Jurnal Teknologi

Full paper

Determination of the Required Sample Size with Assurance for Three-Arm Non-Inferiority Trials

Nor Afzalina Azmee[a*], Zulkifley Mohamed[a] ,Azhar Ahmad[a]

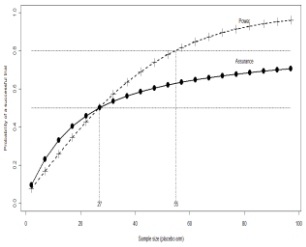
[a]*Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, 35900, Tanjung Malim, Perak*

*Corresponding author: afzalina@fsmt.upsi.edu.my

Article history

Received :21 January 2013
Received in revised form :
7 May 2013
Accepted :25 June 2013

Graphical abstract



Abstract

The concept of assurance in the two-arm non-inferiority trials has been explored, expressing the non-inferiority margin as a clinically meaningful treatment difference. This short paper focuses on developing an assurance formula in the three-arm non-inferiority trial, based on the ratio of means. The discussion starts with the simple case of known variances and then extends to the case of unknown but equal variances. To avoid complicated integration, assurance for the latter case was studied using Bayesian Clinical Trial Simulation (BCTS). The findings indicate that assurance allows the experimenter to formally take into account the uncertainty surrounding the parameter estimates by using the prior distributions. Furthermore, BCTS can be easily implemented to find the required sample size without having to resort to complex integration.

Keywords: Assurance; power; non-inferiority trial; three-arm design; BCTS

Abstrak

Kajian kaedah jaminan di dalam ujian tidak-inferior dua-kumpulan telah pun dibuat dengan mewakilkan margin tidak-inferior sebagai perbezaan rawatan yang bermakna (dalam konteks klinikal). Fokus kertas ini pula adalah terhadap pembentukan formula jaminan di dalam ujian tidak-inferior tiga-kumpulan, yang mewakilkan margin tidak-inferior sebagai nisbah min. Perbincangan dimulakan dengan kes asas di mana varians populasi dianggap diketahui dan dilanjutkan pada kes varians populasi tidak diketahui. Oleh kerana kes yang kedua ini melibatkan kamiran kompleks, kaedah jaminan telah diaplikasikan dengan bantuan Bayesian Clinical Trial Simulation (BCTS). Hasil kajian mendapati bahawa kaedah jaminan membolehkan penyelidik mengambil kira secara formal ketidakpastian berkenaan anggaran parameter yang terlibat di dalam pencarian saiz sample, iaitu dengan mewakilkan ketidakpastian tersebut menggunakan taburan prior yang bersesuaian. Selain itu, BCTS boleh diaplikasikan dengan mudah untuk mencari saiz sampel yang diperlukan tanpa perlu menyelesaikan masalah kamiran yang kompleks.

Kata kunci: Jaminan; kuasa; ujian tidak-inferior; rekabentuk 3-kumpulan, BCTS

1.0 INTRODUCTION

Sample size calculation is an essential matter in any clinical trial design. In practice a necessary sample size has to be determined before any trial is executed and has to be made clear in the protocol as well. Generally it is considered unethical to enroll patients in a small trial when it may not yield any positive outcome. If the trial is small, the desired treatment difference may not be demonstrated and patients may be exposed to unnecessary risk. Such a small trial is also termed as under-powered trial. The widely held view that an under-powered trial is unethical was challenged, arguing that a small trial is ethical given that it is methodogically strong and that the methods and the results are published properly, irrespective of having negative or positive outcomes.[1] The study also seems to suggest that if the available patients will only give power of 50 percent or 60 percent to detect a particular minimum treatment difference, as opposed to the conventional 80 or 90 percent, the ethical committee should consider granting the permission to run this trial. This view seems to be in line with earlier studies, which suggested that under-powered trials can be useful in meta-analysis study.[2-3]

The frequentist approach and the notion of having 80 percent or 90 percent power used to be dominant in the design of clinical trial. However, the Bayesian approach or the mixed Bayesian-frequentist approach has started to garner considerable interest among the researchers.[4-10] In particular, the mixed Bayesian-frequentist is regarded as a favorable approach as it enables the investigators to use a proper prior at the designing stage and then switches to a weak prior in the final analysis. This then fulfills the regulatory requirement to have the data speaking for themselves. This study focuses on the implementation of assurance, considered to fall into the mixed Bayesian-frequentist category.

As opposed to the concept of power, assurance is the unconditional probability that the trial will yield a positive outcome. This concept is seen appealing since it is more natural and less arbitrary that the unknown parameters are expressed in terms of prior distributions rather than the point estimates, although in practice this idea may instigate controversial debates among the members of ethical committee. Previous studies have seen the implementation of assurance in different settings such as on one-sided superiority trials, two-sided superiority trials, two-arm non-inferiority trials and equivalence trials, with examples of normal and binary data.[4-5] These previous studies also seem to imply that a trial can be executed given the available sample size assures an unconditional probability of observing a positive outcome of just around 50 percent. Sample size determination using assurance for the case of two-arm non-inferiority trial has been discussed,[4] expressing the non-inferiority margin as a clinically meaningful treatment difference. The focal point of this paper however is on developing a sample size formula for the case of three-arm non-inferiority trial, based on the ratio of means.

2.0 STATISTICAL DESIGN FOR THREE-ARM NON-INFERIORITY TRIALS

The three-arm non-inferiority trial consists of having an experimental, a reference and a placebo arm, assuming that the inclusion of a placebo arm is properly justified.[11-12] The approach illustrated here considers having the outcome variables that are normally distributed with common but unknown variances and that higher values correspond to better efficacy. The statistical procedure follows the one proposed in earlier study,[12] that is testing superiority of reference against placebo in the first place, and then followed with testing non-inferiority of experimental treatment against the reference, provided that superiority can be established in the first stage. The hypothesis formulation in the second stage is given as follows:

$$H_0: \frac{\mu_E - \mu_P}{\mu_R - \mu_P} \leq \theta \quad \text{versus} \quad H_1: \frac{\mu_E - \mu_P}{\mu_R - \mu_P} > \theta \tag{1}$$

Where μ represents the population mean, E, R and P denote the experimental, reference and placebo group respectively and θ is a positive value, usually ranging between 0.5 to 0.8. The rejection of the null hypothesis in Equation (1) implies non-inferiority of the experimental treatment with respect to reference. To be specific, the new treatment is said to preserve at least $\theta \times 100$ percent of the efficacy of the reference achieved with respect to placebo. Technically, a necessary sample size is derived by first identifying the main objective of a trial, although the statistical procedure outlined may involve other different tests. As in example above, the sample size calculation should meet the objective of showing non-inferiority. The following sample size equation is derived by considering the frequentist approach:

$$n_P = \left(z_{1-\alpha} + z_{1-\beta}\right)^2 \left(\frac{1}{c_E} + \frac{\theta^2}{c_R} + (1-\theta)^2\right)\left(\frac{\varepsilon}{\rho - \theta}\right)^2 \tag{2}$$

Where $z_{1-\alpha}$ and $z_{1-\beta}$ are the 100(1-α)% and 100(1-β)% significance point of the standard normal distribution respectively, $c_E = n_E/n_P$, $c_R = n_R/n_P$, where n denotes the sample size and $\varepsilon = \sigma/(\mu_R - \mu_P)$. Note that Equation (2) is slightly different from the previous study,[12] since it considers expressing the sample sizes in experimental and reference groups as proportions of those in the placebo group. Thus, the total sample size, N for a clinical study is given as:

$$N = n_P(1 + c_E + c_R) \tag{3}$$

To illustrate the application of frequentist sample size calculation, this study has considered setting the type I error rate, $\alpha = 0.025$ and the type II error rate $\beta = 0.2$. The aim was to demonstrate that the new treatment was at least 80 percent as effective as the reference treatment; hence the non-inferiority margin was set as $\theta = 0.8$. The sample size allocation for experimental and reference group are chosen to be $c_E = 5$ and $c_R = 4$, following the optimal allocation discussed in earlier study.[12] Note that other different allocations may also be used if desired. The population means for experimental and placebo groups (μ_E and μ_P) were fixed while the population mean for reference group (μ_R) was made to vary in a decreasing manner to demonstrate an increased ratio ρ (see Table 1). Any calculation which gave a non integral value of sample size in a placebo group (n_p) was rounded up before calculating the total sample size, N.

Table 1 Sample size determination, assuming $\mu_E = 4.2$, $\mu_P = 3.0$ and that $\sigma_E = \sigma_R = \sigma_P = 1.0$

Population Mean μ_R	True Ratio $\rho = (\mu_E - \mu_P)/(\mu_R - \mu_P)$	Sample Size n_P	Total Sample Size N	Power $1 - \beta$
4.33	0.9	177	1770	0.817
4.20	1.0	55	550	0.820
4.09	1.1	30	300	0.803
4.00	1.2	20	200	0.786
3.92	1.3	15	150	0.720
3.86	1.4	12	120	0.561
3.80	1.5	11	110	0.550
3.75	1.6	9	90	0.424
3.71	1.7	8	80	0.370
3.67	1.8	8	80	0.325
3.63	1.9	7	70	0.269
3.60	2.0	7	70	0.267

The power column demonstrates the probability of declaring non-inferiority of the new treatment, based on employing the two-stage procedure; that is testing superiority of reference against placebo in the first stage, followed with testing non-inferiority in the second stage. Power was obtained by recording the number of rejections out of 1000 simulated data sets generated independently for each different configuration. Despite using the optimal allocation, power is seen to decrease as the ratio increases. This scenario, termed as the inflation of a type II error rate was also noted in earlier study.[12] However, this paper will not be addressing this as the focus is on the comparison of using power and assurance to determine the required sample size. Details on the implementation of assurance are given in the next section.

3.0 ASSURANCE IN THREE-ARM NON-INFERIORITY TRIALS

The discussion here begins with the case of known variances. Note that the hypothesis formulation in Equation (1) can be re-arranged in a linear form, given as follows:

$$H_0: \mu_E - \theta\mu_R - (1-\theta)\mu_P \leq 0 \quad \text{versus} \quad H_1: \mu_E - \theta\mu_R - (1-\theta)\mu_P > 0 \tag{4}$$

The null hypothesis is rejected when:

$$\bar{X}_E - \theta\bar{X}_R - (1-\theta)\bar{X}_P > z_{1-\alpha}\tau \tag{5}$$

Where $\bar{X}$ represents the sample mean and τ is given as:

$$\tau = \sqrt{\frac{\sigma_E^2}{n_E} + \frac{\theta^2\sigma_R^2}{n_R} + \frac{(1-\theta)^2\sigma_P^2}{n_P}} \tag{6}$$

Assurance, γ can be defined as:

$$\gamma = P(\bar{X}_E - \theta\bar{X}_R - (1-\theta)\bar{X}_P > z_{1-\alpha}\tau\ ) \tag{7}$$

The conventional approach, power will then consider the following sampling distribution:

$$\bar{X}_E - \theta\bar{X}_R - (1-\theta)\bar{X}_P \sim N(\omega, \tau^2) \tag{8}$$

Where $\omega = \mu_E - \theta\mu_R - (1-\theta)\mu_P$. The assurance approach however will consider placing an additional normal prior to ω, say $\omega \sim N(m, v)$, with mean m and variance v. Therefore the sampling distribution is expressed as:

$$\bar{X}_E - \theta\bar{X}_R - (1-\theta)\bar{X}_P \sim N(m, \tau^2 + v) \tag{9}$$

Taking the equation above into consideration gives the final look of assurance, defined as such:

$$\gamma = \Phi\left(\frac{-z_{1-\alpha}\tau + m}{\sqrt{\tau^2 + v}}\right) \tag{10}$$

Where Φ denotes the standard normal distribution function.

To demonstrate the application of assurance, let's consider the following population means $\mu_E = \mu_R = 4.2$ and $\mu_P = 3.0$, which then corresponds to having a ratio of $\rho = 1$ and $\omega = 0.24$. Suppose that the population variances are known and are set equal as 1. In a setting where the non-inferiority margin θ is set at 0.8, the unequal allocation of sample size in the ratio of 5, 4 and 1 for experimental, reference and placebo groups is the most optimal.[12]

Since the true value of parameters are unknown, it is sensible to introduce uncertainty in the sample size calculation. Suppose that ω is assumed to have a normal distribution with mean $m = 0.24$ and variance $v = 0.04$. Note that the choice of v reflects how uncertain the investigator is regarding the value of $\omega = 0.24$ placed in the sample size equation. The previous section has demonstrated that a sample size of $n_P = 55$ is required to achieve power of 80 percent. When it comes to assurance (depending on the value of v, which in this case is 0.04), the same amount of subject would only give 63 percent assurance. To give an 80 percent assurance, a relatively huge sample size $n_P = 322$ is required. This might be unnecessary and a waste of resources to recruit such a large number of patients. Details on these comparisons are depicted in Figure 1.

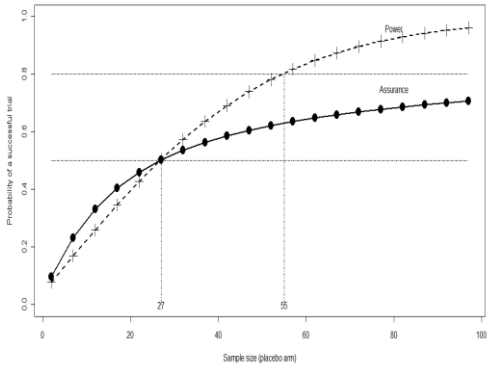


Figure 1 Probability of having a successful trial, based on assurance and power

Several authors seemed to indicate that assurance of around 50 percent is good enough to have the trial executed.[4,5] Figure 1 demonstrates that assurance of 50 percent is attained when the sample size in the placebo arm is 27, which corresponds to having a total sample size of 270. It is about half the size required by the frequentist method to obtain a power 80 percent. To carry out a trial with this relatively small sample size may cause a controversial debate and that the permission to run a trial may not be granted. However, this paper argues that the sponsor should have given a chance to express the sponsor's belief when it comes to designing a trial. The number of sample size adopted to run a trial should not be an issue, albeit large or small or negative or positive as these results can still be used in the meta-analysis study.

Now, let's consider a more practical scenario where the population variances are unknown but are assumed to be equal. Thus, an additional prior has to be specified for the unknown parameter σ^2. To demonstrate its application, the problem outlined above is considered. Assume that a prior belief for σ^2 can be represented by using a log normal prior, ln $\sigma^2 \sim N(0, 0.0625)$, with mean $a = 0$ and variance $b = 0.0625$. As discussed in the earlier study,[4] assurance can be easily computed using the Bayesian Clinical Trial Simulation and the steps implemented here have been modified accordingly, given as follows:

i. Define the counters; $I = 0$ and $S = 0$, where I corresponds to a number of repetition and S corresponds to a number of successful event.
ii. Define the number of repetition, $J = 1000$.
iii. Define the number of subjects in the placebo arm, say n_P and considers allocating the subjects optimally across the treatment arms. Since the non-inferiority margin is chosen as $\theta = 0.8$, the optimal allocation is 5:4:1

(c_E:c_R:1) where c_E and c_R are the proportions of sample size in the experimental and reference groups with respect to those in the placebo group.

iv. Sample $\omega = \mu_E - \theta\mu_R - (1-\theta)\mu_P$ from a normal distribution $N(m, v)$, with mean m and variance v.

v. Sample σ^2 from a log normal prior, ln $\sigma^2 \sim N(a, b)$, that is with mean a and variance b.

vi. Using the results in (iii), (iv) and (v), sample $\bar{X}_E - \theta\bar{X}_R - (1-\theta)\bar{X}_P$ from a normal distribution with mean ω and variance:

$$\sigma^2[(1/n_E) + (\theta^2/n_R) + ((1-\theta)^2/n_P)]$$

vii. Using the results of (iii) and (v), $\hat{\sigma}^2$ can be obtained by sampling from chi-square distribution, that is $\hat{\sigma}^2(n_E + n_R + n_P - 3)/\sigma^2 \sim \chi^2_{n_E+n_R+n_P-3}$.

viii. Using the results of (vi) and (vii), calculate the test statistic:

$$T = \frac{\bar{X}_E - \theta\bar{X}_R - (1-\theta)\bar{X}_P}{\hat{\sigma}\sqrt{\frac{1}{n_E} + \frac{\theta^2}{n_R} + \frac{(1-\theta)^2}{n_P}}}$$

ix. If $T > t_{1-\alpha, n_E+n_R+n_P-3}$ where α = 0.025, update $S = S + 1$

x. Update $I = I + 1$

xi. While $I \leq N$, repeat the following steps (iii) – (x)

xii. An assurance, given a particular total sample size, $n_P(1 + c_E + c_R)$ is calculated by $\gamma = S/J$

Consider an example of known variances constructed in the early part of Section 3. The prior distribution for ω is assumed as $N(0.24, 0.04)$, with mean $m = 0.24$ and variance $v = 0.04$. Now, suppose that a prior belief for σ^2 can be represented as $N(0, 0.0625)$, where the mean $a = 0$ and variance $b = 0.0625$. As revealed in Figure 2, the assurance derived from having this additional uncertainty is not very much different than those which assumed that σ^2 is known.

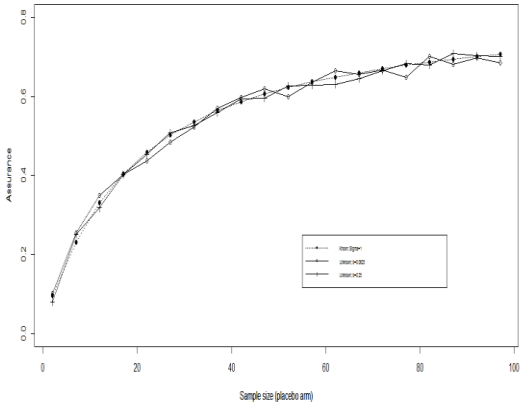


Figure 2 The impact of different values of variance b on the assurance curve

Further exploration of using different values of b has also been explored and some of them are illustrated in Figure 2, which sees no significant difference to the assurance curve. Given a fixed sample size, a reduction of assurance will be observed if only the investigator is willing to place an unrealistically large value of variance b.

4.0 CONCLUSION

Overall, this study indicates that a mixed classical-Bayesian approach, specifically the concept of assurance is appealing. This is because there is always at least some amount of prior information about the unknown parameters but not enough for us to give a reliable point estimate. This could possibly reflect the problem of conducting a trial in situations where the responses of the treatment arms are seen to be varied across many different trials and so it may be difficult to specify point estimates required in the conventional sample size equation. In these circumstances, it is best to represent the information using a prior distribution. Since the decision to run a trial is related to the sponsor's risk (should the result of the trial is negative), the sponsors should be given a chance to express their prior belief at the design stage. At the analysis stage, the frequentist analysis is carried out, to conform to the regulatory bodies.

The study also has demonstrated that Bayesian Clinical Trial Simulation (BCTS) can be easily implemented to find the required sample size, even in complicated cases which involved complex integration. Naturally, the choice of a sample sized based on assurance is a subjective matter. Unlike power, it is not possible to fix an assurance of say γ for all situations. This is because the sample size which gives an assurance of 60 percent may give an assurance of 50 percent when a detailed prior is considered. Ideally, the decision to adopt the sample size at assurance γ depends on the sponsor's judgement, but it may also be indirectly influenced by the available resources. Thus, it is important that the sample size calculation and the priors' specifications truly reflect the sponsor's belief and are conducted transparently. The proposal to run a trial may be rejected by the ethics committee if the sample size is thought to be too small compared to the conventional power analysis. Perhaps it might be best in practice to compare the sample size based on assurance with those based on power to strike some sort of balance.

Acknowledgement

This research was supported by the Fundamental Research Grant Scheme FRGS (2011-02-) and Universiti Pendidikan Sultan Idris.

References

[1] Schulz, K. F. & Grimes D. A. 2005. Sample Size Calculations in Randomized Trials: Mandatory and Mystical. *The Lancet*. 365: 1348–1353.

[2] Sackett, D. L. & Cook, D. J. 1993. Can We Learn Anything from Small Trials? *Annals of the New York Academy of Sciences*. 703: 25–31.

[3] Chalmers, T.C., Levin, H. S. H. S., Reitman, D., Berrier, J. & Nagalingam, R. 1987. Meta-analysis of Clinical Trials as a Scientific Discipline, I: Control of Bias and Comparison with Large Co-operative Trials. *Statistics in Medicine*. 6: 315–328.

[4] O'Hagan, A. Stevens, J. W. & Campbell, M. J. 2005. Assurance in Clinical Trial Design. *Pharmaceutical Statistics*. 4: 187–201.

[5] O'Hagan, A. & Stevens, J. W. 2001. Bayesian Assessment of Sample Size for Clinical Trials of Cost-effectiveness. *Medical Decision Making*. 21: 219–230.

[6] Simon, R. 2000. Clinical Trials and Sample Size Considerations: Another Perspective: Comment. *Statistical Science*. 15: 103–105.

[7] Pezeshk, H. 2003. Bayesian Techniques for Sample Size Determination in Clinical Trials: A Short Review. *Statistical Methods in Medical Research*. 12: 489–504.

[8] Wang, H. Chow, S. C. & Chen, M. 2005. A Bayesian Approach on Sample Size Calculation for Comparing Means. *Journal of Biopharmaceutical Statistics*. 15: 799–807.

[9] Grieve, A. P. 2007. 25 Years of Bayesian Methods in the Pharmaceutical Industry: A Personal, Statistical Bummel. *Pharmaceutical Statistics*. 6: 261–281.

[10] Daimon, T. 2008. Bayesian Sample Size Calculations for a Non-Inferiority Test of Two Proportions in Clinical Trials. *Contemporary Clinical Trials*. 29: 507–516.
[11] Temple, R. & Ellenberg, S. S. 2000. Placebo-controlled Trials and Active-control Trials in the Evaluation of New Treatments. Part 1: Ethical and Scientific Issues. *Annals of Internal Medicine*. 133: 455–463.
[12] Pigeot, I., Schäfer, J., Röhmel, J. & Hauschke, D. 2003. Assessing Non-Inferiority of a New Treatment in a Three-Arm Trial Including A Placebo. *Statistics in Medicine*. 22: 883–889.