Collection of Biostatistics Research Archive
COBRA Preprint Series

Year 2008 *Paper* 37

The Calculation of the 97.5% Upper Confidence Bound: Application to Clustered Binary Data in a Binomial Non-Inferiority Two-Sample Trial.

William F. McCarthy*

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

http://biostats.bepress.com/cobra/art37

The Calculation of the 97.5% Upper Confidence Bound: Application to Clustered Binary Data in a Binomial Non-Inferiority Two-Sample Trial.

William F. McCarthy

Abstract

This paper will discuss the analysis of a cluster randomized binomial non-inferiority two-sample trial. The determination of the intra-cluster correlation coefficient (ICC) and its use in the calculation of the 97.5% upper confidence bound for delta, the true difference in binomial proportions between the active control and the experimental treatment groups, will be outlined.

As noted in the EMEA guideline on the choice of the non-inferiority margin (2005), "Many clinical trials comparing an experimental treatment with an active control are designed as non-inferiority trials. The term 'non-inferiority' is now well established, but if taken literally could be misleading. The objective of a non-inferiority trial is sometimes stated as being to demonstrate that the experimental treatment is not inferior to the active control. However, only a superiority trial can demonstrate this. In fact, a non-inferiority trial aims to demonstrate that the experimental treatment is not worse than the active control by more than a pre-specified, small amount. This amount is known as the non-inferiority margin, or delta (δ_0)."

It is common practice to adopt the 2.5% significance level for a non-inferiority test, since the test is one-sided. Thus, we would consider the 97.5% upper confidence bound for δ, the true difference in binomial proportions between the active control and the experimental treatment groups.

Example

This paper will discuss the analysis of a cluster randomized binomial non-inferiority two-sample trial. An earlier paper discussed the design and sample size requirements of a cluster randomized binomial non-inferiority two-sample trial (McCarthy, 2008). A cluster will be defined as a private practice whose elements are patients. Each cluster will be specific to a treatment assignment (i.e., either the experimental treatment or the active control). The goal was to have an equal allocation of clusters between the two treatment assignments but an imbalance occurred: 17 clusters for the experimental treatment and 12 for the active control $\rightarrow$ a total of 29 clusters. The size of each cluster varied (from 6 to 35 patients; an average cluster size of 23.3). Lets assume the event of interest is one that denotes patient harm (e.g., death).

Denote π_c as the event rate for the active control group and π_t as the event 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 an event 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 $\pi_t = \pi_c$). Let $\delta_1 = 0$.

The active control group event rate is assumed to be π_c = 0.500.

The non-inferiority margin is assumed to be δ_0 = - 0.050, i.e., π_t =0.550.

Thus, to demonstrate non-inferiority, the 97.5% upper confidence bound when clustering is considered needs to be within the prescribed non-inferiority margin of - 0.050.

Lets illustrate the result of each cluster$_i$ of the trial as $\frac{y_i}{n_i}$. The numerator y_i is the number of events in cluster $_i$ and the denominator n_i is the size of cluster$_i$.

For the experimental treatment arm we have:

```
yi = { 0  0  0  6  2  4  1  2  3  1  6  9  8  6 10  4  4}
ni = {30 22 19 30 30 30 30 30 30 30 30 30 22 22 32 31 20}
```

For the active control group we have:

```
yi = { 2  3  6  1 3  3 2  5 18 21 19 21}
ni = {16 11 35 10 9 11 6 12 25 25 23 24}
```

It should be noted that we could convert this analysis from one of "harm" (death) to one of "benefit" (still alive) as follows (*this is done to make the analysis compatible with standard statistical software*):

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

The active control group event rate is assumed to be π_c = 1- 0.500 = 0.500.

The non-inferiority margin is assumed to be δ_0 = + 0.050, i.e., π_t =1 - 0.550 = 0.450.

Thus, to demonstrate non-inferiority, the 97.5% upper confidence bound when clustering is considered needs to be within the prescribed non-inferiority margin (i.e., $\pi_t > \pi_c - \delta_0$).

For the experimental treatment arm we have:

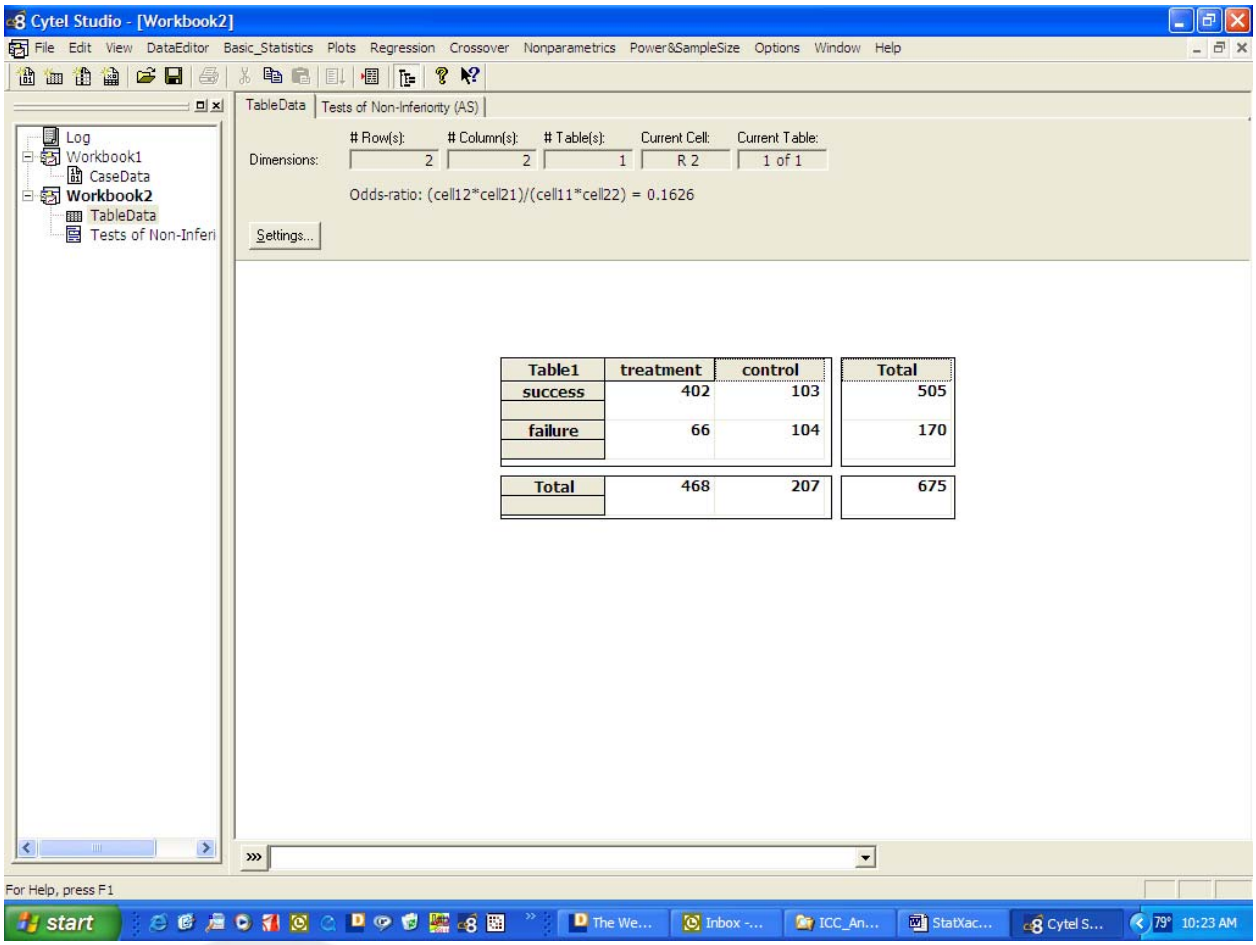
```
yi = {30 22 19 24 28 26 29 28 27 29 24 21 14 16 22 27 16}
ni = {30 22 19 30 30 30 30 30 30 30 30 30 22 22 32 31 20}
```

For the active control group we have:

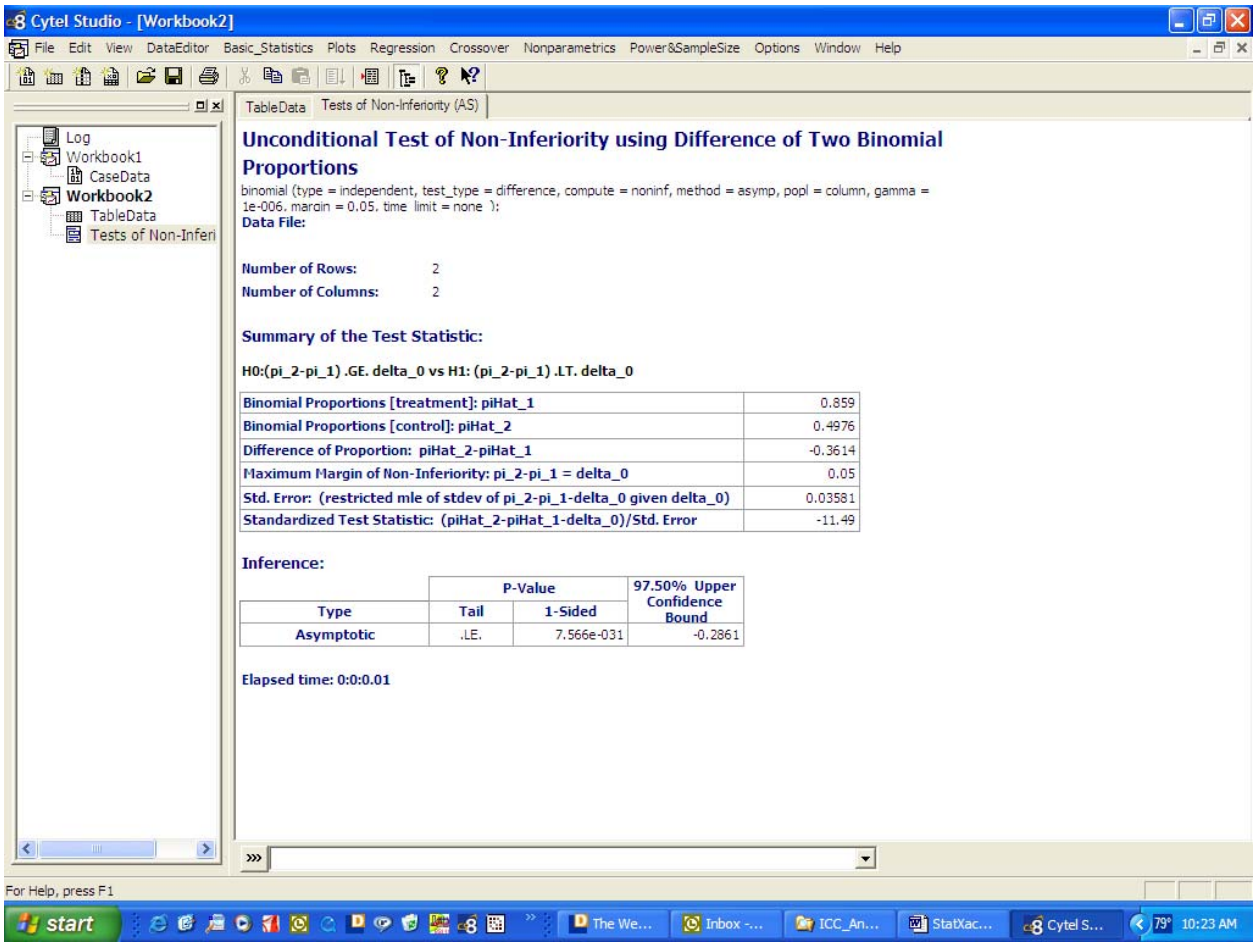
```
yi = {14  8 29  9 6  8 4  7  7  4  4  3}
ni = {16 11 35 10 9 11 6 12 25 25 23 24}
```

Steps of Analysis:

We use the data that represents "benefit". We first analyze the data as if no clustering was involved. We put the data in table form, as shown below. We will use StatXact 7 for our analysis.



Using StatXact 7 we compute the unconditional test of non-inferiority using the difference of the two binomial proportions. The results are below:



Thus, we have:

π_c **= 0.4976**

π_t **= 0.8590**

δ **= - 0.3614**

se($\pi_2 - \pi_1 - \delta_0$)**= 0.03581**

95% upper confidence bound = - 0.2861.

The results above assumed no clustering was involved. We get a difference of proportions of δ **= - 0.3614 and a se(**$\pi_2 - \pi_1 - \delta_0$**) = 0.03581.**

Next, we want to compute the 97.5% upper confidence bound for δ **with clustering considered.**

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)x0.001 = 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 $\sim$ 0 to 0.3	**(large ICC)**
medium cluster $\sim$ 0 to 0.05	**(medium ICC)**
large cluster $\sim$ 0 to 0.001	**(small ICC)**

Thus, to account for clustering, we will analyze the data as if there were no clusters, compute the ICC and the Deff, and finally we will multiply the standard error by the square root of the Deff. Donner and Klar (2000) and Reed (2004) have also discussed the impact of ICC on analysis.

Ridout et al. (1999) evaluated 20 different methods for estimating the ICC with binary outcomes and found that the kappa-type method originally proposed by Fleiss and Cuzick (1979) performed best under a variety of conditions. Zou and Donner (2004) have written an IML SAS program for estimating the ICC for binary outcomes using the Fleiss and Cuzick method.

Using the IML program of Zou and Donner –> see Appendix A, we compute the intra-cluster correlation coefficient (ICC) for our data set (refer to page 2). The ICC = 0.32.

To get the SE when clustering is considered, we multiply the SE from StatXact by the $\sqrt{Deff}$.

Deff = 1 + ($\bar{n}$ -1)(ICC), where $\bar{n}$ is the average cluster size and ICC is the intra-cluster correlation coefficient.

For this data set, $\bar{n}$ =23.3 and ICC=0.32.

Thus Deff = 1 + (23.3-1)(0.32) = 8.14.

$\sqrt{Deff}$ **=** $\sqrt{8.14}$ **= 2.85.**

$SE\left(\sqrt{Deff}\right)$ **= 0.03581(2.85) = 0.10206.**

Thus, the 97.5% upper confidence bound when clustering is considered is:
- 0.3614 + 1.96 (0.10206) = - 0.1614.

	Upper Limit
not clustered	- 0.2861
clustered	- 0.1614

Since the 97.5% upper confidence bound when clustering is considered is - 0.1614, and it is comfortably within the prescribed non-inferiority margin of + 0.50, we can claim non-inferiority.

Also notice how the effect of clustering increased the se($\pi_2 - \pi_1 - \delta_0$), from 0.03581 (not clustered) to 0.10206 (clustered). This in turn increased the upper limit from – 0.2861 to – 0.1614, but not enough to be $\geq$ + 0.05 (our non-inferiority margin).

We can also compute the p-value for this one-sided test:

The standardized test statistic is $\frac{\pi_2 - \pi_1 - \delta_0}{se(\pi_2 - \pi_1 - \delta_0)}$ **=** $\frac{0.4976 - 0.8590 - 0.050}{0.10206} = \frac{-0.4114}{0.10206} = -4.03$**.**

This gives a p-value = 0.000028 which is less than the nominal 0.025. Thus, we can reject the null hypothesis $H_0 : \delta = \delta_0$. This allows us to claim non-inferiority.

References

EMEA/CPMP/EWP/2158/99. Guideline on the choice of the non-inferiority margin. London, 27 July 2005.

McCarthy WF. The design and sample size requirement for a cluster randomized non-inferiority trial with two binary co-primary outcomes. Maryland Medical Research Institute, Technical Report. July 1, 2008.

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

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.

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

Reed JF. Adjusted chi-square statistics: application to clustered binary data in primary care. Annals of Family Medicine 2004; vol. 2, no. 3: 201-203.

Ridout MS, Demetrio CG, Firth D. Estimating intraclass correlation for binary data. Biometrics 1999; 55: 137-148.

Fleiss JL, Cuzick J. The reliability of dichotomous judgements: unequal numbers of judges per subject. Appl. Psychological Measurement 1979; 3: 537-542.

Zou G, Donner A. Confidence interval estimation of the intraclass correlation coefficient for binary outcome data. Biometrics; 60: 807-811.

Appendix A.

```
*******************************************************************************;
* Zou and Donner (2004)Confidence interval estimation of the intraclass correlation *;
* coefficient for binary outcome data Biometrics Vol 60 (3) pp 807-811          *;
*                                                                              *;
*                                EXAMPLE                                       *;
*                                                                              *;
*                        ***** Fleiss-Cuzick *****                             *;
*                                                                              *;
*          Clusters 1-17 intervention group, clusters 18-29 control group      *;
*          Cluster denoted i, patient within cluster denoted j                 *;
*          Binary outcome data: yij=1 if event, yij=0 if non-event             *;
*          yi is the number of events in cluster i                             *;
*          ni is the total number of patients in cluster i                     *;
*                                                                              *;
*                    Modified by W.F. McCarthy 7/02/2008                       *;
*******************************************************************************;


***** SAS IML codes *******;

proc iml;

start cubic(a, b, c, d);
 l=.; u=.;
  if (a = 0) then do;
         if c**2 - 4*b*d > 0 then do;
           l = (-c - sqrt(c**2 - 4*b*d))/(2*b);
           u = (-c + sqrt(c**2 - 4*b*d))/(2*b);
         end;
  end;
  else do;
   a1 = b/a; a2=c/a; a3 = d/a;
   r = a1*a2/6 - a3/2 - a1**3/27;
   q = a2/3 - a1**2/9;
      if q < 0 then do;
        f= r/sqrt(-q**3);
        if abs(f) <=1 then do;
         *if f >1 then f = 1;
         *if f <-1 then f =-1;
         theta = arcos(f);
         pi = 3.1415926;
         k = 2*sqrt(-q)*cos(theta/3)  -a1/3;
         l = 2*sqrt(-q) * cos ( (theta+2*pi)/3  ) - a1/3;
         u = 2*sqrt(-q) * cos ( (theta+4*pi)/3  ) - a1/3;
         end;
     end;
 end;
return(l||u||k);

finish cubic;



start interval4ICC(alpha, data);
 crit = probit( 1- alpha/2);
 chi = crit**2;
 yi = data[,1];
 ni = data[,2];

 bigN = sum(ni);
 piest= sum(yi)/bigN;
 piHat = piest;
 index = piest * (1 - piest);
 k = nrow(ni); * 29; *clusters;
```

```
****** Fleiss-Cuzick **********;
rhoHat = 1 - sum ( yi#(ni-yi)/ni )/((bigN- k)* piest*(1-piest));


part1 = ((1/index - 6) * sum(1/ni))/(bigN - k)**2 + ((2 *bigN
                + 4 * k - k/index) * k)/(bigN * (bigN - k)**2);
part2 = sum(ni##2)/(bigN**2 * index)-
              (3 * bigN - 2 * k) * (bigN - 2 * k) * sum(ni##2)/
              (bigN**2* (bigN - k)**2) - (2 * bigN - k)/(bigN - k)**2;
part3 = (4-1/index)*(sum(ni##2) - bigN)/bigN**2;
vFC  =  (1 - rhoHat) * (part1 + part2*rhoHat + part3*rhoHat**2);

** CUBIC ****;
A2 = - part3;
B2 = part3 - part2;
C2 = part2 - part1;
D2 = part1;


A22 =chi*A2;
B22 =chi*B2 - 1;
C22 =chi*C2 + 2* rhoHat;
D22 =chi*D2 - rhoHat**2;
solution2 = cubic(A22, B22, C22, D22);
lower = solution2[,1];
upper = solution2[,2];

  print alpha piHat rhoHat lower upper;

finish interval4ICC;


* Clusters 1-17 intervention group, clusters 18-29 control group;
* cluster denoted i, patient within cluster denoted j;
* Binary outcome data  yij=1 if event, yij=0 if non-event;
* yi is the number of events in cluster i;
* ni is the total number of patients in cluster i;

yi = {0   0  0  6  2  4  1  2  3  1  6  9  8  6 10  4  4  2  3  6  1 3  3 2  5 18 21 19 21};
ni = {30 22 19 30 30 30 30 30 30 30 30 30 22 22 32 31 20 16 11 35 10 9 11 6 12 25 25 23 24};

yi = t(yi);
ni = t(ni);

print 'Example';
data =yi||ni;
run interval4ICC(0.05, data);

quit;



/* OUTPUT from SAS Program */

/* RHOHAT is the ICC */

/* HARM IS CONSIDERED, HOWEVER THE ICC FOR BENEFIT IS THE SAME */


                    Example

   ALPHA     PIHAT      RHOHAT        LOWER       UPPER
    0.05  0.2518519  0.315556     0.1262249   0.5592528
```