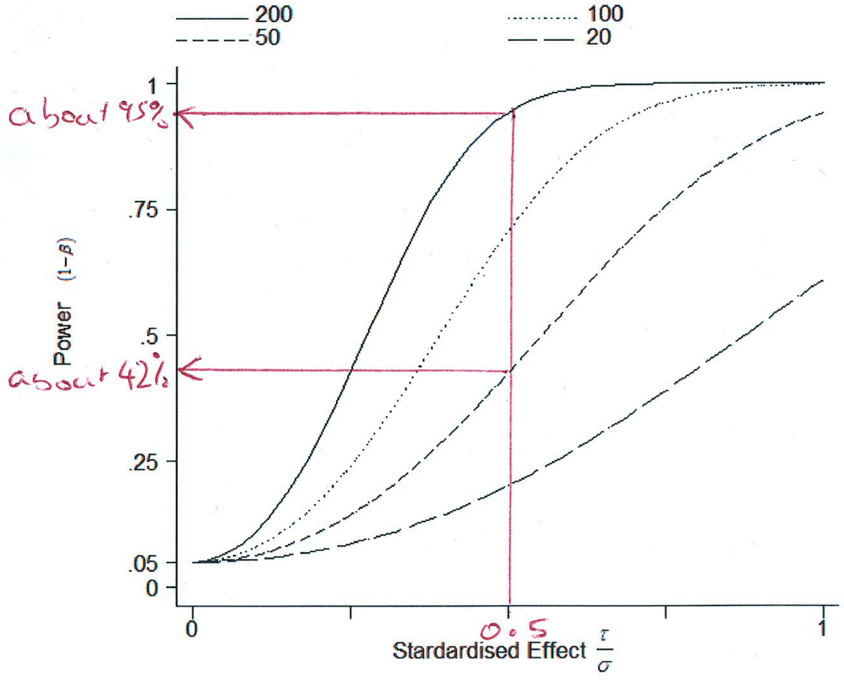
Rearrangement gives $n = \frac{2\sigma^2}{\tau^2}\left(z_{\alpha/2} + z_\beta\right)^2$ as required■

Power

From the above derivation the power of a trial with two groups of size n_T and n_C to detect an treatment effect of magnitude τ using a two group t-test with an α two-sided significance level is

$1-\Phi\left(z_{\alpha/2}-\dfrac{\tau}{\sigma\lambda}\right)$ where Φ is the cumulative distribution function of $N[0,1]$ and $\lambda=\sqrt{1/n_T+1/n_C}$.

Figure 4.2 Plot of Power (1 - β) against standardised effect τ/σ for various total sample sizes for a two sample t-test assuming a 5% significance level and equal size groups.



Ex 4.1 A clinical trial is planned to compare cognitive behavioural therapy (CBT) and a drug therapy for the treatment of depression. The primary outcome measure is the HoNOS scale, which is a measure of impairment due to psychological distress. From published data the *within group standard deviation* of HoNOS is estimated to be 5.7 units.

(i) Calculate the sample size required for each treatment to detect a treatment effect of 2 units on the HoNOS scale with 80% power and a two group t-test with a 0.05 two-sided significance level.

$\tau = 2$ $\qquad\sigma = 5.7$

$\alpha = 0.05$ $\qquad\beta = 1 - \text{power} = 0.2$

$z_{\alpha/2} = 1.960$ $\qquad z_\beta = 0.842$ from tables or page 42

Using the formula, $n=\dfrac{2\sigma^2}{\tau_S^2}\left(z_{\alpha/2}+z_\beta\right)^2 = \dfrac{2\times 5.7^2}{2^2}(1.960+0.842)^2$

Sample size per group = 127.54. Minimum for 80% power is 128

Software assuming t-distribution give sample size per group= 129

(ii) Assuming the same significance level, what power would the study have with only 50 patients into each treatment?

$$Power = 1-\Phi\left(z_{\alpha/2}-\frac{\tau}{\sigma\lambda}\right) = 1-\Phi\left(1.96-\frac{2}{5.7\sqrt{\frac{1}{50}+\frac{1}{50}}}\right)$$

$$= 1-\Phi\left(1.96-\frac{10}{5.7}\right) = 1-\Phi(0.2056)$$

$$\cong 1-0.5815 = 0.4185$$

Power is approx 42%

Note Power determine using statistical software that assumes a t-distribution rather than a normal approximation equals to 0.41.

The Effect of Unequal Randomisation on Sample Size

Suppose the allocation ratio between treatment groups is 1:*k*. i.e. for every patient allocated to one group on average *k* are allocated to the other. For a continuous outcome, it can be shown that the total sample size to give the same power is increased by

$$\frac{(k-1)^2}{4.k}N$$

where *N* is the sample size assuming equal allocation. This result is set as an exercise.

Table 4.1 Increase in total sample size required to maintain power when allocation is unequal

Allocation Ratio	***k***	Percentage Increase in sample size $\frac{(k-1)^2}{4.k}$
3:2	1.5	4.2%
2:1	2	12.5%
3:1	3	33.3%
4:1	4	56.3%

From table 4.1 it can be seen that as the allocation ratio increases the sample size to achieve the same power is increased, but the effect is not great until the allocation ratio exceeds 2:1.

Practical Considerations when Calculating Sample size

- To estimate sample size we need to choose a value of τ. One might take τ to be the minimum difference that is thought to be clinically important, which is called the *minimum clinically important difference* (MCID). Alternatively, one may have an idea of the size of the treatment effect τ and choose that instead

- An estimate of σ is needed to complete the calculation. This is often obtained from previous trials using the outcome in a similar population.

- Power (1-β) = 0.8 or 0.9 and a significance level of 5% are generally used.

- The above formula is for a two-side significance test. The sample size formula for a one-sided test is obtained by replacing α/2 by α in the formulae derived.

- Where a trial compares several outcomes, it is usual to specify one measure as the **primary outcome measure** for which sample size is then determined.

4.4 Sample Size Calculation for Binary Outcome Measures

For a binary outcome measure the approximate number of subjects required in each of two equal sized groups to have power (1-β) to detect a treatment effect $\tau = \pi_T - \pi_c$ using a two sample z-test of proportions with a two-sided significance level α is

$$n = \frac{\left(z_{\alpha/2}\sqrt{2\pi(1-\pi)} + z_\beta\sqrt{\pi_T(1-\pi_T)+\pi_C(1-\pi_C)}\right)^2}{\tau^2}$$

Suppose $p_T = r_T/n_T$ and $p_C = r_C/n_C$ are the observed proportion of successes in each group, the test statistic for the two-tailed hypothesis test of $H_0 : \tau = 0$ vs $H_1 : \tau \neq 0$ is

$$T = \frac{|p_T - p_C|}{\sqrt{(p(1-p))}.\lambda} \quad \text{with } \lambda = \sqrt{1/n_T + 1/n_C} \text{ and } p = \frac{r_T + r_C}{n_T + n_C}.$$

$\hat{\tau} = p_T - p_C$

The distribution of ~~T~~ is approximately $N\left[0, \pi(1-\pi)\left(\frac{1}{n_T}+\frac{1}{n_C}\right)\right]$ under the null hypothesis, with critical values of T are $z_{\alpha/2}$ and $-z_{\alpha/2}$ for an α level two-sided test.

Suppose $\tau = \pi_T - \pi_c$ is the effect under the alternative hypothesis.

Without loss of generality assume that $\tau > 0$. The power $1-\beta(\alpha,\tau)$ equals

$$\Pr\left[p_T - p_C < -z_{\alpha/2}\lambda\sqrt{\pi(1-\pi)}\right] + \Pr\left[p_T - p_C > z_{\alpha/2}\lambda\sqrt{\pi(1-\pi)}\right].$$

See illustration of power calculation on page 61

The distribution of $p_T - p_c$ under the alternative hypothesis is

$$N\left[\tau, \frac{\pi_T(1-\pi_T)}{n_T} + \frac{\pi_C(1-\pi_C)}{n_C}\right].$$

Since $\tau > 0$, $\Pr\left[p_T - p_C < -z_{\alpha/2}\lambda\sqrt{\pi(1-\pi)}\right]$ will be negligible.

Therefore

Power $= 1-\beta(\alpha,\tau) = 1-\Phi\left(\frac{z_{\alpha/2}.\lambda.\sqrt{(\pi(1-\pi))}-\tau}{\sqrt{\frac{\pi_T(1-\pi_T)}{n_T}+\frac{\pi_C(1-\pi_C)}{n_C}}}\right)$ where Φ is the cumulative density function of $N[0,1]$.

Since $\Phi^{-1}(\beta) = -z_\beta$, it follows that

$$-z_\beta = \frac{z_{\alpha/2}.\lambda.\sqrt{(\pi(1-\pi))}-\tau}{\sqrt{\frac{\pi_T(1-\pi_T)}{n_T}+\frac{\pi_C(1-\pi_C)}{n_C}}}.$$

Assuming equal size groups $(n_T = n_C = n)$, then $\lambda = \sqrt{2/n}$.

Rearrangement gives

$$-z_\beta\frac{\sqrt{\pi_T(1-\pi_T)+\pi_C(1-\pi_C)}}{\sqrt{n}} = z_{\alpha/2}\frac{\sqrt{2\pi(1-\pi)}}{\sqrt{n}} - \tau$$

Further rearrangement gives

$$\sqrt{n} = \frac{z_{\alpha/2}\sqrt{2\pi(1-\pi)} + z_\beta\sqrt{\pi_T(1-\pi_T)+\pi_C(1-\pi_C)}}{\tau}$$

so that

$$n = \frac{\left(z_{\alpha/2}\sqrt{2\pi(1-\pi)} + z_\beta\sqrt{\pi_T(1-\pi_T)+\pi_C(1-\pi_C)}\right)^2}{\tau^2}$$

giving the required result ■

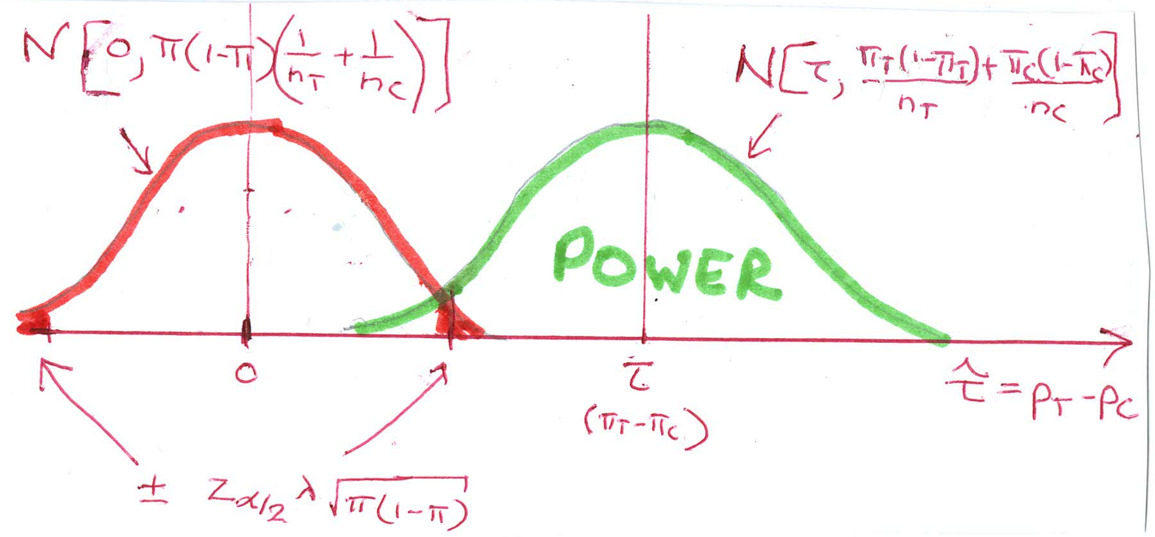
This formula assumes a normal approximation to the binomial i.e. $n.\pi \geq 5, n(1-\pi) \geq 5$. It may be inaccurate if π_T or π_C, close to either 0 or 1.

The power of a trial with two groups of size n_T or n_C to detect a treatment effect $\tau(=\pi_T - \pi_C)$ using a two sample z-test of proportions with an α size two-sided significance level is

$$1-\Phi\left(\frac{z_{\alpha/2}\lambda\sqrt{(\pi(1-\pi))}-\tau_S}{\sqrt{\frac{\pi_T(1-\pi_T)}{n_T}+\frac{\pi_C(1-\pi_C)}{n_C}}}\right)$$

where Φ is the cumulative distribution function $N[0,1]$ and

$\lambda = \sqrt{1/n_T + 1/n_C}$ and $\pi = \dfrac{\pi_T n_T + \pi_C n_C}{n_T + n_C}$.


$N\left[0, \pi(1-\pi)\left(\frac{1}{n_T}+\frac{1}{n_C}\right)\right]$
$N\left[\tau, \frac{\pi_T(1-\pi_T)}{n_T}+\frac{\pi_C(1-\pi_C)}{n_C}\right]$
POWER
0
τ
$(\pi_T-\pi_C)$
$\hat{\tau} = p_T - p_C$
$\pm z_{\alpha/2}\lambda\sqrt{\pi(1-\pi)}$

Ex 4.2 In a placebo controlled clinical trial the placebo response is 0.3 and we expect the response in the drug group to be 0.5. How many subjects are required in each of two equal siz group so that we have an 90% power at a 5% significance level? and a two sided test.

$\pi_T = 0\cdot 5 \qquad \pi_C = 0\cdot 3 \qquad \tau = 0\cdot 5 - 0\cdot 3 = 0\cdot 2$

$\pi = \dfrac{n_T\pi_T + n_C\pi_C}{n_T + n_C} = \dfrac{\pi_T + \pi_C}{2}$ Since $n_T = n_C$

$\sqrt{2\pi(1-\pi)} = \sqrt{2 \times 0\cdot 4 \times 0\cdot 6} = \sqrt{0\cdot 48}$

$\sqrt{\pi_T(1-\pi_T)+\pi_C(1-\pi_C)} = \sqrt{0\cdot 5 \times 0\cdot 5 + 0\cdot 3 \times 0\cdot 7} = \sqrt{0\cdot 46}$

From statistical table $z_{\alpha/2} = z_{0\cdot 025} = 1\cdot 96$ $z_\beta = z_{0\cdot 1} = 1\cdot 282$

$$n = \frac{\left(z_{\alpha/2}\sqrt{2\pi(1-\pi)} + z_\beta\sqrt{\pi_T(1-\pi_T)+\pi_C(1-\pi_C)}\right)^2}{\tau^2}$$

$$= \frac{\left(1\cdot 96 \times \sqrt{0\cdot 48} + 1\cdot 282\sqrt{0\cdot 46}\right)^2}{0\cdot 2^2} = 124\cdot 03$$

Minimum sample size to achieve 90% power is $\underline{\underline{125}}$ per group

4.5 Other Considerations Affecting Trial Size

In many trials it is not possible to follow-up every patient to obtain outcome data. This can be due to many factors such as length of follow-up, commitment of patients, or severity of the condition. In these situations sample size calculation needs to take account of the potential loss of patients to follow-up. Similarly, patient recruitment into a trial may only be a small fraction of the percentage of patients that are potentially eligible.

Ex4.3 The total sample size of patients for a trial has been estimated to be 248 to achieve the required power.

(i) It is thought that outcome data may not be obtained for 15% of patients randomised. How many patients need to be randomised?

85% followed up.

Numbers randomised $= \frac{248}{0.85} = 291.76$

Need to randomise 292 patients

(ii) It is thought that only 20% of patients screened for the trial will be eligible and agree to join the trial. How many patients need to be screened?

Number screened $= \frac{248}{0.85 \times 0.2} = 1458.8$

Need to screen 1459 patients