Enliven Archive
Intensifying Thoughts

Research Article | www.enlivenarchive.org | Enliven: Biosimilars and Bioavailability

# M-Multiple 2x2 Crossover Block Design: An Alternative Criterion Bioequivalence Assessment

Junrong Lin*, and Shein-Chung Chow

*Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, North Carolina*

***Corresponding author**: Junrong Lin, Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC 27705, Email: junrong.lin@duke.edu

**Received Date**: 1st November 2021
**Accepted Date**: 15th December 2021
**Published Date**: 26th December 2021

**Citation**: Lin J, Chow S-C. M-Multiple 2x2 Crossover Block Design: An Alternative Criterion Bioequivalence Assessment. Enliven: Biosimilars Bioavailab. 2020; 5(1): 006.



## 1) Abstract

An alternative method for assessing average bioequivalence (ABE) using multiple crossover experiment is studied. Under a standard M-multiple sequential 2x2 crossover experiment, the test treatment is considered not equivalent to the reference if the difference of treatment effects is greater than a pre-specified limit in any two consecutive 2x2 crossover experiments (block) of the total M experiments. In this paper, power function for assessing ABE under a M 2x2 crossover this new design is studied by solving related characteristic polynomial. In addition, sample size calculation is also developed. A comparison in sample size requirement between traditional gold method and this new method is presented. It is found that when the 2 x 2 crossover block experiment is conducted 6 times with power less than 97.162%, the new method requires a smaller sample size.

## 2) Keywords: Crossover design; Bioequivalence test; Clinical trials

## 3) Introduction

In health-related research, crossover designs are widely used and accepted to health authorities, such as the United States Food and Drug Administration (FDA) [1-3], for evaluating bioequivalence. A crossover experiment is usually conducted on healthy volunteers for assessment of bioequivalence between innovative drug and its generic drug. For a standard two-sequence, two period (2 × 2) crossover experiment, qualified subjects are randomly assigned to one treatment at the first dosing period. Then, after a sufficient length of washout, the same subjects are switched to receive the other treatment. The purpose of washout is to wear off possible residual effect that might be carried from previous dosing period to the following period. One of the primary advantages for a crossover experiment is each subject serves as its own control, which removes the inter-subject variability. In some situations, a crossover experiment is necessarily carried out repeatedly for comparing the performance of treatments within the same subject. In this article, we refer to such a repeated experiment with the same crossover design as a multiple crossover experiment.

Let (TR,RT) denote a standard two-sequence, two-period (2 × 2) crossover design, where T is a test treatment and R is a reference treatment. A M-multiple 2 × 2 crossover design refers to the design of the form $(\overset{M}{\overbrace{RT|\ldots|RT}}, \overset{M}{\overbrace{TR|\ldots|TR}})$ in which (TR,RT) and (RT, TR) appears with the probability of 0.5.

In practice, bioequivalence is usually tested by a two one sided test (TOST) procedure [4].We can conclude that bioequivalence when the confidence interval is within bioequivalence limits $-\Delta_L$ and $\Delta_U$. Alternatively, in this article, we propose a test under M-multiple 2x2 crossover block design, for assessing bioequivalence. The proposed method does not depend on confidence interval but only on the number of consecutive difference of average treatment effect in each 2x2 crossover block beyond the limit.

In the next Section, the statistical model for assessing BE under a multiple crossover experiment is briefly outlined. In Section 5, an estimate of the probability of failing the experiment for a given day is obtained with a given power of 1- $\beta$. Procedures for sample size calculations for continuous and discrete cases were derived in Section 6 and 7, respectively. An example and a brief discussion are given in Section 8 and 9.

## 4) Statistical Model

Let (TR,RT) denotes a standard two-sequence, two-period (2 × 2) crossover design. That is, in the first sequence, qualified subjects are randomly assigned to receive the test treatment (T) first and then crossovered to receive the reference treatment (R) after a sufficient length of washout.

Consider a M-multiple 2 × 2 crossover experiment with n1 subjects in the first sequence and n2 in the second (Table 1).

Table 1 M-multiple 2x2 crossover experiment

| | Day1 | Day2 | … | Day M |
|---|---|---|---|---|
| $n_1$ | T R | R T | … | T R |
| $n_2$ | R T | T R | … | R T |

In each block (e.g. day), one of the two 2×2 crossover design, i.e., (TR,RT) or (RT,TR) is independently chosen with probability 0.5. To simplify the procedure, we just let different designs to be conducted in turn below.

Let $\overline{Y_T}$ and $\overline{Y_R}$ be the sample mean of the test treatment and the reference treatment respectively. Assume $\overline{Y_T} - \overline{Y_R}$ are independent identically distributed (i.i.d) among different days.

$$p = P\{|\overline{Y_T} - \overline{Y_R}| > \Delta \; in \; a \; particular \; day\}$$

If $|\overline{Y_T} - \overline{Y_R}| > \Delta$ for a pre-specified Δ, test treatment T is considered to be different from reference treatment R. If T is different from R in two consecutive days, then we consider the experiment fails and T is not equivalent to R. The objective of this article is to study the power of this type of experiment design and develop procedure for sample size calculation.

## 5) The Estimation of *P*

For testing whether two product are equivalent, the following hypotheses can be set up

$$H_0 : -\Delta \leq |Y_T - Y_R| \leq \Delta,$$

$$H_A : |Y_T - Y_R| > \Delta,$$

Where Δ is a pre-specified bioequivalence limit.
Let

$$X_d = \begin{cases} 1 \; if \; |\bar{Y}_{dT} - \bar{Y}_{dR}| > \Delta \\ 0 \; \; else, \end{cases}$$

Where d= 1,…,M. Then $X_d$ are i.i.d Bernoulli random variables with success rate *p*. The power of this experimental design is the probability that 2 consecutive 1's are observed. If we let g(m) denote the probability that no two consecutive 1's in *m* days, the power is given by 1-g(m). A deductive formula for g(m) can be obtained by considering the first two days' observations (i.e. $X_1$ and $X_2$).

$$g(m) = pq.g(m-2) + q.g(m-1)$$

The above formula shows that the probability of no two consecutive 1's are observed in *m* days, which can be divided into two parts according to $X_1$ and $X_2$. If $X_1$ = 1, then $X_2$ must be 0. Otherwise, two consecutive 1's are observed. The condition is met if and only if in the rest $m - 2$ days two consecutive 1's does not happen. The probability of no two consecutive 1's in a sequence of length $m - 2$ is $g\,(m - 2)$. Finally, this part of probability is given by

$$P(X_1 = 1, X_2 = 0, no \; two \; consequetive 1' \sin X_t, 3 \leq t \leq m) = pq.g(m-2)$$

When $X_1$= 0, no two consecutive 1's is in the whole sequence if and only if no two consecutive 1's in the rest $m - 1$ days.

This probability is $q.\,g\,(m - 1)$. Thus, (1) holds.
Let

$$s = \frac{q - \sqrt{q^2 + 4pq}}{2} \qquad (2)$$

$$t = \frac{q - \sqrt{q^2 + 4pq}}{2}$$

From (1), if can be verify that

$$\frac{g(m) - tg(m-1)}{g(m-1) - tg(m-2)} = s$$

and

$$\begin{aligned} g(m) - tg(m-1) &= s^{m-1}\big(g(1) - tg(0)\big) \\ &= s^{m-1}(1-t) \\ &= s^{m-1} - ts^{m-1} \end{aligned} \qquad (3)$$

Reformulate equation 3, we have

$$g(m) - tg(m-1) = t^{m-1}(1-t)\left(\frac{s}{t}\right)^{m-1}$$

$$tg(m-1) - t^2 g(m-2) = t^{m-1}(1-t)\left(\frac{s}{t}\right)^{m-2}$$

$$\ldots = \ldots$$

$$t^{m-1}g(1) - t^m g(0) = t^{m-1}(1-t)$$

$$g(m) - t^m g(0) = t^{m-1}(1-t)\frac{1 - (s/t)^m}{1 - s/t}$$

$$= \frac{1-t}{s-t}s^m + \frac{1-s}{t-s}t^m$$

The last equation is in form of roots of characteristic polynomial.
Thus, the power is

$$1 - \beta = 1 - g(m) = 1 - t^m - t^{m-1}(1-t)\frac{1-(s/t)^m}{1-s/t} \qquad (4)$$

where t and s are given in equation (2).

From equation (4), it can seen that it is a strictly monotone increasing of *m* and *p*. Thus, once *p* is determined, we can choose appropriate sample size to achieve the desired *p*.

For more general cases, suppose the probability that there is no k consecutive 1's observed in a Bernoulli sequence with length *m* is *g* (*m*), which can be calculated by a similar deductive formula: [5]

$$g(m) = \sum_{i=0}^{k-1} p^i q g(m-i-1) \qquad (5)$$

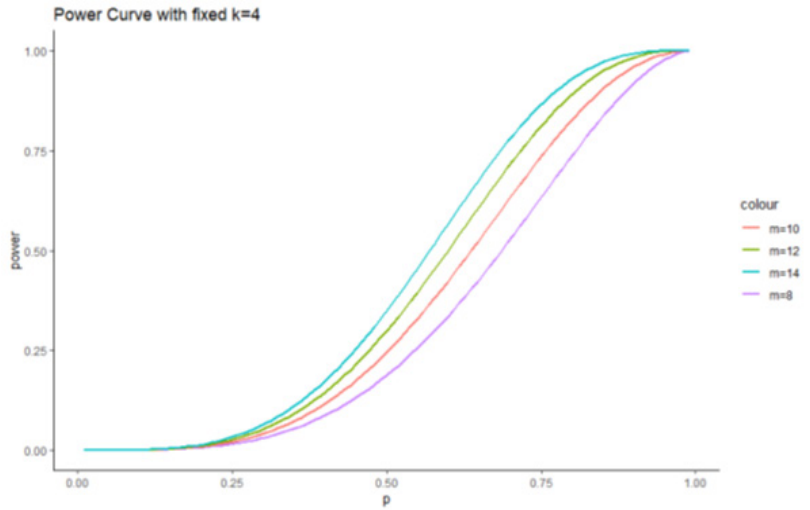
with initial conditions $g\,(j) = 1, \forall\, 0 \leq j \leq k-1$.

Note that in this general case, equation (5) is a homogeneous linear recurrence equation with constant coefficients (C-recursive equation)5 of order *k*. Explicit solution can be found up to degree 4 by directly finding the roots of characteristic polynomial.

When $k \geq 5$, no explicit expression for $g(m)$ can be found.

When $k \geq 5$, no explicit expression for $g(m)$ can be found As discussed above, when the number of consecutive tolerable days $k$ and the number of total experiment days $m$ are fixed, $1 - g(m)$ is an increasing function of p. Thus, for desired power $1 - \beta$, p can be obtained by a binary search. Figure 1 plots power versus $p$ in case $k = 4$ and $m = 8, 10, 12, 14$. It's clear that $1 - g(m)$ is an increasing function in p and m. Figure 1 below illustrates the case when $k = 4$.

Figure 1 The power curve corresponding to $k = 4$, $m = 8, 10, 12, 14$




## 6) Within Block Analysis

Without assuming $p$ directly, for a given day, assuming $n_1$ subjects in sequence 1 and $n_2$ subjects in sequence 2, consider the following model when a M-multiple M 2 × 2 crossover experiment in a given day,

$$y_{ijk} = \mu + P_k + Q_i + F_{(i,k)} + S_{ij} + e_{ijk}$$

where

- $\begin{cases} sequence & i = 1,2 \\ subject\ in\ sequence & j = 1, \ldots, n_i \\ period & k = 1,2 \end{cases}$
- $\mu$: *the overall mean*
- $P_k$: *the fixed effect of kth period and* $\sum P_k = 0$;
- $Q_i$: *the fixed effect of the ith sequence* $Q_1 + Q_2 = 0$;
- $S_{ij}$: *the random effect of the jth subject in the ith sequence;*
- $e_{ijk}$: *the within subject random error observing* $y_{ijk}$ *and* $e_{ijk} \sim N(0, \sigma^2)$
- $F_{ik}$: *the direct fixed of the formulation in the ith sequence which is administrated at the kth period and* $\sum F_{(i,k)} = 0$

| $n_1$ | T | R |
|---|---|---|
| $n_2$ | R | T |

For a given day with the following experimental sequence (referred as design α)
we have,

$$\bar{Y}_T = \frac{n_1 \bar{Y}_{T1} + n_2 \bar{Y}_{T2}}{n_1 + n_2}$$

where

$$\bar{Y}_{T1} = \mu + Q_1 + P_T + F_{T1} + \frac{1}{n_1}\sum_{i=1}^{n_1} S_{i1} + \frac{1}{n_1}\sum_{i=1}^{n_1} ei11$$

And

$$\bar{Y}_{T2} = \mu + Q_2 + P_T + F_{T2} + \frac{1}{n_2}\sum_{i=1}^{n_1} S_{i2} + \frac{1}{n_1}\sum_{i=1}^{n_1} ei22$$

Then the average test treatment effect

$$\bar{Y}_T = \mu + \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} + \frac{n_1 Q_1 + n_2 PQ_2}{n_1 + n_2} + \frac{n_1 F_{T1} + n_2 F_{T2}}{n_1 + n_2} + \frac{\sum_{i=1}^{n_1} S_{i1} + \sum_{i=1}^{n_2} S_{i2}}{n_1 + n_2} + \frac{\sum_{i=1}^{n_1} e_{i11} + \sum_{i=1}^{n_2} e_{i22}}{n_1 + n_2}$$

Similarly, the average reference treatment effect

$$\bar{Y}_R = \mu + \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} + \frac{n_1 Q_1 + n_2 PQ_2}{n_1 + n_2} + \frac{n_1 F_{R1} + n_2 F_{R2}}{n_1 + n_2} + \frac{\sum_{i=1}^{n_1} S_{i1} + \sum_{i=1}^{n_2} S_{i2}}{n_1 + n_2} + \frac{\sum_{i=1}^{n_1} e_{i21} + \sum_{i=1}^{n_2} e_{i12}}{n_1 + n_2}$$

Thus

$$\bar{Y}_T - \bar{Y}_R = \frac{n_1 - n_2}{n_1 + n_2}(P_1 - P_2) + \frac{n_1}{n_1 + n_2}(F_{T1} - F_{R1}) + \frac{n_2}{n_1 + n_2}(F_{T2} - F_{R2}) + \frac{1}{n_1 + n_2}\sum e_{ijk} \quad (6)$$

Similarly, if the alternative design $\beta$ is conducted, that is

| $n_1$ | R | T |
|---|---|---|
| $n_2$ | T | R |

the equation (6) becomes

$$\bar{Y}_T - \bar{Y}_R = \frac{n_1 - n_2}{n_1 + n_2}(P_2 - P_1) + \frac{n_2}{n_1 + n_2}(F_{T1} - F_{R1}) + \frac{n_1}{n_1 + n_2}(F_{T2} - F_{R2}) + \frac{1}{n_1 + n_2}\sum e_{ijk}$$

Therefore,

$$E(\bar{Y}_T - \bar{Y}_R) = E[E(\bar{Y}_T - \bar{Y}_R | \, design)]$$
$$= \frac{1}{2}E(\bar{Y}_T - \bar{Y}_R | \, design\ \alpha) + \frac{1}{2}E(\bar{Y}_T - \bar{Y}_R | \, design\ \beta)$$
$$= 0.5\,(F_{T1} - F_{R1} + F_{T2} - F_R) \coloneqq \delta$$

is an unbiased estimator of the direct formulation effects ($\delta$). Then we can note that

$$\bar{Y}_T - \bar{Y}_R =_d \begin{cases} N\left(a + \delta_1, 2\sigma^2/n_1\right) & w.p.\ 1/2 \\ N\left(-a + \delta_2, 2\sigma^2/n_2\right) & w.p.\ 1/2 \end{cases}$$

where $=_d$ denotes "distributed as", "w.p." is "with probability",

$$a = \frac{n_1 - n_2}{n_1 + n_2}(P_1 - P_2) \qquad (7)$$

And

$$\delta_1 = \frac{n_1}{n_1 + n_2}(F_{T1} - F_{R1}) + \frac{n_2}{n_1 + n_2}(F_{T2} - F_{R2})$$
$$\delta_2 = \frac{n_2}{n_1 + n_2}(F_{T1} - F_{R1}) + \frac{n_1}{n_1 + n_2}(F_{T2} - F_{R2})$$

When there is no random formulation effect: $\delta 1 = \delta 2 = 0$. The probability of committing the type I error is given by

$$\frac{1}{2}P\left(\left|N\left(a, \frac{2\sigma^2}{n_1}\right)\right| > \Delta\right) + \frac{1}{2}P\left(\left|N\left(-a, \frac{2\sigma^2}{n_2}\right)\right| > \Delta\right)$$

From this expression we can see that the probability of committing the type I error could be very big, if $|a|$ is far from 0. In what follows, statistical method for sample size calculation under various situation (i.e., $a = 0$) are proposed to reduce the probability of committing the type I error.

## 6.1) When There Are No Period Effects

A carefully designed experiment can eliminate period effects (i.e., $p1 = p2 = 0 \Rightarrow a = 0$). In this case, we have

$$\bar{Y}_T - \bar{Y}_R =_d \begin{cases} N\left(\delta_1, 2\sigma^2/n_1\right) & w.p.\ 1/2 \\ N\left(\delta_2, 2\sigma^2/n_2\right) & w.p.\ 1/2 \end{cases}$$

Then, the power function can be obtained as

$$P(|\bar{Y}_T - \bar{Y}_R| > \Delta) = \frac{1}{2}P\left(\left|N\left(\delta_1, \frac{2\sigma^2}{n_1}\right)\right| > \Delta\right) + \frac{1}{2}P\left(\left|N\left(\delta_2, \frac{2\sigma^2}{n_2}\right)\right| > \Delta\right)$$
$$= \frac{1}{2}P\left(Z < -\frac{\Delta + \delta_1}{\sigma}\sqrt{\frac{n_1}{2}}\right) + \frac{1}{2}P\left(Z > \frac{\Delta - \delta_1}{\sigma}\sqrt{\frac{n_1}{2}}\right) + \frac{1}{2}P\left(Z < -\frac{\Delta + \delta_2}{\sigma}\sqrt{\frac{n_2}{2}}\right)$$
$$\frac{1}{2}P\left(Z > \frac{\Delta + \delta_2}{\sigma}\sqrt{\frac{n_2}{2}}\right)$$

## 6.2) Equal Number of Subjects in Each Sequence

From the expression for $a$ (5), if $n_1 = n_2 = n$, in this case,

$$\bar{Y}_T - \bar{Y}_R \sim N\left(\delta, \frac{2\sigma^2}{n}\right)$$

The power function is given by

$$P\left(\left|N\left(\delta, \frac{2\sigma^2}{n}\right)\right| > \Delta\right) = P\left(Z < -\frac{\Delta + \delta}{\sigma}\sqrt{\frac{n}{2}}\right) + P\left(Z > \frac{\Delta - \delta}{\sigma}\sqrt{\frac{n}{2}}\right)$$

Under the null hypothesis that $-\Delta \le \delta \le \Delta$, where $\delta > 0$, the probability $P\left(Z < -\frac{\Delta + \delta}{\sigma}\sqrt{n/2}\right)$ may be negligible for a large n. Thus for a given power of $1 - \beta$, a conservative half sample size estimation can be get by solving

$$P\left(Z > \frac{\Delta - \delta}{\sigma}\sqrt{n/2}\right) = 1 - \beta$$

This leads to

$$n = 2\frac{z_{1-\beta}^2 \sigma^2}{(\delta - \Delta)^2}$$

## 6.3) Alternative Estimates for $\mu_T$ and $\mu_R$

Suppose for a given day, the sequences of a crossover experiment is as follows

| | | |
|---|---|---|
| **$n_1$** | **T** | **R** |
| **$n_2$** | R | T |

If we define $\bar{Y}_T = \frac{1}{2}(\bar{Y}_{T1} + \bar{Y}_{T2})$ and $\bar{Y}_R = \frac{1}{2}(\bar{Y}_{R1} + \bar{Y}_{R2})$, we have

$$\bar{Y}_{T1} = \mu + \frac{1}{2n_1}\sum S_{i2} + \frac{1}{2n_1}\sum S_{i2} + \frac{1}{2}(F_{T1} + F_{T2}) + \frac{1}{2n_1}\sum e_{i11} + \frac{1}{2n_2}\sum e_{i22}$$

And

$$\bar{Y}_R = \mu + \frac{1}{2n_1}\sum S_{i1} + \frac{1}{2n_2}\sum S_{i2} + \frac{1}{2}(F_{R1} + F_{R2}) + \frac{1}{2n_1}\sum e_{i21} + \frac{1}{2n_2}\sum e_{i12}$$

Then

$$\bar{Y}_T - \bar{Y}_R = \delta + \frac{1}{2n_1}\sum (e_{i11} - e_{i21}) + \frac{1}{2n_2}\sum (e_{i22} - e_{i12})$$

On the other hand, if the design is

| | | |
|---|---|---|
| **$n_1$** | **R** | **T** |
| **$n_2$** | T | R |

then we will have

$$\bar{Y}_T - \bar{Y}_R = \delta + \frac{1}{2n_1}\sum (e_{i21} - e_{i11}) + \frac{1}{2n_2}\sum (e_{i12} - e_{i22})$$

In this case, for a given day, we have

$$\bar{Y}_T - \bar{Y}_R \sim N\left(\delta, \left(\frac{1}{2n_1} + \frac{1}{2n_2}\right)\sigma^2\right)$$

regardless of the design (either design $\alpha$ or design $\beta$) used. As a result,

$$P\left(\left|N\left(\delta, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\delta^2\right)\right| > \Delta\right) = P\left(Z < -\frac{\Delta + \delta}{\sigma}\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\right) + P\left(Z > \frac{\Delta + \delta}{\sigma}\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\right)$$

## 7) Binary Outcome

In this section, for simplicity, we assume that there are no carryover effects and formulation effects across days. Suppose our response yijT is a Bernoulli random variable associated with the jth subject in the ith sequence under treatment (T). yiR is similarly defined. Suppose $p_{it} = P(y_{ijT} = 1)$ and $p_{iR} = P(y_{ijR} = 1)$. If two responses come from two different subjects, they are assumed independent. This can happen because two responses come from two different subjects within the same sequence (having different j but same i) or two subjects come from different sequence (having different *i*, but

may same *j*). But if two responses come from the same subject by different formulation (T or R), then may not be independent. Suppose the crossover design for a given day is

| | | |
|---|---|---|
| $n_1$ | T | R |
| $n_2$ | R | T |

Suppose $p_{iT} = P(y_{ijT} = 1)$ *and* $p_{iR} = P(y_{ijR} = 1)$. Then $\bar{Y}_T - \bar{Y}_R$ can be expressed as

$$\bar{Y}_T - \bar{Y}_R = \frac{1}{n_1 + n_2}[(n_1 \bar{Y}_{T1} + n_2 \bar{Y}_{T2}) - (n_1 \bar{Y}_{R1} + n_2 \bar{Y}_{R2})]$$
$$= \frac{1}{n_1 + n_2}[n_1(\bar{Y}_{T1} - \bar{Y}_{R1}) + n_2(\bar{Y}_{T2} - \bar{Y}_{R2})]$$
$$= \frac{1}{n_1 + n_2}\left[\sum_{j=1}^{n_1}(y_{1jT} - y_{1jR}) + \sum_{j=1}^{n_2}(y_{2jT} - y_{2jR})\right]$$

Then $(y_{1jT} - y_{1jR})$ and $(y_{2jT} - y_{2jR})$ are independent. When n1 and n2 are sufficiently large, central limit theorem can be applied. As a result,

$$\bar{Y}_T - \bar{Y}_R \sim \frac{n_1}{n_1 + n_2} N(\mu_1, \sigma_1^2/n_1) + \frac{n_2}{n_1 + n_2} N(\mu_2, \sigma_2^2/n_1) \sim N(\mu^*, \sigma^*)$$

Where

$$\mu_k = E(y_{ijT} - y_{ijR}) = p_{iT} - p_{iR}$$

$$\sigma_k^2 = V\,ar(\bar{Y}_{ijT} - \bar{Y}_{ijTk})$$

And

$$\sigma^{*2} = \frac{n_1\mu_1 + n_2\mu_2}{n_1 + n_2}$$

$$\sigma^{*2} = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{(n_1 + n_2)}$$

Thus, the sample size can be determined by solving

$$P(|N(\mu^*, \sigma^{*2})| > \Delta) = 1 - \beta$$

Note that when $n_1 = n_2$, we have

$$\bar{Y}_T - \bar{Y}_R \sim N\left(\frac{\mu_1 + \mu_2}{2}, \frac{\sigma_1^2 + \sigma_2^2}{4n}\right)$$

This leads to

$$n = \frac{z_{1-\beta}^2(\sigma_1^2 + \sigma_2^2)}{[2\Delta - (p_{1T} - p_{1R} + p_{2T} - p_{2R})]^2}$$

## 8) An Example

In practice, a multiple crossover experiment is often conducted in mice studies. For example, we may of interest to evaluate for evaluation of the odor intensity from animal housing hardware. The purpose of the experiment is to verify that a newly designed animal housing hardware or an animal housing hardware with a new filter for the space shuttle will not contain odors that are perceived as strong, unpleasant or annoying for a time period of 1 to 21 days. As a result, a M-multiple crossover experiment for evaluation of odor from animal housing hardware is usually conducted on daily basis for a consecutive 21 days. On each days, housing hardware with an old filter or with new filter, which contains no animals or a different number of animals (e.g. 3, 6, 9 or 12 mice) is assessed using a 5-scale odor intensity rating (i.e., 0 for none, 1 for barely detectable, 2 for weak, 3 for moderate, 4 for strong and very strong). If a statistically significant difference between the housing hardware with the old filter and the housing hardware with the new filter is detected for two consecutive days, then we claim that the new filter is not equivalent to old filter, and hence conclude that the newly designed filter for the housing hardware fails.

### 8.1) Sample Size Calculation Compared with Traditional Gold Method

In traditional test for equivalence in a 2x2M crossover design, response in each sequence and each period are summed up to do a single t-test. Chow and Jun [6] gives a conservative approximation of sample size *n* which can be done since the actual power is greater than

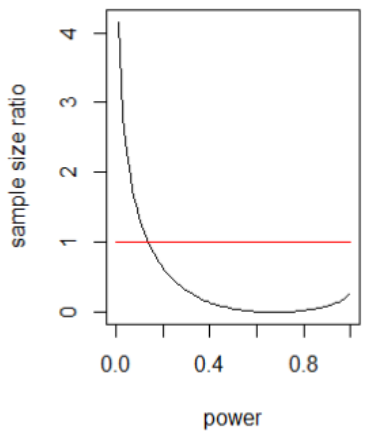
$$1 - 2T_{2n-2}\left(t_{\alpha,2n-2}|\frac{\sqrt{2n}(\Delta - |\varepsilon|)}{\sigma_m}\right)$$

Where Δ is the limit, ε is truth difference in treatment effect and $(\sigma_m^2/2n$ is the variance of the unbiased estimator of mean. When n is large, the resulting sample size to achieve desired power is

$$n = \frac{\left(z_\alpha + z_{\beta/2}\right)^2 \sigma_m^2}{2(\delta - |\varepsilon|)^2}$$

Compared with the sample size requirement above, set $k = 2$ and $m = 6$ in Section 5. Let $\alpha = 0.05$. This property holds whenever power ≥ 0.138 while keep the same power. Figure 2 shows the ratio of new sample size requirement to the old method.

Figure 2 Comparison of sample size requirement, black curve is the ratio of the sample size requirement of new method to traditional method. Red horizontal line is ratio=1




## 9) Discussion

Crossover experiment is a type of model that have been used widely in clinical research. In this article, we propose an alternative criterion/method for evaluation of treatment effect under a M-multiple 2x2 crossover design. Power analysis for sample size calculation under the M-multiple 2x2 crossover design was also developed. It is suggested that the proposed method for bioequivalence assessment under multiple 2x2 crossover design be used in the following situations:

1. The primary endpoint is a continuous variable, which is discussed in section 6. Procedure for sample size calculation is also given. Note that period effect is also permitted.

2. For a long term experiment, especially when there is a large number of periods. In this case, subjects often cannot complete the entire trial. In addition, a serious imbalance cause (i.e. $n1 \neq n2$) may occur. The proposed method described in section 6.3 is useful for obtaining valid results.

3. Note that the proposed criterion/method does not require the calculation of confidence interval to conclude bioequivalence. The proposed criterion/method is mathematically easy and user friendly.

## References


1. FDA. Guidance for Industry on Statistical Approaches to Establishing Bioequivalence. Center for Drug Evaluation and Research, Rockville: Food and Drug Administration; 2001.

2. FDA. Guidance on Statistical Procedures for Bioequivalence Studies Using a Standard Two-Treatment Crossover Design; Office of Generic Drugs, Center for Drug Evaluation and Research, Rockville: Food and Drug Administration; 1992.

3. FDA. Guidance for Industry: Bioavailability and Bioequivalence Studies for Orally Administered Drug Products—General Considerations; Center for Durg Evaluation and Research, Rockville: Food and Drug Administration; 2002.

4. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J Pharmacokinet Biopharm. 1987, 15: 657-680.

5. Rosen KH. Discrete Mathematics and Its Applications 7th. New York: McGraw Hill Education; 2011.

6. Chow SC, Jun S. Sample Size Calculation in Clinical Research 3rd. New York: Chapman & Hall/CRC Press; 2018. 510.

7. Senn S. Cross-over Trials in Clinical Research. New York: John Wiley and Sons; 1993.

8. Chow SC, Liu JP. Design and Analysis of Bioavailablility and Bioequivalence Studies 2nd. New York: Marcel Dekker; 1999.

9. Wang H, Chow SC. On statistical power for average bioequivalence testing under replicated crossover designs. J Biopharm Stat. 2002, 12: 295-309.

10. Chow SC, Wang H. On sample size calculation in bioequivalence trials. J Pharmacokinet Pharmacodyn. 2001, 28: 155-169.

11. Wang H, Chow SC. A practical approach for comparing means of two groups without equal variance assumption. Stat Med. 2002, 21: 3137-3151.

12. Chow SC, Shao J, Wang H. Individual bioequivalence testing under 2x3 designs. Stat Med. 2002, 21: 629-648.