NIHR Clinical Trials Unit Support Funding for Efficient and Innovative Methodologies

The SampSize App to assist in the calculation of sample sizes for clinical trials

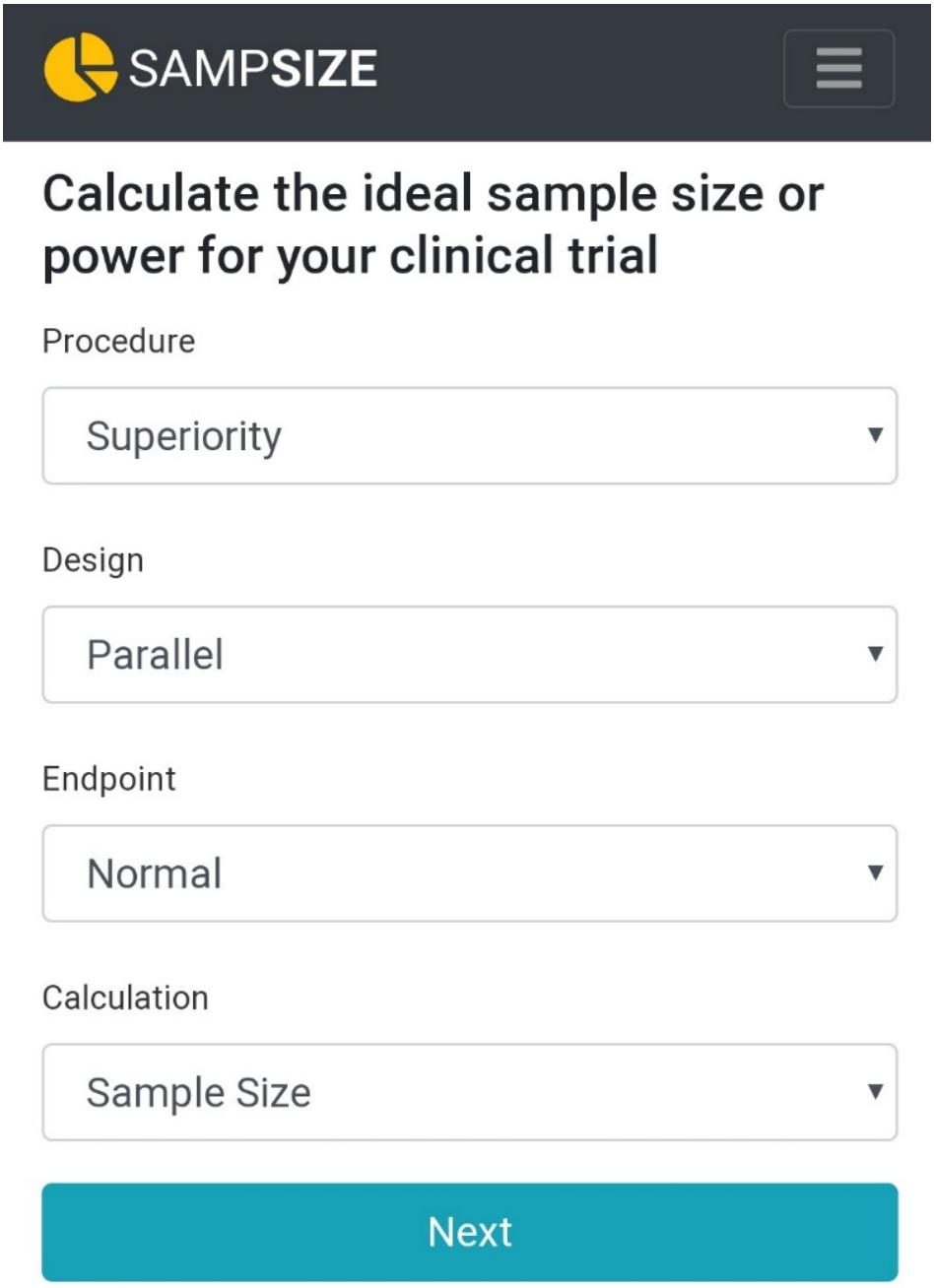
Steven A. Julious

March 2020

Professor Steven A. Julious, Design, Trials and Statistics, School of Health and Related Research (ScHARR, The University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA

Email: S.A.Julious@sheffield.ac.uk

Total Word Count: 3509


SAMPSIZE
Calculate the ideal sample size or power for your clinical trial
Procedure
Superiority
Design
Parallel
Endpoint
Normal
Calculation
Sample Size
Next

Abstract

Background: An important step in the design of a study is the sample size calculation which impacts both on the timelines and the costs of a clinical trial. To facilitate efficient study design it is important therefore to have as an accurate estimate of the sample size as possible.

Methods: The App SampSize has been developed to assist health service researchers in their sample size calculations.

Results: The App does sample size calculations for the study objectives of: superiority; non-inferiority; precision; equivalence and bioequivalence for clinical trials which have Normal; binary or survival endpoint data.

Summary: It is hoped that the App SampSize will help service researchers in the efficient design of clinical trials of new health technologies by assisting them in their estimation of the sample size.

Table of Contents

1 Introduction

An important step in the design in the design of any study is the sample size calculation. Once the sample size is calculated other practical aspects of the study design are considered such as: the number of study centres; the length of the recruitment window and numbers of research staff. These practical aspects subsequently impact on the costs of the study when making a grant application.

It is important therefore to have as accurate estimate of the sample size as possible. Too few patients and resource would have been spent with reduced chance of finding an effect. Whilst with too many patients: patients will be unnecessarily be randomised to a treatment which could already be shown to be effective – or ineffective.

The App SampSize will facilitate efficient study designs by assisting health service researchers in their sample size calculations. It does sample size calculations for different types of endpoint data

- Normal;
- Binary;
- Survival.

For different study objectives:

- Superiority;
- Non-inferiority;
- Equivalence;
- Bioequivalence (for Normal data);
- Precision.

For binary and Normal outcome data the App gives the number patients required per arm. For survival outcomes it gives the number of events

Table 1 details how the App can be obtained and used. The App works on either desktop or mobile devices

Table 1. How to obtain the SampSize App

The sample size App SampSize is available for free at the web site: https://app.sampsize.org.uk/ (date last access 11 Mar 2020) If you access the App through an Android device on accessing the website you will be asked if you wished to install the App. By agreeing SampSize will be installed as an App on your device and could be run without being on a network. If you do not get the query then while on the web site if "Add to Home Screen" the App will be added to your phone and again you will not need to be on a network For an IOS a devices. Tap the Safari icon and then navigate to the website for SampSize given above. After the website has loaded tap on the share icon. The icon is an arrow coming out of a box. Select "Add to Home Screen" to open the Add to Home dialog box. SampSize will look and work like an App on your device and you do not need to have access to a network to run it. If you do not wish to have the App on your phone you do not need to install it as SampSize will work from the web site

2 Practical guide articles

Sample size calculations using the SampSize App have been described in a series of three articles which were published by the journal *Pharmaceutical Statistics* and for which the author submitted version of the papers are freely available on the White Rose Repository which will be highlighted.

The papers describe calculations for the situation with a Normal outcome for trials looking to investigate objectives of superiority, non-inferiority and equivalence with a parallel group design.

2.1 Practical guide to sample size calculations: an introduction

A sample size justification is a vital step when designing any trial. However, estimating the number of participants required to give a meaningful result is not always straightforward. A number of components are required to facilitate a suitable sample size calculation. In this paper, the general steps are summarised for conducting sample size calculations with practical advice and guidance on how to utilise the SampSize app.

Flight L and Julious SA. Practical guide to sample size calculations: an introduction. *Pharmaceutical Statistics* 2016 **15(1)** 75-79

Available on the White Rose Repository at: http://eprints.whiterose.ac.uk/97115/ (date last access 11 Mar 2020)

2.2 Practical guide to sample size calculations: superiority trials

In this paper, the steps for conducting sample size calculations for superiority trials are summarised. Practical advice and examples are provided illustrating how to carry out the calculations by hand and using the SampSize app.

Flight L and Julious SA. Practical guide to sample size calculations: superiority trials. *Pharmaceutical Statistics* 2016 **15(1):** 80-89

Available on the White Rose Repository at: http://eprints.whiterose.ac.uk/97114/ (date last access 11 Mar 2020)

2.3 Practical guide to sample size calculations: non-inferiority and equivalence trials

In this paper, the steps for conducting sample size calculations for non-inferiority and equivalence trials are summarised. Practical advice and examples are provided that illustrate how to carry out the calculations by hand and using the SampSize app.

Flight L and Julious SA. Practical guide to sample size calculations: non-inferiority and equivalence trials. *Pharmaceutical Statistics* 2016:**15(1)** 68-74

Available on the White Rose Repository at: http://eprints.whiterose.ac.uk/97113/ (date last access 11 Mar 2020)

3 Description of the sample size methods by data type

The App SampSize uses the methods for binary and Normal data from two journal articles in the journal *Statistics in Medicine.* The author submitted version of both of these papers are freely available on the White Rose Repository which will be highlighted.

For trials with a survival endpoint the methods are introduced for studies where the objective is to demonstrate superiority. More detailed calculations are given in the appendix for other study objectives.

3.1 Sample sizes for clinical trials with Normal data

A detailed description of the methods for Normal outcomes can be found in a Tutorial in Biostatistics paper in the journal *Statistics in Medicine*. The article gives an overview of sample size calculations for parallel group and cross-over studies where the primary outcome is anticipated to take a Normal form. Sample size derivation is given for trials where the objective is to demonstrate: superiority, equivalence, non-inferiority, bioequivalence and estimation to a given precision, for different types I and II errors. It is demonstrated how the different trial objectives influence the null and alternative hypotheses of the trials and how these hypotheses influence the calculations. Sample size tables for the different types of trials and worked examples are given.

The methods in this tutorial article were uses as the basis for the sample size calculation for Normal data in the SampSize App

Julious SA. Tutorial in Biostatistics: Sample Sizes for clinical trials with Normal Data. *Statistics in Medicine* 2004 **23:**1921-86

Available on the White Rose Repository at: http://eprints.whiterose.ac.uk/145474/ (date last access 11 Mar 2020)

3.2 Sample sizes for parallel group clinical trials with binary data

For trials with Binary outcomes methods are described in a second Tutorial in Biostatistics paper in the journal *Statistics in Medicine*. This article gives an overview of sample size calculations for a single response and a comparison of two responses in a parallel group trial where the outcome is binary. Sample size derivation is given for trials where the objective is to demonstrate: superiority, equivalence, non-inferiority and estimation to a given precision. For each type of trial the null and alternative hypotheses are described and how the impact these have on the sample size calculations. Also, for each type of trial the calculations are highlighted through worked examples. Sample size tables for the different types of trials and worked examples are given to assist in future calculations

The methods in this tutorial article were uses as the basis for the sample size calculation for binary data in the SampSize App

Julious SA and Campbell MJ. Tutorial in Biostatistics: Sample Sizes for clinical trials with binary data. *Statistics in Medicine* 2012;**31**:2904–36

Available on the White Rose Repository at: http://eprints.whiterose.ac.uk/145472/ (date last access 11 Feb 2020)

3.3 Sample sizes for parallel group clinical trials with a survival endpoint

At the moment there is no tutorial article describing the sample size calculations for trials with a superiority outcome

When calculating the sample sizes at the simplest level the calculations described for binary endpoints could be used. However, this approach would ignore the survival times. A more plausible approach would be to use the methodologies for Normal data for the (probably logged) survival times. However, this approach would ignore the censored subjects meaning the sample size would be just for the number of events and not the total sample size.

A common method for estimating the sample size for a survival endpoint is to assume that we have exponential survival. Under this assumption if we let T be the survival time random variable such that for treatment A we have

(1) $$S(t) = P(T \geq t) = e^{-\lambda_A t},$$

where λ_A is constant and does not change with t.

Suppose the survival distributions for the two arms of the trial have instantaneous death rates of λ_A for treatment A and λ_B for treatment B. From this the hazard ratio (HR) is defined as

(2) $$HR = \lambda_A/\lambda_B.$$

From (1) we also get

(3) $$M_A = \log_e 2/\lambda_A.$$

A similar result for M_B can be derived for where λ_B. Assuming an exponential survival an alternative formula for the Hazard Ratio is to derive it in terms of the median survival terms for each treatment

(4) $$HR = \frac{M_B}{M_A}.$$

If the hazard ratio does not change with time, then it can also be estimated by

(5) $$HR = \frac{\log \pi_A}{\log \pi_B},$$

where π_A and π_B are two survival rates at some fixed time poin

With an estimate of the Hazard Ration the number of events, *E*, required in each patient group can be estimated from

(6) $$E = \frac{2(Z_{1-\alpha/2}+Z_{1-\beta})^2}{(\log HR)^2}.$$

Sample sizes from (6) are given in Table 1 for different hazard ratios

The result (6) is implemented in the app SampSize for superiority sample sizes for superiority trials with survival outcome.

The appendix details the other methods use in the SampSize App for trials with a survival outcome where the objective is to demonstrate: non-inferiority, equivalence or precision around the outcome. The results are also described by Julious [7]

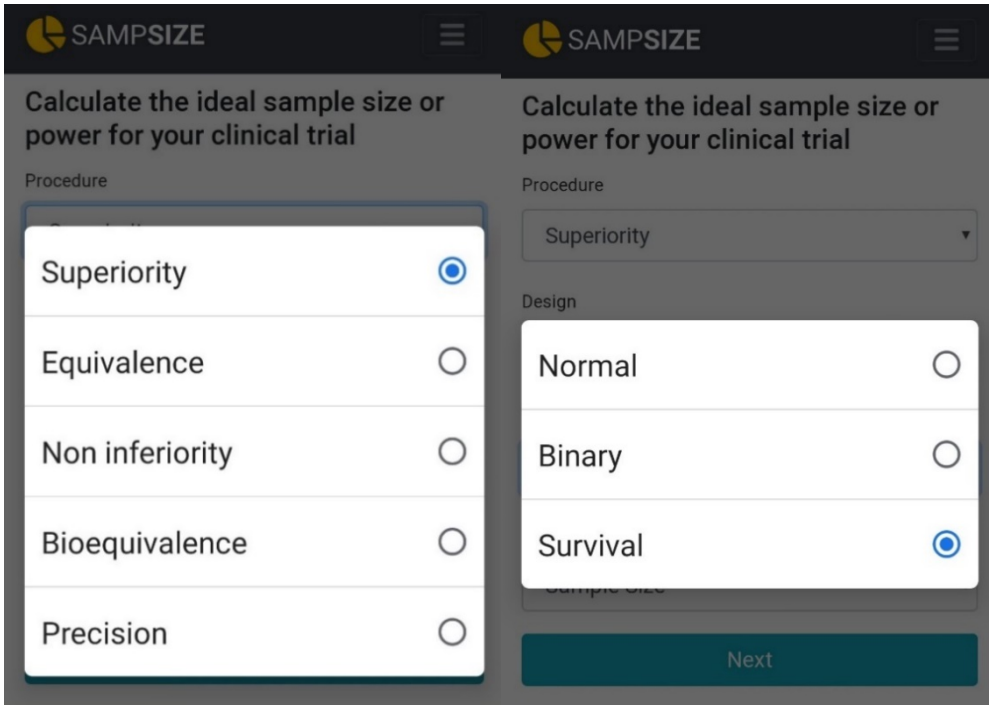
Table 1. Number of events for different hazard ratios for a two sided 5% significance level and 90% power

Hazard Ratio	Number of Events
0.6	81
0.7	166
0.8	423
0.9	1894
1.1	2314
1.2	633
1.3	306
1.4	186
1.5	128
1.6	96
1.7	75
1.8	61
1.9	52
2.0	44

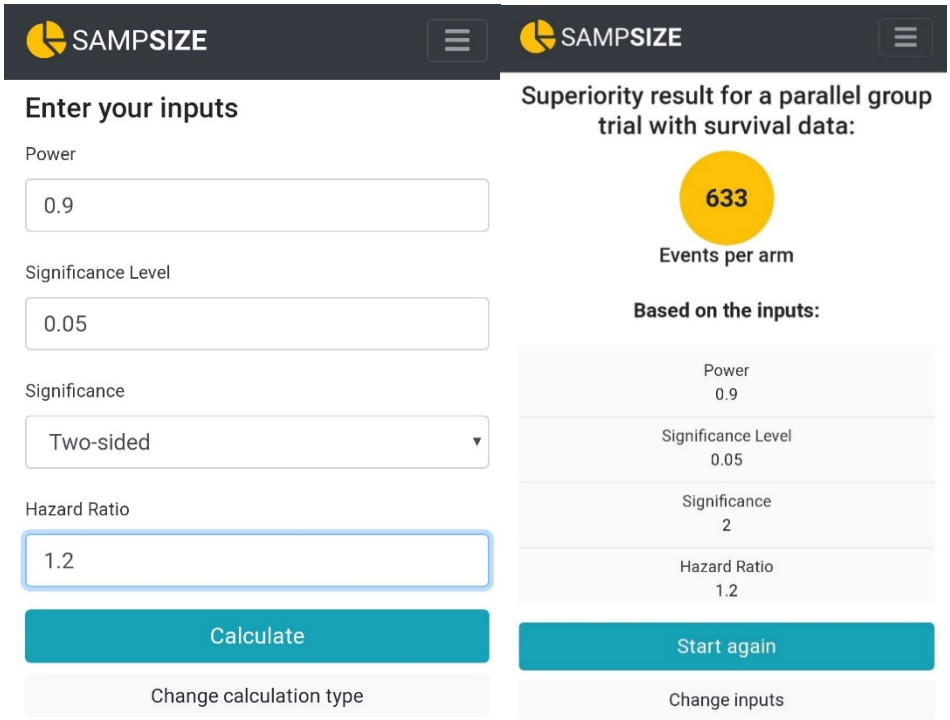
4 Worked Example with Using SampSize

Suppose we wish to design a study investigating a new investigative treatment against control where the primary endpoint is progression free survival and a hazard ratio is the effect size of interest. The target effect size of interest is a hazard ratio of 1.2 against the control treatment (or 0.83 in favour of the investigative treatment). For a two tailed level of significance of 5% and 90% power the sample size be (in terms of number of events) assuming exponential survival would be 633 events per arm (from (14)).

To repeat the calculations in SampSize, select the options for Superiority and Survival as below.



The inputs would be as per the left panel below.



SampSize gives 633 events per arm as the sample size which is the same as using Table 1 in the worked example earlier.

5 Guidance on effect sizes

Randomised controlled trials are considered to be the best method to assess comparative clinical efficacy and effectiveness, and can be a key source of data for estimating cost effectiveness. Central to the design of a randomised controlled trial is an *a priori* sample size calculation, which ensures that the study has a high probability of achieving its pre-specified main objective. There is the DELTA^2 guidance to assist in determining the target difference and sample size calculation for randomised controlled trials. Recommendations for the subsequent reporting of the sample size calculation are also provided in the paper.

Cook JA, Julious SA, Sones W, Hampson LV, Hewitt C, Berlin JA, et al. DELTA2 guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. *BMJ* 2018;**363:**k3750 http://dx.doi.org/10.1136/bmj.k3750

The paper is available at: https://www.bmj.com/content/363/bmj.k3750 (date last access 11 Mar 2020)

It is also available on the White Rose Repository at: http://eprints.whiterose.ac.uk/139500/ (date last access 11 Mar 2020)

The work is also detailed in a Health Technology Assessment monograph

Cook JA, Julious SA, Sones W, Hampson LV, Hewitt C, Berlin JA, Ashby D, Emsley R, Fergusson DA, Walters SJ et al. Practical help for specifying the target difference in sample size calculations for RCTs : the DELTA2 five-stage study, including a workshop Health Technology Assessment 2019 2**3(60):**1-88 01

Available at: https://www.journalslibrary.nihr.ac.uk/hta/hta23600#/abstract last access 11 Mar 2020)

Available on the White Rose Repository at: http://eprints.whiterose.ac.uk/152823/ (date last access 11 Mar 2020)

6 Sample sizes for pilot trials

Although not covered in the App a sample size justification for a pilot trial is important.

When the outcome is a continuous variable, the sample size calculation requires an accurate estimate of the standard deviation of the outcome measure. A pilot trial can be used to get an estimate of the standard deviation, which could then be used to anticipate what may be observed in the main trial. The paper by Whitehead *et al* looks at how you

can choose an external pilot trial sample size in order to minimise the sample size of the overall clinical trial programme, that is, the pilot and the main trial together.

Whitehead AL, Julious SA, Cooper CL and Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. *Statistical Methods in Medical Research* 2016 **25(3)** 1057-1073 (DOI: 10.1177/0962280215588241)

Available at: https://journals.sagepub.com/doi/full/10.1177/0962280215588241 last access 11 Mar 2020)

Available on the White Rose Repository at: http://eprints.whiterose.ac.uk/87982/ (date last access 11 Mar 2020)

7 Summary

It is hoped that the App SampSize will help service researchers in the efficient design of clinical trials of new health technologies by assisting them in their estimation of the sample size.

8 Acknowledgments

The App SampSize was developed with a assistance of a Crowdfunding appeal and the support of NIHR CTU Support Funding Supporting efficient / innovative delivery of NIHR research. We are grateful for all the support

9 References

1. Cook JA, Julious SA, Sones W, Hampson LV, Hewitt C, Berlin JA, et al. DELTA2 guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. *BMJ* 2018;**363:**k3750 http://dx.doi.org/10.1136/bmj.k3750

2. Cook JA, Julious SA, Sones W, Hampson LV, Hewitt C, Berlin JA, Ashby D, Emsley R, Fergusson DA, Walters SJ et al. Practical help for specifying the target difference in sample size calculations for RCTs : the DELTA2 five-stage study, including a workshop Health Technology Assessment 2019 2**3(60):**1-88 01

3. Flight L and Julious SA. Practical guide to sample size calculations: an introduction. *Pharmaceutical Statistics* 2016 **15(1)** 75-79

4. Flight L and Julious SA. Practical guide to sample size calculations: superiority trials. Pharmaceutical Statistics 2016 15(1) 80-89

5. Flight L and Julious SA. Practical guide to sample size calculations: non-inferiority and equivalence trials. Pharmaceutical Statistics 2016:15(1) 68-74

6. Julious SA. Tutorial in Biostatistics: Sample Sizes for clinical trials with Normal Data. *Statistics in Medicine* 2004 **23:**1921-86

7. Julious, SA. Sample sizes for clinical trials. Chapman and Hall, 2009

8. Julious SA and Campbell MJ. Tutorial in Biostatistics: Sample Sizes for clinical trials with binary data. *Statistics in Medicine* 2012;**31**:2904–36

10 Appendix

The appendix gives an overview of sample size calculations and methodologies for parallel group studies where the primary outcome is a survival endpoint used in the App SampSize.

10.1 Superiority Trials

Suppose the event of interest is a negative: for example death or recurrence such that the primary objective of the trial is to delay the event from happening. The objective of the trial would be to slow down the time to the primary outcome and the primary analysis for such a response would be a log-rank test.

Now suppose the survival distributions for the two arms of the trial have instantaneous death rates of λ_A for treatment A and λ_B for treatment B. Now from this the hazard ratio (HR) is defined as

(A1) $$HR = \lambda_A/\lambda_B.$$

In terms of the hazard ratio the null and alternative hypothesis would be of the form

H_0: The survival experience for both treatment groups is the same (HR=1).

H_1: The survival experience for both treatment groups differs (HR≠1).

If the hazard ratio does not change with time, then it can be estimated by

(A2) $$HR = \frac{\log \pi_A}{\log \pi_B},$$

where π_A and π_B are two survival rates at some fixed time point. Assuming an exponential survival an alternative formula for the Hazard Ratio is to derive it in terms of the median survival terms for each treatment

(A3) $$HR = \frac{M_B}{M_A},$$

where M_A and M_B are the median survival times on A and B respectively.

For a given hazard ratio (HR) the number of events, *E*, required in each patient group in SampSize is estimated from

(A4) $$E = \frac{2(Z_{1-\alpha/2}+Z_{1-\beta})^2}{(\log HR)^2}.$$

Where $Z_{1-\alpha/2}$ and $Z_{1-\beta}$ are standardised Normal values for α (the Type I error) and β (the Type II error. Here 1-β is the power of the study. The significance level is taken as two sided and so α/2 is used.

10.2 Non-inferiority Trials

Assuming that a HR < 1 favours the experimental treatment the hazard ratio for the null and alternative hypothesis is now of the form

> H_0: The survival experience for the new treatment is inferior to the control (HR≥d).
>
> H_1: The survival experience for new treatment is the same or favours it compared to control (HR<d).

The number of events, *E*, required in each patient group is estimated within SampSize from

(A5) $$E = \frac{2(Z_{1-\alpha}+Z_{1-\beta})^2}{(\log HR-\log(d))^2}.$$

where d is the non-inferiority limit in terms of a hazard ratio and Z1-α and Z1-β are defined as for the superiority trials. The significance level is taken as one sided and so α is used.

10.3 Equivalence Trials

In the case of an equivalence trial the aim is to show that the experimental treatment does not differ in efficacy (or safety) from the current treatment in either direction. An

example of the use of this design is in the case of biosimilar trials. The hazard ratio for the null and alternative hypothesis is now of the form

> H_0: The survival experience for either treatment groups is inferior to the other (HR≥d or HR≤d).
>
> H_1: The survival experience for both treatment groups is the same (d<HR<d).

For equivalence trials a direct estimate of the sample size is not possible. However, for a given number of events the power of the study can be estimated. Hence, for given a number of events, *E*, SampSize estimates the power from the following result.

(A6) $1-\beta = \Phi\left(\frac{\sqrt{E}|\log HR-\log(d)|}{\sqrt{2}} - Z_{1-\alpha}\right) + \Phi\left(\frac{\sqrt{E}|\log HR+\log(d)|}{\sqrt{2}} - Z_{1-\alpha}\right) - 1.$

where d is the equivalence limit in terms of a hazard ratio. To estimate the sample size you would need to iterate on E to obtain the required to get a E which gave the requisite power. $Z_{1-\alpha}$ and $Z_{1-\beta}$ are defined as for the non-inferiority and superiority trials. As for non-inferiority trials the significance level is taken as one sided and so α is used.

Note that for the special case of HR=1 we can estimate the sample size from.

(A7) $$E = \frac{2(Z_{1-\alpha}+Z_{1-\beta/2})^2}{(\log(d))^2}.$$

10.4 Precision Trials

Precision trials are designs employed to consider the precision of for example a devise in diagnosing a disease of interest. The trials are less about proving there is a treatment difference than estimating plausible treatment differences with a view to undertaking a definitive trial later.

To obtain a sample size to have required precision w about the hazard ratio SampSize uses the following result

(A8) $$E = \frac{2Z_{1-\alpha/2}^2}{(\log(1-w))^2}.$$

The precision for a given n can be estimate from

(A9) $$\mathrm{w} = 1 - exp - \left(\sqrt{\frac{2Z_{1-\alpha/2}^2}{E}}\right).$$