INSTITUTE OF RESEARCH ADVANCES

IRA-International Journal of Applied Sciences
ISSN 2455-4499 Vol. 03 Issue 02 (May, 2016)
Paper DOI: https://dx.doi.org/10.21013/jas.v3.n2.p5

Sample Size Estimation and Power Analysis for Research Studies Using R

Suma AP[1] , KP Suresh[2]
[1] Research Scholar, Jain University, Jayanagar, 9th Block, Bengaluru, India
[2] Senior Scientist, National Institute of Veterinary Epidemiology and Disease Informatics (NIVEDI), India.

ABSTRACT

Sample size estimation is very crucial in any research design. A research design with less sample size may give a biased result or inconclusive result. A research design with very large sample size than required results is waste of resources, time and energy. So, it is very essential to determine 'ideal' or 'optimum' sample size. This article gives formulae and R code for determining sample size for single mean, two means, single proportion, two proportions, proportion in survey type data, case control studies, cohort studies, correlation coefficient and difference between correlation coefficients.

Key words: Sample size, level of significance, one and two tailed tests, statistical power, Superiority study, Equivalence study, Non inferiority study, R code.

Introduction

The design of research studies is essentially a plan for collecting information. A commonly asked question in any research activity is 'what should be the sample size required for the study?' This is one of the critical research methodological issues to be sorted out at the beginning of the study. Research studies can be classified into surveys, experiments, case-control studies, observational studies etc. They need to be carefully planned to achieve the objective of research study. Planning of a good research has many steps, from defining the problem, defining the population by setting inclusion and exclusion criteria, setting up null and alternative hypothesis, defining the outcome variables, measurement techniques, determining the optimum sample size, randomisation, proper selection of sampling units, data management, use of appropriate statistical methods etc. The basic aim of any research study is to achieve the internal and external validity by managing the confounding bias, selection bias and information bias.

A research study is said to be scientific and systematic if it has good sampling design. A sampling design describes the techniques and procedures adopted to select the sample.

Method of selection of sample forms the primary unit of any research study. Determination of sample size depends on the objective of the research. In other words it depends on the type of inference expected of the research. The sample size is affected more by the variability i.e. standard deviation of the outcome variable, study design and power of research.
In addition, sample size is important for economic reasons, as undersized study can result in waste of resources since it may not produce useful results while oversized study uses more resources than necessary.

Factors that influence the sample size:

- Study design, outcome variable and sample size
- level of significance α
- Power 1-β
- The minimum detectable difference
- The standard deviation

Study design, outcome variable and sample size
Researches of Descriptive type require large sample size. Experimental type of researches requiresmall sample size. Cross over designs require one-fourth of size of the control group. A pre-post type of study in a single group requires half the size of a similar study with control group. A one tailed hypothesis study design requires 20% lesser sample as compared to two tailed hypothesis study design.

Level of significance α

Any research study should include statistical inference. Statistical inference provides tools and techniques required for decision making regarding the characteristics of interest with the help of sample drawn from the population. Statistical inference has two main divisions

- Theory of estimation
- Testing of hypothesis

In theory of estimation, a population parameter is estimated. In testing of hypothesis, population parameter is tested for its validity.
Hypothesis is a statement regarding the parameters of the population. There are two types of hypothesis – Null hypothesis and Alternative hypothesis. The hypothesis which is often being tested for possible rejection is said to be Null hypothesis, denoted by H_0.
Alternative hypothesis is the hypothesis that is accepted when the Null hypothesis is rejected. It is denoted by H_1.
The decision taken to accept or reject the Null hypothesis is based on the inference drawn from the sample drawn from the population. So, if the decisions are not taken properly, errors are committed. The following table-1 explains the errors.

Table 1 (See Tables & Figures Section)

P[committing Type1 error] =P[reject H_0 /H_0 is true] is called Level of significance.

It is denoted by α.

P[committing Type 2 error] =P[accept H_0 /H_0 is not true] =β.
1-β = P[reject H_0 /H_0 is not true] is called POWER of the test which is to be maximized.

Power

Statistical power is the probability of a statistical test indicating significant difference when it is actually true.

One tailed and Two tailed tests

While testing H_0 , if the critical region or the rejection region is considered on one tail of the null distribution of the test statistic, the test is said to be One tailed. On the other hand, if the critical region is considered on both the sides of the null distribution of the test statistic, the test is Two tailed.
$H_1 : \mu_1 > \mu_2$ or $H_1 : \mu_1 < \mu_2$ are One tailed tests, in particular upper tailed and lower tailed tests respectively where as $H_1 : \mu_1 \neq \mu_2$ is a Two tailed test.
If the test is One tailed and is lower tailed, then the critical value is $-Z_\alpha$ and if the test is One tailed and is upper tailed, the critical value is Z_α. If the test is Two tailed, the critical values are ($-Z_{\alpha/2}$, $Z_{\alpha/2}$).
Normal deviates for Type 1 error(α)

Table 2 (See Tables & Figures Section)

Normal deviates for statistical power

Table 3 (See Tables & Figures Section)

The minimum detectable difference

This is the minimum difference between the groups under study that the researcher would be interested in detecting. This is also known as the 'effect size'. The effect size may be determined from pilot studies or prior studies.

The standard deviation

The standard deviation or variance may be obtained from pilot studies or prior studies.

Sample size determination and R code for various study designs.

Any research study can be classified under three categories namely Superiority study, Equivalence study and Non inferiority study.

Superiority study: In this type of study, the researcher tries to establish supremacy of Effect of 'B' over 'A', or may like to find out whether Treatment B is better than Treatment A, or the researcher may like to find out whether new drug is better than the existing one. Here, to determine the sample size, the researcher must decide in advance the margin of difference (d) between 'A' and 'B'.
The sample size is estimated using the formula

$$N = \frac{(Z_\alpha + Z_{1-\beta})^2 s^2}{d^2}$$

Where
Z_α is the normal deviate,
$Z_{1-\beta}$ is the normal deviate at (1-β)% power.
s^2 is the variance and $s^2 = p_1(1-p_1)+p_2(1-p_2)$ where p_1 and p_2 are the proportions of the study of interest of A and B.
d is the margin of difference between A and B

Example 1.In naive cases of chronic hepatitis C genotype 1 pegylated interferon along with ribavirinproduced sustained virologic response in about 40%. Estimate the sample size required to test a new regimen that can increase the sustained response in this type of patients to 60% with a power of 80% at 5% level of significance[17].

Here, p1=0.4 and p2=0.6
$s^2 = p_1(1-p_1)+p_2(1-p_2) = 0.48$

$N = \frac{(1.64+0.84)^2 * 0.48}{0.2^2}$ =74 for each treatment.
R code:

```
Nestimate<-function(Ztab, Zpower, p1,p2, d)
{
s=sqrt((p1*(1-p1))+(p2*(1-p2)))
N=((Ztab+Zpower)^2*s^2)/d^2
return(N)
}
```

For the above example, type the R code in R console/R studio and specify values in the statement
Nestimate(1.64,0.84,0.4,0.6,0.2) to get 'N'.

Equivalence study: In this type of study, the researcher tries to find out if the effect of 'B' is same as that of 'A', i.e. whether the effect of treatment B is the same as that of treatment A, or the effect of new drug is same as that of existing one.Here, to determine the sample size, the researcher must decide the small allowable interval of difference (-d, d) which is not going to affect the quality of A as compared to that of B. The researcher must note that if 'd' is taken as zero, sample size cannot be determined. Here, the sample size is estimated

using the same formula as the superiority trial, except that the test is two tailed and so, Z_α is replaced by $Z_{\alpha/2}$ in the formula. Hence the sample size is estimated using the formula

$$N = \frac{(Z_{\alpha/2}+Z_{1-\beta})^2 s^2}{d^2}$$

Where
$Z_{\alpha/2}$ is the the normal deviate,
$Z_{1-\beta}$ is the normal deviate at (1-β)% power.
s^2 is the variance and $s^2 = p_1(1\text{-}p_1)+p_2(1\text{-}p_2)$ where p_1 and p_2 are the proportions of the study of interest of A and B.
d is the interval of difference between A and B.

Example 2.The therapeutic equivalence between the present regimen of pegylatedinterferonplus ribavirin and a new regimen with fewer side effects needs to be tested in the patients described in example 1 using an RCT. The number of patients necessary for this trial is required to be estimated. The power of the trial should be 80%. The level of significance is 5%. The therapies would be considered equivalent if the confidence interval for the difference in proportion with sustained response falls within the interval ±0.10%[17].

Here, p1=0.4 and p2=0.6
$s^2 = p_1(1\text{-}p_1)+p_2(1\text{-}p_2) = 0.48$

$N = \frac{(1.96+0.84)^2 \ 0.48}{0.1^2}$ =376 for each treatment.

R code:

```
Nestimate<-function(Ztab, Zpower, p1,p2, d)
{
s=sqrt((p1*(1-p1))+(p2*(1-p2)))
N=((Ztab+Zpower)^2*s^2)/d^2
return(N)
}
```

For the above example, type the R code in R console/R studio and specify values in the statement
Nestimate(1.96,0.84,0.4,0.6,0.1) to get 'N'.

Non inferiority study: In this type of study, the researcher tries to explain that effect of 'B' is not worse than 'A' that is new drug is not worse than the existing one. That means to say that the new drug may not be better than the existing one but definitely not inferior to the existing one.Here, the researcher tries to explain that the difference in effect between B and A is not less than '-d' [17].
The test is one tailed. The sample size is estimated using the formula

N = $\frac{(Z_\alpha + Z_{1-\beta})^2 s^2}{d^2}$

Where
Z_α is the normal deviate,
$Z_{1-\beta}$ is the normal deviate at (1-β)% power.
s^2 is the variance and $s^2 = p_1(1-p_1)+p_2(1-p_2)$ where p_1 and p_2 are the proportions of the study of interest of A and B.
d(rather –d) is the value above which the difference in effects should lie.

Example 3.We shall now conduct the trial described in Example 2 as a non- inferiority trial. So the trial should be one-sided. The only difference would be that one should use Z_α instead of $Z_{\alpha/2}$[17].

Here, p1=0.4 and p2=0.6
$s^2 = p_1(1-p_1)+p_2(1-p_2) = 0.48$

N = $\frac{(1.64+0.84)^2 * 0.48}{0.1^2}$ =295 for each treatment.

R code:

```
Nestimate<-function(Ztab, Zpower, p1,p2, d)
{
s=sqrt((p1*(1-p1))+(p2*(1-p2)))
N=((Ztab+Zpower)^2*s^2)/d^2
return(N)
}
```

For the above example, type the R code in R console/R studio and specify values in the statement
Nestimate(1.64,0.84,0.4,0.6,0.1) to get 'N'.

Figure 1 (See Tables & Figures Section)

(From Erik Christensen.Methodology of superiority vs. equivalence trialsand non-inferiority trials.Journal of Hepatology 46 (2007) 947–954.)

Sample size estimation for testing Mean of a population:

Suppose a researcher is interested in testing whether mean of a population is a value μ_0, the sample size can be estimated using the formula
N = $\frac{(Z_{\alpha/2})^2 s^2}{d^2}$, for a Two tailed test

N = $\frac{(Z_\alpha)^2 s^2}{d^2}$, for a One tailed test

Where $Z_{\alpha/2}$ and Z_α are the normal deviates, s is the sample standard deviation and d is the margin of difference from the true mean.

Example : In a study involving mean weight of containers, containing a particular health drink, the researcherwants to estimate the mean weight of the containers and is deciding the margin of difference in weight to be 0.5 kgs, with sample standard deviation of 2 kgs and level of significance 5%, the sample size is estimated as

N = $\frac{(1.96)^2 2^2}{0.5^2}$ =61.5 =62

R code:

```
Nestimate<-function(Ztab, s, d)
{
N=((Ztab)^2*s^2)/d^2
return(N)
}
```

For the above example, type the R code in R console/R studio and specify values in the statement

Nestimate(1.96,2,0.5) to get 'N'.

Sample size estimation for two means:

Suppose a researcher is interested in testing whether means of two populations are the same, H_0 will be $H_0 : \mu_1 = \mu_2$ against the alternative, $H_1 : \mu_1 = \mu_2$+d where d is the difference between the means. The sample size is estimated as

N = $\frac{(r+1)(Z_{\frac{\alpha}{2}}+Z_{1-\beta})^2\sigma^2}{rd^2}$, for a two tailed test

N = $\frac{(r+1)(Z_\alpha+Z_{1-\beta})^2\sigma^2}{rd^2}$, for a one tailed test

Where r=$\frac{n1}{n2}$, n1 and n2 are the sizes of the first sample and the second sample from the first population and the second population respectively and N =n1+n2.

$Z_{\alpha/2}$and$Z_\alpha$are the normal deviate at level of significance α.

$Z_{1-\beta}$is the normal deviate at (1-β)% power.

σ is the common standard deviation.

Example: A researcher gave intelligence test to equal groups of boys and girls and found that the mean marks scored by boys was 80 and that by girls was 85, with a common standard deviation of 7. The total sample size required for the study with α=0.05 and power=90% is

N = $\frac{(1+1)(1.96+1.28)^2 7^2}{1*(85-80)^2}$ = 41.2=41

R code:

```
Nestimate<-function(r,Ztab,Zpower,sigma,d)
{
N=((r+1)*(Ztab+Zpower)^2*sigma^2)/(r*d^2)
return(N)
}
```

For the above example, type the R code in R console/R studio and specify values in the statement

Nestimate(1,1.96,1.28,7,5) to get 'N'.

Sample size estimation for Proportion:

Suppose a researcher is interested in testing for a population proportion P, the sample size can be estimated using the formula

$N = \frac{(Z_{\frac{\alpha}{2}})^2 P(1-P)}{E^2}$, for a two tailed test

$N = \frac{(Z_{\alpha})^2 P(1-P)}{E^2}$, for a one tailed test

Where

$Z_{\alpha/2}$ and Z_{α} are the normal deviate at level of significance α.

P is the proportion of event of interest.

E is the margin of error allowed by the researcher.

Generally, E will be 10% of P.

Example: Past experience shows that 25% girls viewed cricket match. Assuming 10% margin of error, at 5% level of significance, the sample size becomes

$N = \frac{(1.96)^2 0.25(1-0.25)}{(0.1*0.25)^2} = 1152.5 = 1153$

R code:

```
Nestimate<-function(Ztab,P,E)
{
N=((Ztab)^2*P*(1-P))/E^2
return(N)
}
```

For the above example, type the R code in R console/R studio and specify values in the statement

Nestimate(1.96,0.25,0.025) to get 'N'.

Sample size estimation with two Proportions:

Suppose a researcher is interested in testing whether population proportions of two populations are the same, H_0 will be H_0 :P1=P2 against the alternative H_1 :P1=P2+d where d is the difference in proportions. The sample size is estimated using the formula

$N = \frac{(Z_{\frac{\alpha}{2}}\sqrt{2P(1-P)}+Z_{1-\beta}(\sqrt{p1(1-p1)+p2(1-p2)})^2}{(p1-p2)^2}$, for a two tailed test.

$N = \frac{(Z_{\alpha}\sqrt{2P(1-P)}+Z_{1-\beta}(\sqrt{p1(1-p1)+p2(1-p2)})^2}{(p1-p2)^2}$, for a one tailed test.

Where *p1* and *p2* are the sample proportions for sample1 and sample2 drawn from population 1 and population 2 respectively.

P=(p1+p2)/2

$Z_{\alpha/2}$andZ_{α} are the normal deviates at level of significance α.

$Z_{1-\beta}$is the normal deviate at (1-β)% power.

Example:Past experience shows that 15% of products manufactured by company A is defective and 20% of the product manufactured by company B is defective. Determine the sample size required at 5% level of significance and 90% power to test whether there is significant difference in proportion of defectives between the two companies.

$N = \frac{(1.96\sqrt{2(0.175)(1-0.175)}+1.28(\sqrt{0.15(1-0.15)+0.2(1-0.2)})^2}{(0.15-0.2)^2}$=1210

R code:

```
Nestimate<-function(Ztab,Zpower,p1,p2)
{
P=(p1+p2)/2
N=((Ztab*sqrt(2*P*(1-P)))+(Zpower*sqrt(p1*(1-p1)+p2*(1-p2))))^2/(p1-p2)^2
return(N)
}
```

For the above example, type the R code and specify values in the statement Nestimate(1.96,1.28,0.15,0.2) to get 'N'.

Sample size estimation for Proportion in survey type of data:

In survey type of data, sample size depends on the sampling design used. It is '1' for Simple Random Sampling and higher values(1 or 2) for Cluster sampling, Stratified and Systematic sampling. The sample size is estimated using the formula

$N = \frac{(Z_{\frac{\alpha}{2}})^2 P(1-P)*D}{E^2}$

Where D is design effect.

$Z_{\alpha/2}$is the normal deviate at level of significance α.

P is the proportion of event of interest.

E is the margin of error allowed by the researcher.

Generally, E will be 10% of P.

Example:Previous literature gives the estimate of obesity at 20% in the population to be surveyed, assuming 5% level of significance and 10% margin of error, the sample size is calculated as[1]

N = $\frac{(1.96)^2\ 0.20(1-0.20)*1}{(0.1*0.2)^2}$=1536.64=1537

R code:

```
Nestimate<-function(Ztab,P,D,E)
{
N=((Ztab)^2*P*(1-P)*D)/E^2
return(N)
}
```

For the above example, type the R code in R console/R studio and specify values in the statement

Nestimate(1.96,0.2,1,0.02) to get 'N'.

Sample size estimation for Case control studies:

In Case control studies, two independent samples are identified based on a binary(yes/no) response variable. The Odds ratio is a useful measure of association in Case control studies. If P_1 and P_2 are the proportions of cases and controls respectively,

Odds Ratio OR =$\frac{P_1(1-P_2)}{P_2(1-P_1)}$

The sample is estimated using the formula

N= $\frac{(1+r)^2\ (Z_{\frac{\alpha}{2}}+Z_{1-\beta})^2}{r(lnOR)^2[p(1-p)]}$

Where

$Z_{\alpha/2}$is the normal deviate at level of significance α.

$Z_{1-\beta}$is the normal deviate at (1-β)% power.

p is the prevalence of exposure in the general population.

Example:The prevalence of vertebral fracture on a population is 25%.When the study is to estimate the effect of smoking on the fracture, with an odds ratio of 2 and at the significance level of 5% (one sided test) and power of 80% the total sample size for the study of equal sample size can be estimated by[1]

N=$\frac{(1+1)^2(1.64+0.84)^2}{1(ln2)^2[0.25(1-0.25)]}$= 274

R code:

```
Nestimate<-function(r,Ztab,Zpower,Or,p)
{
N=((1+r)^2*(Ztab+Zpower)^2)/(r*(log(Or))^2*(p*(1-p)))
return(N)
}
```

For the above example, type the R code in R console/R studio and specify values in the statement
Nestimate(1,1.64,0.84,2,0.25) to get 'N'.

Sample size estimation for Cohort studies:

In Cohort (Prospective) studies, two samples are identified based on the presence or absence of an explanatory factor. The two samples are observed in future time for the binary response variable. The measures of Relative Risk are useful in Cohort studies. Here, Relative Risk is estimated to within ε of the true population mean[6].
The sample size is estimated using the formula

$$N=\frac{(Z_{\frac{\alpha}{2}})^2[\frac{(1-P_1)}{P_1}+\frac{(1-P_2)}{P_2}]}{[ln(1-\varepsilon)]^2}$$

Where
$Z_{\alpha/2}$ is the normal deviate at level of significance α.
P_1 and P_2 are the proportions of those who develop the disease under exposed and unexposed groups respectively.

Relative Risk $RR=\frac{P_1}{P_2}$

Example:Suppose an outcome is present in 20% of the unexposed group of a cohort study, how large a sample would be needed to estimate the relative risk to within 10% of the true value, which is believed to be around 1.75, with 95% confidence?

Here P_2=0.2
Therefore P_1= RR(P_2)=1.75(0.2)=0.35.

So, $N=\frac{1.96^2[\frac{(1-0.35)}{0.35}+\frac{(1-0.2)}{0.2}]}{[ln(1-0.1)]^2}$ =2027.

R code:

```
Nestimate<-function(Ztab,RR,p2,epsilon)
{
p1=RR*p2
N=((Ztab)^2*(((1-p1)/p1)+((1-p2)/p2)))/((log(1-epsilon))^2)
return(N)
```

}

For the above example, type the R code in R console/R studio and specify values in the statement
Nestimate(1.96,1.75,0.2,0.1) to get 'N'.
If P_1 is available, then write p2=p1/RR instead of p1 within flower brackets and replace p2 with p1 in function command.

Sample size estimation with correlation coefficients:

In a study involving estimation of correlation coefficient 'r', between variables X and Y, for testing the hypothesis $H_0 : r = 0$ against $H_1 : r \neq 0$, the sample size is estimated using the formula

$$N=\frac{\left(\frac{Z_\alpha}{2}+Z_{1-\beta}\right)^2}{\frac{1}{4}[log_e\left(\frac{1+r}{1-r}\right)]}+3$$

Where $Z_{\alpha/2}$ is the normal deviate at level of significance α.
$Z_{1-\beta}$ is the normal deviate at (1-β)% power.
r is the correlation coefficient.
Example: The correlation coefficient between salt intake and systolic blood pressure is 0.3. Estimate the sample size required to test this correlation in a population at 1% level of significance and power 90%.

$$N=\frac{(2.58+1.28)^2}{\frac{1}{4}[log_e\left(\frac{1+0.3}{1-0.3}\right)]}+3 =99$$

R code:

```
Nestimate<-function(Ztab, Zpower ,r)
{
N = (Ztab+Zpower)^2/(0.25*(log((1+r)/(1-r))))+3
return(N)
}
```

For the above example, type the R code in R console/R studio and specify values in the statement
Nestimate(2.58,1.28,0.3) to get 'N'.

Sample size estimation for difference between correlation coefficients:

If the study is involving estimating the difference between two correlation coefficients r_1 and r_2 that are obtained from two independent samples of sizes n_1 and n_2 respectively for testing the hypothesis
$H_0 : r_1 = r_2$ against $H_1 : r_1 \neq r_2$ then the sample size is estimated using the formula

$$N = \frac{4(Z_{\frac{\alpha}{2}} + Z_{1-\beta})^2}{(Z_1 - Z_2)^2}$$

Where

$$Z_1 = \frac{1}{2} log_e \left(\frac{1+r_1}{1-r_1}\right)$$

$$Z_2 = \frac{1}{2} log_e \left(\frac{1+r_2}{1-r_2}\right)$$

Example : Sample size required to estimate the difference between two correlation coefficients 0.4 and 0.8 is at 5% level of significance and 90% power is

$$Z_1 = \frac{1}{2} log_e \left(\frac{1+0.4}{1-0.4}\right) = 0.424$$

$$Z_2 = \frac{1}{2} log_e \left(\frac{1+0.8}{1-0.8}\right) = 1.098$$

$$N = \frac{4(1.96+1.28)^2}{(0.424-1.098)^2} = 92$$

R code:

```
Nestimate<-function(Ztab, Zpower ,r1,r2)
{
Z1=0.5*log((1+r1)/(1-r1))
Z2=0.5*log((1+r2)/(1-r2))
N = 4*(Ztab+Zpower)^2/(Z1-Z2)^2
return(N)
}
```

For the above example, type the R code in R console/R studio and specify values in the statement
Nestimate(1.96,1.28,0.4,0.8) to get 'N'.

Conclusion

A research design with less sample size is unscientific and inconclusive where as a research design with larger than required sample size is unnecessary as it leads to waste of resources, time and energy. So, it is very essential to have optimum sample size. A sampling design which is scientific with the optimum sample size and minimum of 80%

power is highly reliable. This paper gives R code for estimating sample size to facilitate the researcher to maintain high standard of research. Many examples of sample size estimation have beenpresented with the formulae for N and the R code.

References:

1. Suresh KP, Chandrashekara S. Sample size estimation and power analysis for clinical research studies. J Hum ReprodSci 2012;5:7-13.
2. Shusterjj. Handbook of sample size guidelines for clinical trials. Buca Raton, FL: CRC Press; 1990.
3. Altman DC. Practical statistics for Medical Research London. U K: Chapman and Hall. 1991.
4. Wittes J. Sample size calculations for randomized control trials. Epidemiol Rev 2002; 24:39-53.
5. Desu M. Raghavarao D. Sample size methodology. Boston. MA: Academic press. Inc; 1990.
6. LwangaSK ,Lemenshow S. Sample size determination in health srudies. A Practical manual. Geneva- World Health Organization. 1991; P.1-3.
7. Fleiss JL. Statistical methods for rates and proportions. 2nded.New York, Ny:Wiley; 1981;p.45.
8. Hintze JL. Power analysis and sample size system (PASS) for windows User's Guide 1. NCSS. Kaysville, Utah, USA. 2008.
9. Whitley E.Ball J. Statistics review 4: Sample size calculations, Crit Care 2002; 6:335-41.
10. James EB, II Joe WK, Chadwick CH.Organizational research: Determining appropriate sample size in survey research. InfTechnol Learn Performance J 2001; 19-43-50.
11. Johnson, PO. Development of the sample survey as a scientific methodology. J ExpEduc 1959;27:167-76.
12. Wunsch D. Survey research. Determining sample size and representative response. Bus Educ Forum 1986;40:31-4.
13. Lachin JM, Introduction to sample size determination and power analysis for clinical trials. Control Clin Trials 1981;2:93-113.
14. Donner A. Approaches to sample size estimation in the design of clinical trials - A review. Stat Med 198;3:199-214.
15. MarcelWolbers. Basel Institute for Clinical Epidemiology and Biostatistics. Briel, M. et al. Arch Intern Med 2008; 168:2000-2007.
16. Les Huson. Statistical Assessment of Superiority, Equivalence and Non-inferiority in Clinical Trials. Orfocus. July 2001.Number 5.Volume 12.
17. Erik Christensen. Methodology of superiority vs. equivalence trials and non-inferiority trials. Journal of Hepatology. 2007;46. 947–954.
18. Michael Tracy. Methods of Sample Size Calculation for Clinical Trials.theses.gla.ac.uk/671/1/MT_MSC_Feb_2009.pdf
19. Stanley Lemeshow, David W HosmerJr, Janelle Klar, and Stephen K. Lwanga. Adequacy of sample size in health studies. World Health Organization. John Wiley & Sons Ltd. Baffins Lane, Chichester. West Sussex P019 1 UD, England.
20. EmmanuelLesaffre, Dr.Sc. Superiority, Equivalence, and Non-Inferiority Trials. Bulletin of the NYU Hospital for Joint Diseases 2008; 66(2):150-4.

21. Ben EwaldAustralian Prescriber.OCTOBER 2013. Number 5 : Volume 36.
22. Dominik-Non-Inferiority Trials. https://tracs.unc.edu/docs/bios_seminars/pastyear/Dominik-Non-InferiorityTrials.pdf
23. Topic11. Estimation of sample size (1): http://ykhoa.net/baigiang/Topic11.%20Estimation%20of%20sample%20size

Tables & Figures

Actual fact	Decision taken based on the sample	Decision	Error	Notation
H_0 is true	Accept H_0	right	-	-
H_0 is true	Reject H_0	wrong	Type 1	α
H_0 is not true	Accept H_0	wrong	Type 2	β
H_0 is not true	Reject H_0	right	-	1-β

Table 1

α	Z_α (One tailed)	$Z_{\alpha/2}$ (Two tailed)
0.20	0.84	1.28
0.15	1.04	1.44
0.10	1.28	1.64
0.05	1.64	1.96
0.01	2.33	2.58
0.001	3.09	3.29

Table 2

Power	$Z_{1-\beta}$
0.70	0.52
0.75	0.67
0.80	0.84
0.85	1.03
0.90	1.28

Table 3

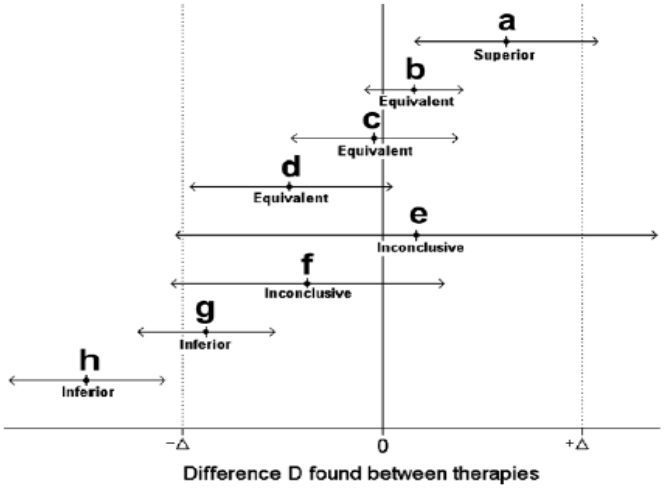

Fig 1