

Collection of Biostatistics Research Archive
COBRA Preprint Series

Year 2008

Paper 36

The Design and Sample Size Requirement for
a Cluster Randomized Non-Inferiority Trial
with Two Binary Co-Primary Outcomes.

William F. McCarthy*

*Maryland Medical Research Institute, dr.w.f.mccarthy@gmail.com

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://biostats.bepress.com/cobra/art36>

Copyright ©2008 by the author.

The Design and Sample Size Requirement for a Cluster Randomized Non-Inferiority Trial with Two Binary Co-Primary Outcomes.

William F. McCarthy

Abstract

This paper will discuss the design and sample size requirement for a cluster randomized non-inferiority trial with two binary co-primary outcomes. A hypothetical study (the EXAMPLE Trial) will be considered.

Lets assume the EXAMPLE Trial will consist of two separate binomial non-inferiority two-sample trials. Trial 1: the Coronary Artery Disease known population (co-primary 1) and Trial 2: the Coronary Artery Disease unknown population (co-primary 2). A physician-month cluster randomization scheme will be used. That is, for each trial (trial 1 and trial 2) every month for a 12-month period, each physician participating in the EXAMPLE Trial will be allocated a randomized cluster of size 10. The physician will need to consent and enroll 10 patients each month for the 12-month period for each trial (trial 1 and trial 2). Each cluster will be specific to a treatment group (either EXPERIMENTAL or CONTROL).

The design and sample size method discussed by Bland (2003) and Donner and Klar (2000) will be used.

The EXAMPLE Trial will be declared a success if statistical significance is demonstrated at the pre-specified nominal alpha-level for both co-primary outcomes.

Lets assume the EXAMPLE Trial will consist of two separate binomial non-inferiority two-sample trials. Trial 1: the CAD known population (co-primary 1) and Trial 2: the CAD unknown population (co-primary 2). A physician-month cluster randomization scheme will be used. That is, for each trial (trial 1 and trial 2) every month for a 12-month period, each physician participating in the EXAMPLE Trial will be allocated a randomized cluster of size 10. The physician will need to consent and enroll 10 patients each month for the 12-month period for each trial (trial 1 and trial 2). Each cluster will be specific to a treatment group (either EXPERIMENTAL or CONTROL).

The design and sample size method discussed by Bland (2003) and Donner and Klar (2000) will be used.

The EXAMPLE Trial will be declared a success if statistical significance is demonstrated at the pre-specified nominal α -level for all co-primary outcomes.

Therefore, non-inferiority must be demonstrated with respect to both co-primary 1 and co-primary 2 in order for the EXAMPLE Trial to be considered a success. Since both co-primary outcomes are required to be met, no adjustment to the significance level ($\alpha = 0.025$; 1-sided) is required (EMEA: CPMP/EWP/908/99; 2002; section 2.1.1 and Sankoh AJ et al.; 2003; section 3.1.1). The EMEA document states: “ In this situation, there is no intention or opportunity to select the most favourable result and, consequently, the individual type I error levels are set equal to the overall type I error level α , i.e., no reduction is necessary”.

However, the EMEA document notes: “This procedure inflates the relevant type II error (here: falsely accepting that at least one null hypothesis is true), which in the worst case scenario is the sum of the type II errors connected with the individual hypotheses. This inflation must be taken into account for a proper estimation of the sample size for the trial”.

Therefore, we will need to set the type II error rate for each co-primary outcome at a value which, in the worst case scenario, will sum to a value no greater than 0.20 (since, for this trial, we want an overall type II error rate = 0.20; that is power = 80%). Thus, we will use $\beta = 0.10$ for each co-primary outcome.

EAST 5.2.0 will be used to determine the sample size (East, 2008).

Trial 1: Sample Size Calculation for Co-Primary 1:

Lets assume the study design consists of two treatment groups. CONTROL is the active control group and EXPERIMENTAL is the experimental treatment group.

Co-primary 1 will be treated as a binomial non-inferiority two-sample trial. The goal is to establish that the death rate of the experimental treatment group is no worse than that of the active control group. A difference in proportions is considered.

Co-Primary 1: Coronary Artery Disease (CAD) known: Death rate at 12 months.

Denote π_c as the death rate for the active control group and π_i as the death rate for the experimental treatment group.

Let $\delta = \pi_c - \pi_t$.

Let the non-inferiority margin be denoted as δ_0 .

The null hypothesis is $H_0 : \delta = \delta_0$ and is tested against a one-sided alternative hypothesis.

Since the occurrence of a response denotes patient harm rather than benefit, then $\delta_0 < 0$ and the alternative hypothesis is $H_1 : \delta > \delta_0$ or equivalently as $H_1 : \pi_t < \pi_c - \delta_0$.

For any given π_c , the sample size is determined by the desired power at a specified value of $\delta = \delta_1$. A common choice is $\delta_1 = 0$ or (equivalently $\pi_t = \pi_c$). For co-primary 1, $\delta_1 = 0$.

The active control group death rate is assumed to be $\pi_c = 0.042$.

The non-inferiority margin is assumed to be $\delta_0 = -0.011$, i.e., $\pi_t = 0.053$.

Information Used by EAST to Calculate the Sample Size (assuming no clustering):

A non-inferiority trial is considered

Difference of proportions

Binomially distributed data is considered

The active control group event rate is assumed to be 0.042.

The non-inferiority margin is assumed to be -0.011.

The fraction of patients assigned to each study arm is 1:1 (exp. treatment : active control)

The α -level is 0.025

The β -level is 0.10 (90% power)

One-sided hypothesis testing is used



Results from EAST.

The screenshot shows the EAST software interface for a Binomial Non-Inferiority Trial. The main window displays the following parameters and results:

Parameter	Value	Information Fraction	Cumulative Accrual	Alpha Spent	Boundary to Reject H0	Boundary Crossing Probabilities		
						Under H0	Under H1	Under H1/2
1-Sided or 2-Sided Test	1-Sided	1.000	13976.093	0.02500	1.9600	0.025	0.900	0.367
Significance Level (α)	0.025							
Power ($1 - \beta$)	0.9							
Assigned Fraction (Treatment)	0.5							
Planned Number of Looks	1							
Proportion Response (Control: π_c)	0.042							
Non-Inferiority Margin (δ_0)	-0.011							
Difference in Proportions (δ_1)	0.0							
Maximum	13976							
Expected Under H0								
Expected Under H1								
Expected Under H1/2								

A total sample size of 13, 976 is required.

In order to account for the physician-month cluster design to be used in the EXAMPLE Trial, this total sample size of 13, 976 needs to be inflated to account for the effect of clustering.

The ICC method is used to inflate the sample size to account for the effect of clustering. We will use an ICC=0.01.

In a talk presented by J M Bland to the RSS Medical Section and the RSS Liverpool Local Group, 12 NOV 2003, Bland stated “the magnitude of the effect of clustering is measured by the design effect, $Deff$, given by the following: $Deff = 1 + (n - 1)(ICC)$ where n is the number of observations in a cluster and ICC is the intra-cluster correlation coefficient. The ICC is the correlation between pairs of subjects chosen at random from the same cluster. It is usually quite small, 0.04 is a typical figure. This was the median ICC reported in the review by Eldridge *et al.* (2004). If $n=1$, cluster size one, in other words, no clustering, then $Deff=1$, otherwise $Deff$ will exceed 1 (this assumes a positive correlation between pairs of subjects).

We can use this in two ways. In design, if we estimate the required sample size ignoring clustering, we must multiply it by the design effect to get the sample size required for the clustered sample. Alternatively, we can say that if the sample size is estimated ignoring the clustering, the clustered sample has the same power as for a simple sample of size equal to what we get if we divide our sample size by the design effect. In analysis, if we analyse the data as if there were no clusters, the variances of the estimates must be multiplied by *Deff*, hence the standard error must be multiplied by the square root of *Deff*. From this formula, we can see that clustering may have a large effect if the ICC is large or if the cluster size is large. Only one of these conditions need be met. For example, if the ICC is 0.001, a very small correlation, and the cluster size is 500, the design effect will be $1 + (500-1) \times 0.001 = 1.5$ and we would need to increase the sample size by 50% to achieve the same power as an unclustered trial.

In addition, we need to estimate variances both within and between clusters. If the number of clusters is small, the between clusters variance will have few degrees of freedom and we will be using the t distribution in inference rather than the Normal. This too will cost in terms of power.” ICC (ρ) is the correlation between pairs of patients chosen at random from the same cluster.

$$\rho = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2}$$

where σ_B^2 is the between cluster variability

where σ_W^2 is the within cluster variability

the size of the ICC is generally larger for smaller clusters

small cluster ~ 0 to 0.3 (large ICC)

medium cluster ~ 0 to 0.05 (medium ICC)

large cluster ~ 0 to 0.001 (small ICC)

Eldridge et al. (2006) also discuss the impact of the design effect on sample size determination for cluster randomized trials.

The EXAMPLE Trial assumes a cluster size of 10 patients (cluster = physician-month). Using $Deff = 1 + (n - 1)(ICC)$, we have $1 + (10-1)(0.01) = 1.09$. Therefore, we need to inflate the sample size 13, 976 by 1.09 to get the total sample size required for co-primary outcome 1 when clustering is considered.

$13, 976 \times 1.09 = 15, 234$ patients. This would result in approximately 1, 524 clusters.

Trial 2: Sample Size Calculation for Co-Primary 2:

Lets assume the study design consists of two treatment groups. CONTROL is the active control group and EXPERIMENTAL is the experimental treatment group.

Co-primary 2 will be treated as a binomial non-inferiority two-sample trial. The goal is to establish that the death rate of the experimental treatment group is no worse than that of the active control group. A difference in proportions is considered.

Co-Primary 2: Coronary Artery Disease (CAD) unknown: Death rate at 12 months.

Denote π_c as the death rate for the active control group and π_t as the death rate for the experimental treatment group.

Let $\delta = \pi_c - \pi_t$.

Let the non-inferiority margin be denoted as δ_0 .

The null hypothesis is $H_0 : \delta = \delta_0$ and is tested against a one-sided alternative hypothesis.

Since the occurrence of a response denotes patient harm rather than benefit, then $\delta_0 < 0$ and the alternative hypothesis is $H_1 : \delta > \delta_0$ or equivalently as $H_1 : \pi_t < \pi_c - \delta_0$.

For any given π_c , the sample size is determined by the desired power at a specified value of $\delta = \delta_1$. A common choice is $\delta_1 = 0$ or (equivalently $\pi_t = \pi_c$). For co-primary 2, $\delta_1 = 0$.

The active control group death rate is assumed to be $\pi_c = 0.065$.

The non-inferiority margin is assumed to be $\delta_0 = -0.016$, i.e., $\pi_t = 0.081$.

Information Used by EAST to Calculate the Sample Size (assuming no clustering)

A non-inferiority trial is considered

Difference of proportions

Binomially distributed data is considered

The active control group event rate is assumed to be 0.065.

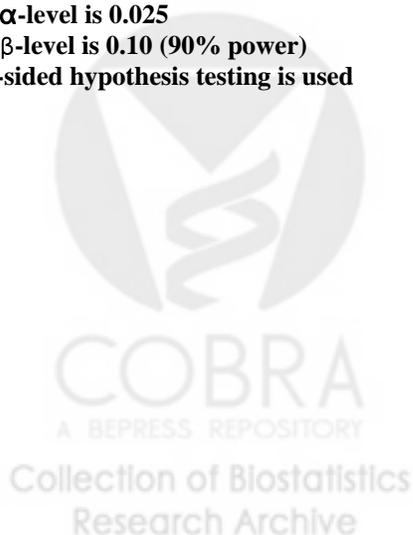
The non-inferiority margin is assumed to be -0.016.

The fraction of patients assigned to each study arm is 1:1 (exp. treatment : active control)

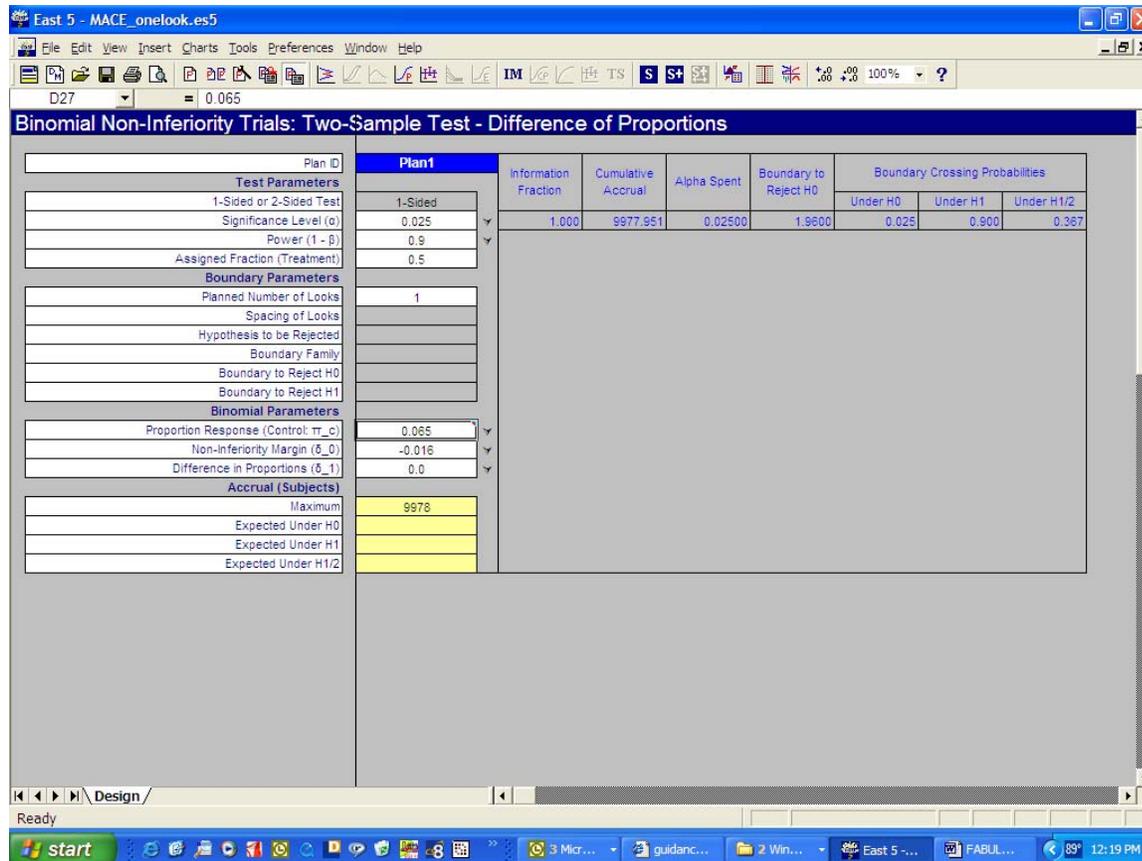
The α -level is 0.025

The β -level is 0.10 (90% power)

One-sided hypothesis testing is used



Results from EAST.



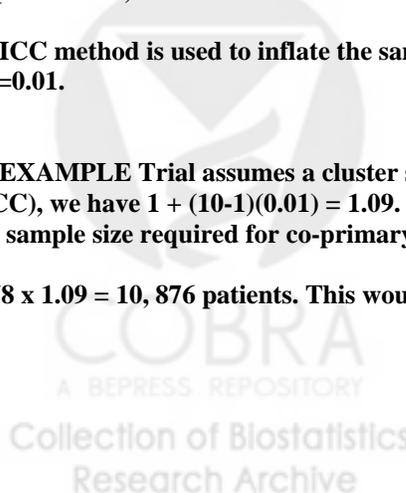
A total sample size of 9, 978 is required.

In order to account for the physician-month cluster design to be used in the EXAMPLE Trial, this total sample size of 9, 978 needs to be inflated to account for the effect of clustering.

The ICC method is used to inflate the sample size to account for the effect of clustering. We will use an ICC=0.01.

The EXAMPLE Trial assumes a cluster size of 10 patients (cluster = physician-month). Using $Deff = 1 + (n - 1)(ICC)$, we have $1 + (10-1)(0.01) = 1.09$. Therefore, we need to inflate the sample size 9, 978 by 1.09 to get the total sample size required for co-primary outcome 2 when clustering is considered.

9, 978 x 1.09 = 10, 876 patients. This would result in approximately 1, 088 clusters.



Sample Size Requirement for the EXAMPLE Trial – One-look approach

Trial 1: Sample Size Calculation for Co-Primary Outcome 1: 15, 234 --> 1, 524 clusters
Trial 2: Sample Size Calculation for Co-Primary Outcome 2: 10, 876 --> 1, 088 clusters

Interim Monitoring of the Co-Primary Outcomes using Lan DeMets

The above sample sizes would have to be increased if one wanted to perform formal interim analyses using the Lan DeMets method. For a 3-look approach, we have the following:

Sample Size Requirement for the EXAMPLE Trial – Three-look approach

Trial 1: Sample Size Calculation for Co-Primary Outcome 1: 15, 416 --> 1, 542 clusters
Trial 2: Sample Size Calculation for Co-Primary Outcome 2: 11, 006 --> 1, 102 clusters

The Lan-DeMets (1983) procedure will be used for assessing each co-primary study outcome when the interim “looks” of the data are taken. This procedure allows for flexible interim monitoring while simultaneously preserving the type-I error of the study. An alpha spending function will be used. The rate at which the alpha is spent is a function of the total information available at the time of the interim analysis (i.e., cumulative patient accrual). A stopping boundary that preserves the spirit of the O’Brien-Fleming stopping boundary (O’Brien and Fleming, 1979) will be utilized. The software package East version 5.2.0 (East, 2008) will be used for the monitoring of the primary outcome.

Because the Lan-DeMets approach is flexible, the number of “looks” does not have to be specified in advance and the time interval between “looks” does not have to be the same throughout the course of the trial.

The formal monitoring for each co-primary outcome could occur independently or at the same time. It would be best if the first ‘look’ occurred when 1/3 of the patients had information available regarding death at 12 months.

Sensitivity Analysis

This section addresses the impact of changes in assumptions and total sample size required for each trial (trial 1 and trial 2) of the EXAMPLE Trial.

Lets assume for this exercise, that the assumed death rates for the EXPERIMENTAL group and the CONTROL group are reasonable. This implies that the non-inferiority margin is reasonable as well.

This exercise will focus on the impact of:

The ICC assumed

The number of patients that do not consent to participate in the trial, and

The number of patients that are Loss-to-Follow-up (LTF)

Impact of ICC:

Single-Look Design

Cluster size for each trial is 10

ICC is the intra-cluster correlation coefficient.

$$Deff = 1 + (n - 1)(ICC)$$

	No Cluster	Large Cluster	Medium Cluster		Small Cluster	
ICC	0	0.001	0.01	0.05	0.1	0.3
Deff	1	1.009	1.09	1.45	1.9	3.7
Trial 1: effective sample size	13, 976	14, 102	15, 234	20, 266	26, 556	51, 712
clusters	NA	1, 411	1, 524	2, 027	2, 656	5, 172
Trial 2: effective sample size	9, 978	10, 068	10, 876	14, 470	18, 960	36, 920
clusters	NA	1, 007	1, 088	1, 447	1, 896	3, 692

As you can see, as the ICC gets stronger in a positive sense (because the cluster size gets smaller), the required effective sample size gets larger. In most cases, a cluster size of 10 would be considered a small cluster --> thus requiring an ICC of ≤ 0.3 . To be safe one would use an ICC=0.3 (unless pilot study data and/or the literature suggests otherwise). In this case, the effective sample size for trial 1 would be 51, 712 patients (5, 172 clusters) and the effective sample size for trial 2 would be 36, 920 patients (3, 692 clusters).

Impact of Loosing Patients:

Now, we need to inflate the effective sample size of each trial to compensate for loosing patients due to non-consent or LTF. Lets assume the following “lost patient” rates: 5%, 10%, 15%, and 20%. Lets also assume the “lost patient” rate is non-differential (i.e., the rate is the same for each treatment group, for each physician-month, etc). NOTE: This probably is not the case, we probably will have differential “lost patient” rates but to simplify this exercise, we will assume they are non-differential. Lets also just focus on the resulting effective sample sizes for trial 1 and trial 2 when a “small cluster” is considered, ICC= 0.3. We will define the total sample size as [effective sample size x (1+”lost patient” rate)]; where the “lost patient” rate is a proportion.

Single-Look Design

	Small Cluster ICC=0.3	Lost patient rate 5%	Lost patient rate 10%	Lost patient rate 15%	Lost patient rate 20%
	Effective sample size Number of clusters	Total sample size Total number of clusters			
Trial 1	51, 712 5, 172	54, 298 5, 430	56, 884 5, 689	59, 470 5, 947	62, 056 6, 206
Trial 2	36, 920 3, 692	38, 766 3, 877	40, 612 4, 062	42, 458 4, 246	44, 304 4, 431

As you can see, as the “lost patient” rate gets larger, the required total sample size gets larger. Thus, if we assume an ICC=0.3 and a “lost patient” rate of 20%, then the total sample size for trial 1 would be 62, 056 patients (6, 206 clusters). The total sample size for trial 2 would be 44, 304 patients (4, 431 clusters). Therefore, 106, 360 patients would be required (10, 637 clusters) for the EXAMPLE Trial.

Lets hope this is not the case! However, this exercise shows you the impact of the ICC that is assumed and the “lost patient” rate that is assumed on the total sample size requirements.

References

Bland JM. Cluster Randomized Trials in the Medical Literature. Talk presented to the RSS Medical Section and the RSS Liverpool Local Group, 12 NOV 2003.

<http://www-users.york.ac.uk/~mb55/talks/clusml.htm>

Donner A, Klar N. Design and analysis of cluster randomization trials in health research. Arnold Publishers: 2000, London.

EMEA/CPMP/EWP/908/99. Points to consider on multiplicity issues in clinical trials. 2002.

Sankoh AJ et al. Efficacy endpoint selection and multiplicity adjustment methods in clinical trials with inherent multiple endpoint issues, *Statist. Med*, 2003; 22:3133-3150.

East, Version 5.2.0. Cytel Software Corporation, Cambridge, MA. 2008.

Eldridge SM, Ashby D, Feder GS, Rudnicka AR, Ukoumunne OC. Lessons for Cluster Randomised Trials in the 21st Century: A Systematic Review of Trials in Primary Care. *Clin Trials*, 2004; 1:80-90.

Eldridge SM, Ashby D., Kerry S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *International Journal of Epidemiology* 2006; 35: 1292-1300.

Lan KKG and DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika*, 1983; 70, (3), pp(659-663).

O’Brien PC and Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*, 1979; 35, pp(549-556).

