國立臺灣大學生物資源暨農學院農藝學研究所

生物統計學組碩士論文

Division of Biometry Graduate Institute of Agronomy

College of Bioresources and Agriculture

National Taiwan University

Master Thesis

尺度平均對等性檢定於基因改造產品統計評估之研究

A Study on Applications of the Scaled Average Equivalence Test
to Statistical Evaluation of Genetically Modified Products

朱亭甄

Ting-Chen Chu

指導教授：劉仁沛 博士

Advisor: Jen-Pei Liu, Ph.D.

中華民國 102 年 6 月

June, 2013

誌謝

在台大農藝所生統組碩士班的這兩年讓我受益良多，由衷感謝我的指導老師劉仁沛教授。自大學部開始加入劉老師的研究室之後，即感受到老師的教學熱忱和對學生的無私奉獻，進入碩士班之後，更是受到劉老師不辭辛勞的指導與孜孜不倦的叮嚀，使我能順利完成碩士班學業並且完成碩士論文，並且給予了我不僅在學業上，甚至在為人處事這方面上學習、成長了許多，再次誠摯感謝劉仁沛老師!也感謝兩位口試委員 林志榮 助理教授和 季瑋珠 教授，感謝您們給予我的建議和指導，使我的碩士論文更為完善!

而在碩士班的兩年生涯中，因為有了研究室學長姐、同儕、以及學弟妹們的關懷與照顧，使我感受到研究室就像一個大家庭一般溫馨。感謝姜杰學長總是很願意幫助學弟妹解決課業上的問題，感謝瑱芳學姐、詩婷學姐總會貼心得提醒我們許多事項，感謝同儕育誠、松林、智揚，因為有了你們一起打拚的好夥伴，使我的碩士生涯不孤單，也感謝宇晴、力維、証群、家康在平時熱心的幫忙!

最後，我要感謝我的家人，感謝爸爸媽媽的支持與栽培，我才能完成台大碩士學業，感謝兩位哥哥在平時不斷給予我鼓勵，並且在我人生的道路上指引方向!

<div style="text-align: right">

亭甄 謹誌於

國立台灣大學農藝學研究所生物統計組

中華民國壹零貳年陸月

</div>

摘要

根據歐盟委員會 (European Commission)，植物或動物經由人工進行基因轉植之後，使其擁有新的特性或性質者，稱為基因改造生物 (genetically modified organisms, GMOs)。然而，隨著基改作物的發展，基改作物的安全性評估成為一大重要議題。基改作物的成分、基因型、外表型皆需要接受評估與鑑定。

歐盟食品安全管理局 (European Food Safety Authority, EFSA) 提出了針對基因改造作物安全性的統計考量評估方法。基改作物團隊 (GMO Panel) 建議同時利用差異性檢定 (proof of difference) 以及對等性檢定 (proof of equivalence) 來進行評估，並建立其信賴區間 (confidence interval)與對等性限度 (equivalence limit) 。

歐盟食品安全管理局 (European Food Safety Authority, EFSA) 對於安全性評估之科學統計考量建議，基改作物與其傳統作物之平均差異值應落於商業品種之自然變異中。因此，我們提出一項新的標準，亦即利用基改作物與其相對應之傳統作物的平均平方差異，與商業品種變異的比例，來進行基改作物安全性之對等性評估。

在混和效應的模式 (mixed-effects model) 之下，我們採用大樣本修正法 (Modified Large-Sample Method, MLS) 建立線性化指標的 95% 信賴上限，作為基改作物與其相對應傳統作物的安全性對等性評估之檢定程序。最後利用模擬資料，以經驗型 I 錯誤率和經驗檢定力來評估此一檢定程序。並以一個例子介紹本方法之應用。

關鍵字：歐盟食品安全管理局，大樣本修正法，基因改造作物，對等性限度，95% 信賴上限

Abstract

According to the European Commission, genetically modified organisms (GMOs) are organisms, such as plants and animals, whose genetic characteristics are being modified artificially in order to give them a new property. However, with the development of genetically modified organisms, the assessment of GMOs for safety is an important issue. The components, genotype and phenotype of GM plants should be identified and evaluated.
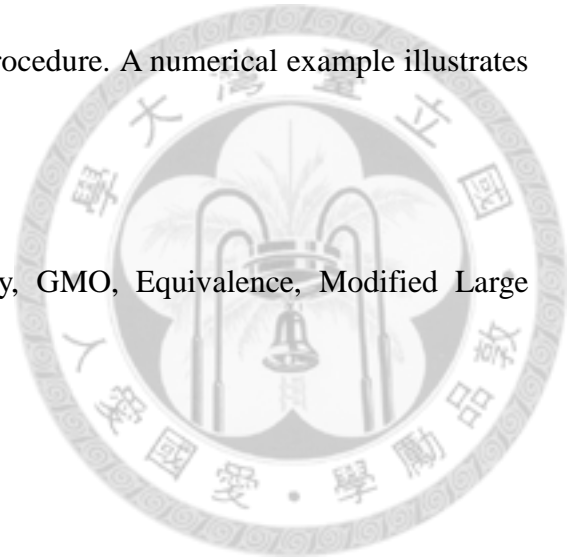
The Panel on GMOs of the European Food Safety Authority (EFSA) issued a guideline of scientific opinion on the statistical considerations for safety evaluation of GMOs in 2010. The GMO Panel of the EFSA indicated that both the proof of difference and the proof of equivalence are required for evaluation of the safety of GMOs.

The EFSA's scientific opinion on statistical considerations for the safety evaluation suggests that the difference in average between the GM crop and its conventional counterpart lie within the natural variability which can be estimated from the commercial varieties. Therefore, we propose a new criterion for assessment of equivalence of the safety profile between the GM crop and conventional crop which is the scaled square mean difference between the GM crop and its conventional crop with the variance of commercial crops as the scaled factor.
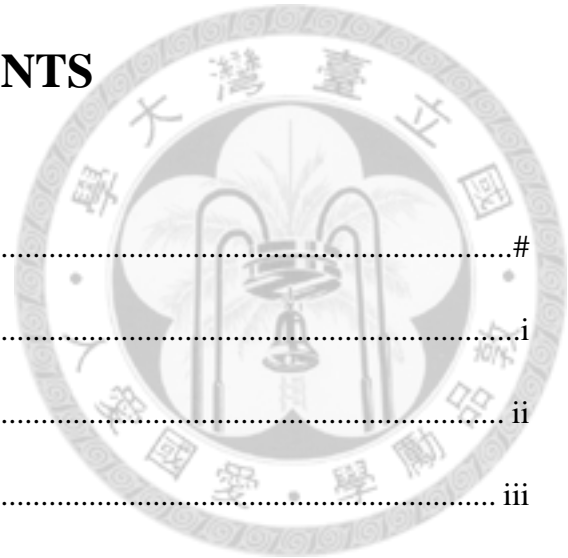
Under the mixed-effects model, we applied the modified large-sample (MLS) method to derive the 95% upper confidence limit of the linearized criterion as the testing procedure for evaluation of equivalence in the safety profile between the GMO and its conventional crop. A simulation was conducted to investigate the performance,

in terms of size and power, for the proposed procedure. A numerical example illustrates the applications of the proposed method.

Key words: European Food Safety Authority, GMO, Equivalence, Modified Large Sample Method, 95% upper confidence limit
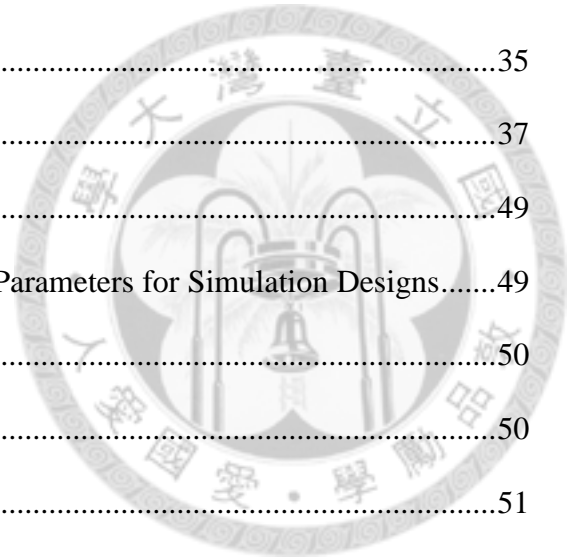
CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction

1.1 Global Development of Genetically Modified Crops
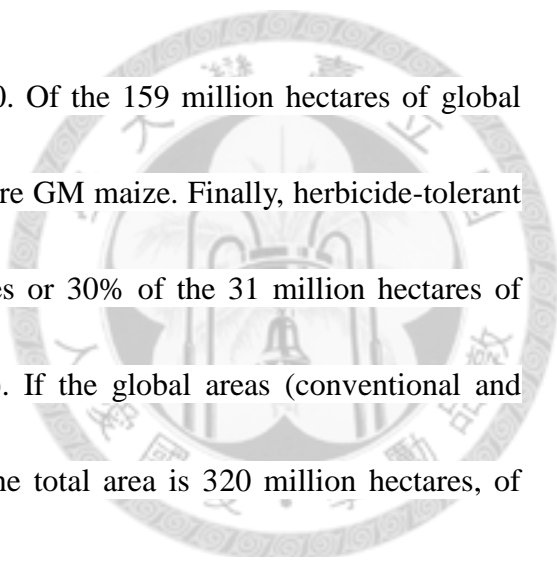
Genetically modified (GM) crops are those organisms what genetic materials are modified in the way through gene technology, rather than based on the natural proliferation or natural recombination under the definition of The Food Standards Committee (Codex) of United Nations Food and Agriculture Organization (FAO) and World Health Organization (WHO), and European Associates and Regulations (WHO, 2002) (EU Directive 2001/18/EC). With transgenic technology such as DNA recombination, GM crops were inserted with target genes in order to develop biotic or abiotic stress-tolerant traits. The purpose is to improve the agronomic traits, to reduce production costs, to increase production, to improve quality, or will increase the nutrients, enhance the effectiveness of the farmer farming households, and give the new value of the crop. Improved agronomic traits include insect-resistance, disease-resistance, cold-tolerance, salt-tolerance, herbicide-resistance, and other kinds of resistance. The mainstream GM traits are herbicides-resistant and insect-resistant traits. In recent years, GM crops are mainly of multiple traits hybrid (stacked traits).

Since GM crops were first commercialized in the U.S. in 1996, millions of

farmers worldwide have made decisions to plant and replant crops featuring the technology on an accumulated area of GM crops plantation. In 2012, the global area of GM crops continued to increase for the 17th year at a sustained growth rate of 6% or 10 million hectares (25 million acres) and to reach 170.3 million hectares or 420 million acres (Figure 1.1) (James, 2012). International Service for the Acquisition of Agri-Biotech Applications (ISAAA) data indicates that, U.S. farmers continued to plant more GM crops than any country in the world in 2011 – a total of almost 70 million hectares or 170 million acres, of which half is in the maize area, and two thirds of cotton had more than one trait, generating multiple benefits. Of the countries that had adopted biotech crops in 2012, 20 were developing countries and 8 were industrialized nations. Figure 1.2 shows the global areas of GM crops by types of countries (James, 2012). China and India lead Asian adoption, Brazil and Argentina lead Latin American adoption, South Africa leads adoption on the continent of Africa. A growth rate for GM crops in developing countries at 11 percent, or 8.7 million hectares during 2012, was significantly stronger than industrial countries at 3 percent or 1.6 million hectares.

In 2012, 81% (80.7 million hectares) of the 100 million hectares of the soybean planted globally were GM crops (Figure 1.3). GM cotton was planted to 24.3 million hectares, 81% of the 30 million hectares of global cotton, an increase from the 21.0
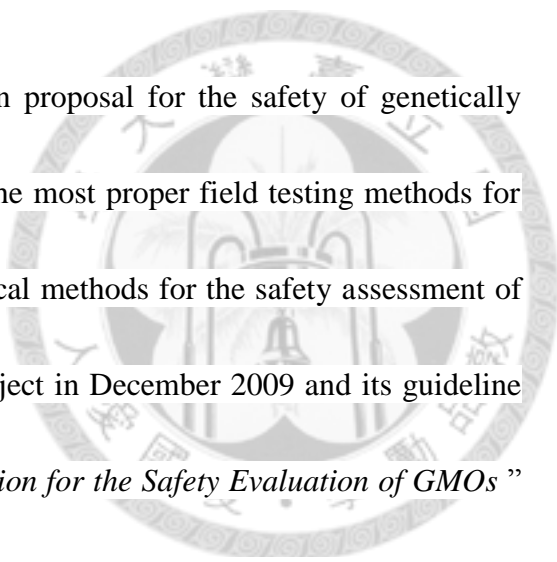
million hectares of GM cotton planted in 2010. Of the 159 million hectares of global maize planted in 2011, 35% or 55.1 million were GM maize. Finally, herbicide-tolerant GM canola was planted in 9.2 million hectares or 30% of the 31 million hectares of canola grown globally in 2012 (James, 2012). If the global areas (conventional and biotech) of these four crops are aggregated, the total area is 320 million hectares, of which 53% or 170.3 million hectares were GM crops, up from 50% in 2011.

1.2 Issues of GM Crops

In spite of benefits of GM crops mentioned above, there are serious doubts whether genetically modified crops are harmful to human, including that if transgenic genes affect human bodies, if the transgenic crops metabolites would endanger human health, and whether GM crops would impact on the environment. GM crops remain controversial because they have not been adequately tested by independent scientists. Most of data that are the basis for government's approval for GM crops are conducted by scientists who work either directly or indirectly for biotech companies. Moreover the data is confidential and not available for objective evaluation by independent and credible experts.

The Genetic Modified Organism Panel (GMO Panel) of the European Food

Safety Authority (EFSA) developed evaluation proposal for the safety of genetically modified crops in 2009. EFSA came up with the most proper field testing methods for safety assessment, as well as the use of statistical methods for the safety assessment of GM crops. GMO Panel has completed this project in December 2009 and its guideline on scientific opinion on "*Statistical Consideration for the Safety Evaluation of GMOs*" has been published in 2010. It is not only the guideline for safety assessment of GM crops, but also the basic for the standardized legal formulation of GM crops by the European Union.

The EFSA scientific opinion on statistical consideration for the safety evaluation suggests the design of experiment for safety assessment, equivalence hypothesis in conjunction with the difference hypothesis, the use of natural variability of commercial varieties for establishment of equivalence limits. In addition, the EFSA's scientific opinion proposes that the difference in average between the GM crop and its conventional counterpart should lie within the natural variability of the commercial varieties.

Following the EFSA's scientific opinion, we propose a new criterion for assessment of equivalence of the safety profile between the GM crop and conventional crop, which is the scaled square mean difference between the GM crop and its

conventional crop with the variance of commercial crops as the scaled factor. Then we suggest using the 95% upper confidence bound as the testing procedure for evaluation of the equivalence between GM crop and its conventional counterpart.

Next chapter will review the statistical methods including designs recommended by the EFSA's scientific opinion. Our proposed criterion and inference procedure are introduced in Chapter 3. Numerical examples are provided in Chapter 4 to illustrate our proposed approach and the EFSA's method. Simulation results are reported in Chapter 5. Chapter 6 provides the discussion and conclusions.
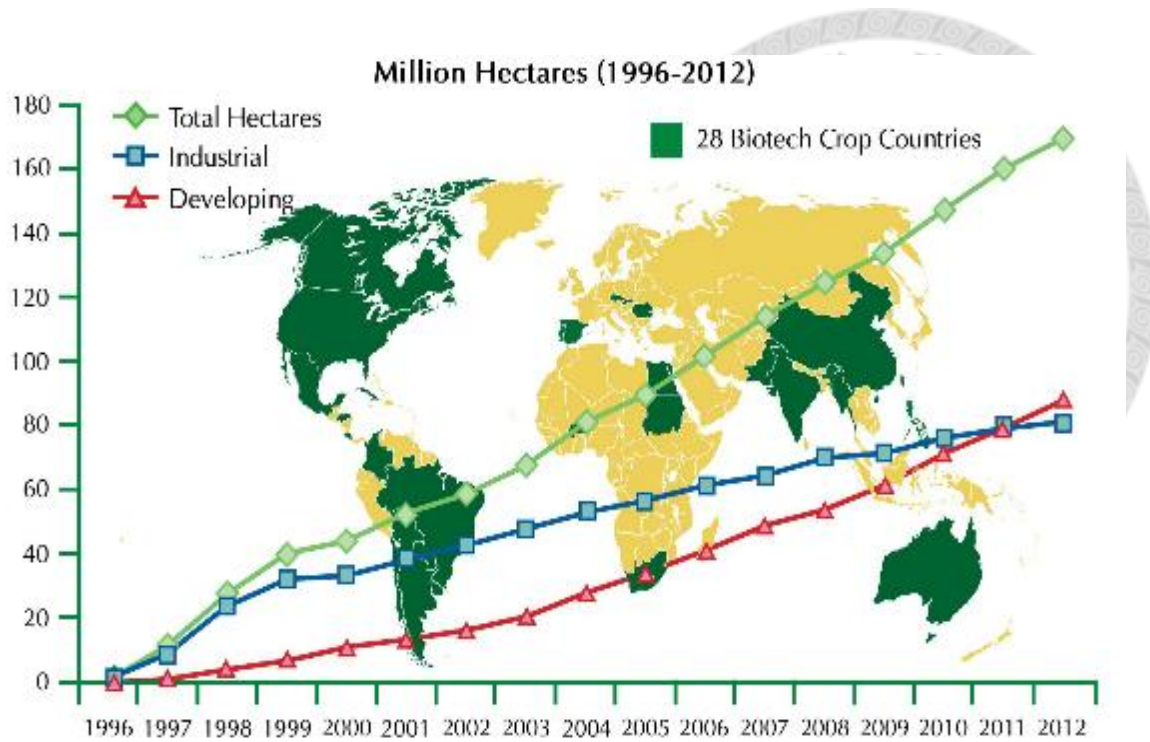
Figure 1.1 Global Area of Biotech Crops, 1996 to 2012 (Million Hectares)
Source: James (2012).
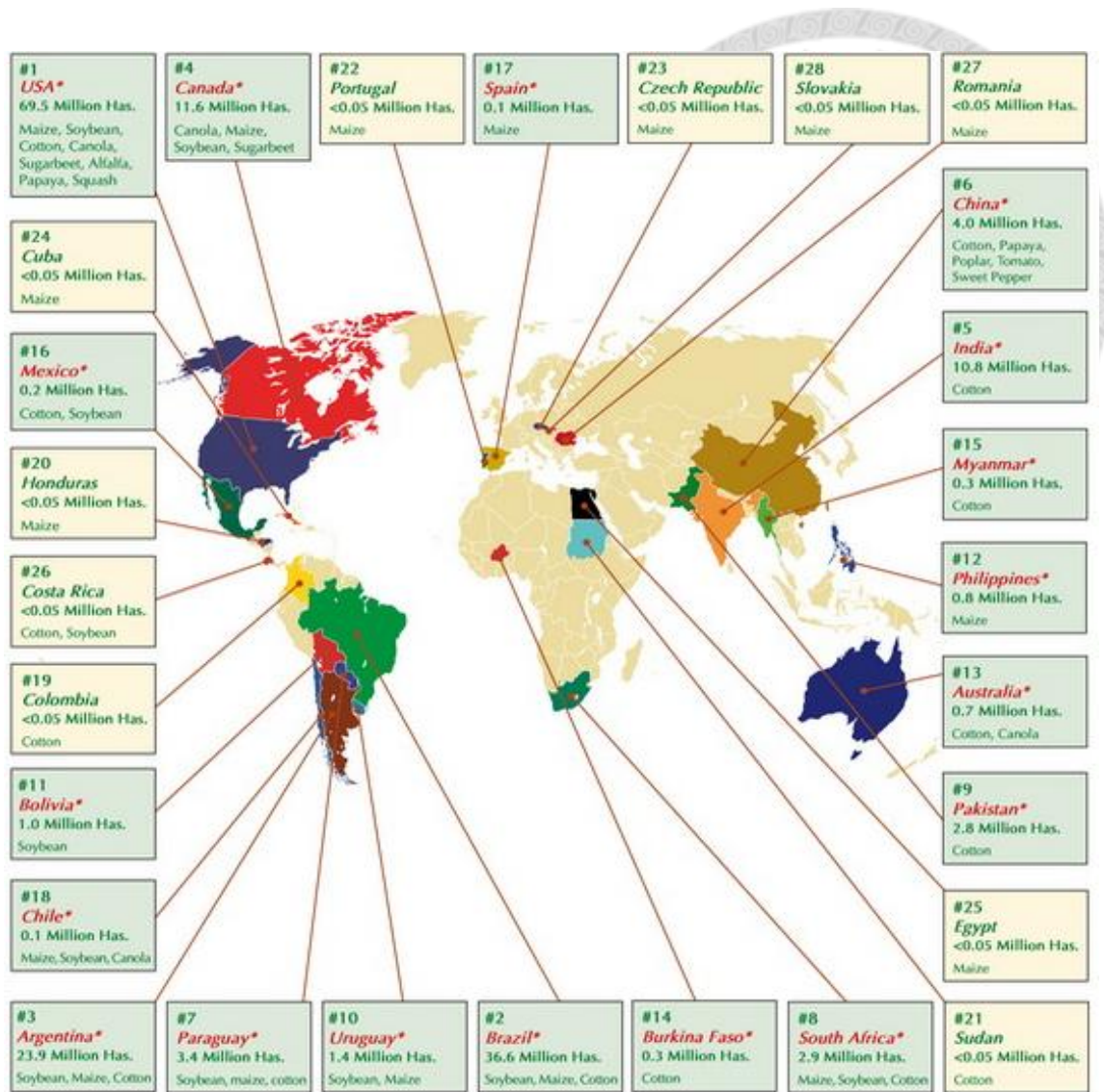
Figure 1.2 Global map of Biotech Crop countries and Mega-countries in 2012
Source: James (2012).
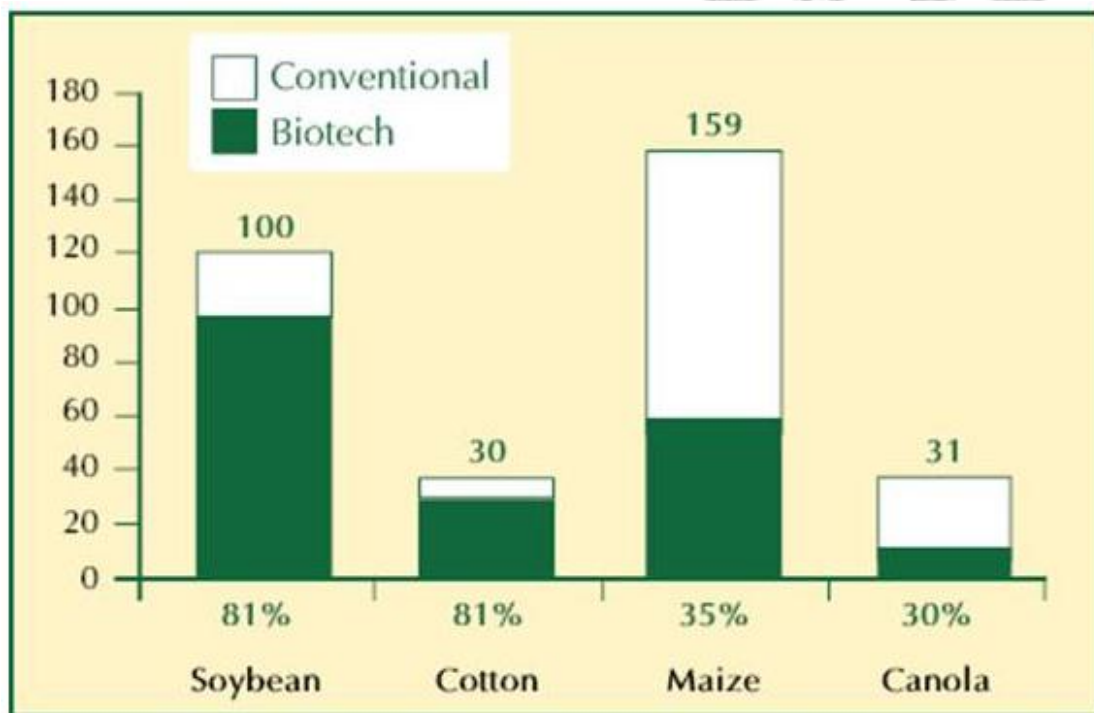
Figure 1.3 Biotech Crop Areas as % of Global Area of Principal Crops, 2012 (Million Hectares)

Source: James (2012)

Chapter 2 Safety Assessment of GM Crops

2.1 Design of Field Experiments

The EFSA proposed that field trials must be set in a number of sites, and each of varieties should be randomized in plot of multiple blocks. The interaction of environment and genotype must be considered. If the interaction exists, the efficiency of statistic test holds under enough replications. As a practical consideration, commercial variables as the basis of the minimum recommended number of sites on the test set. The EFSA suggests each field experiments must be repeated in at least eight sites, where can be representative of the crop growth under environmental conditions. In addition, the environmental variability is not only variation from region to region, as well as the variability between years. The main concern is not environmental variability within each region, but that under different environmental conditions, whether the potential variability of the test material changes. The EFSA concludes if the test sites only cover limited geographical environment, it is necessary to repeat the test more than a year.

It's very important to choose commercial varieties while setting equivalence limits. Varieties should represent the growth regions which in terms represent environments. The EFSA suggests that the commercial varieties should be randomized

in blocks besides GM crops and its conventional counterparts. In addition, the GMO Panel proposed at least six commercial varieties is necessary within a field trial to provide sufficient natural variation.

The EFSA recommended that the GM crops, conventional counterparts, and commercial varieties should be randomized within each of block in the field trial. On the other hand, the EFSA also suggested that (1) each of the appropriate conventional counterpart(s) must always together with its particular GM crop in the same block, (2) all the different GM crops and their conventional counterpart(s) and all the commercial varieties used for test equivalence with those GM crops must be fully randomized within each block. For example, in the same field trial plot, GM1, GM2, and GM3 represent three different GM crops; NIC1, NIC2 NIC3 are relatively non-GM crops; CV1, CV2, CV3, CV4 are four commercial varieties. Based on the trial of a minimum replicate number of four, an example of experiment design is as Table 2.1.

It is worth noting that the terms of the idea of degrees of freedom in the statistical point of view, all the information in the field trials should be placed in a test assessments, which can achieve the most effectively reduce the baseline residual. However, the GMO Panel proposed GM crops should be individually assessed due to the need to maintain transparency and verifiability. Therefore, for example, for the

design given in Table 2.1 to assess GM1, only plot 2,3,6,7,8,10 test data points in a the

area set can be placed into the assessment consideration; To assess GM2, only plot

1,2,3,7,9,10 data points can be placed in the assessment consideration.
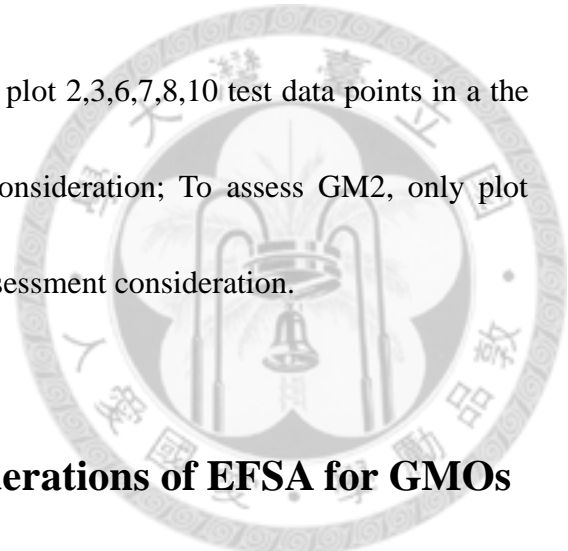
2.2 Concepts of Statistical Considerations of EFSA for GMOs
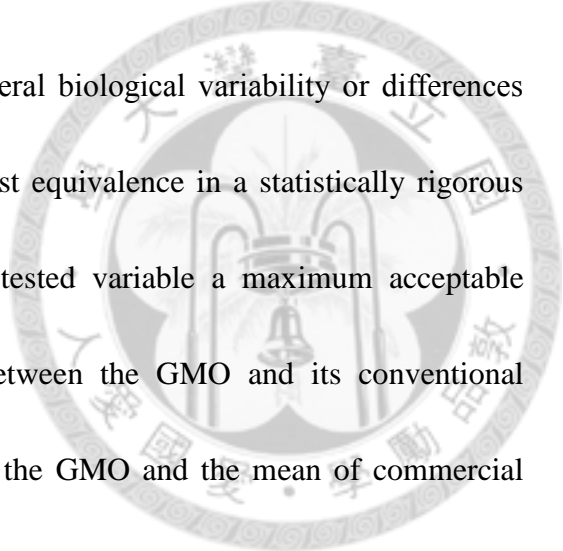
"*Statistical Consideration for the Safety Evaluation of GMOs, EFSA, 2010*"

updated statistical guidelines and possible approaches for the analysis of compositional,

agronomic and phenotypic data from field trials carried out for the risk assessment of

GM plants and derived foods/feeds, and provided minimum requirements that should be

met in the experimental design of field trials such as the inclusion of commercial

varieties, in order to ensure sufficient statistical power and reliable estimation of natural

variability. It suggests the safety assessment of GM products with the use of statistical

methods for comparisons between GM products and non-GM products, and between

GM products and commercial products which have been listed with security history. It

is recommended to quantify natural variability from data on non-GM commercial

varieties treated in the same way and in the same experiments as the GM and the

conventional counterpart test materials. The main concept is the comparative

assessment to demonstrate whether the GM products and/or derived food/feed are

different from their appropriate conventional counterparts and/or equivalent to commercial varieties, apart from the inserted traits. Comparisons should consider genotype, phenotype, composition, gene location characteristics, and other unique agronomic traits.

Instead of completely accepting or completely rejecting the null hypothesis that there is difference between GM crop and its counterpart variety from test results, the EFSA provides a richer frame work within which the conclusions of both types of assessment are allowed because statistical methodology should not be focused exclusively on either differences or equivalences. Statistically significant differences may point at biological changes caused by the genetic modification, but may not be relevant from the viewpoint of food safety. On the other hand, equivalence assessments may identify differences that are potentially larger than normal natural variation, but such cases may or may not be cases where there is an indication for true biological change caused by the genetic modification. A procedure combining both approached can only aid the subsequent toxicological assessment following risk characterization of the statistical results. The GMO Panel proposed the use of confidence intervals can be described more messages between GM crops and its control counterpart varieties.

The concept of equivalence must be clarified clearly, which is defined as

"absence of differences ", rather than the general biological variability or differences caused by genetically modified. In order to test equivalence in a statistically rigorous manner, it is necessary to specify for each tested variable a maximum acceptable difference, set either as the difference θ between the GMO and its conventional counterpart, or as the difference θ' between the GMO and the mean of commercial reference varieties. In principle the limits on the difference can be different in the positive and the negative direction; these are termed, respectively, the 'upper equivalence limit', θ_U, and the 'lower equivalence limit', θ_L。

2.3 Two One-sided Tests (TOST) Procedure

When testing for differences (proof of difference approach) the null hypothesis and alternative hypothesis are:

$$H_0 : \mu_G - \mu_C = 0 \qquad\qquad (2.1)$$

$$H_a : \mu_G - \mu_C \neq 0 ,$$

where $\mu_G(\mu_C)$ is the population mean of the GM (its conventional counterpart) crop. In words, the null hypothesis is "no difference between the GMO and its conventional counterpart" against the alternative hypothesis: "difference between the GMO and its conventional counterpart". This is a common two-tailed test for detecting on the both

direction of incremental and decreasing value. However, the difference of endpoint only to fall in ascending or descending one end, one-tailed test should be used at this time. For each test with significance level $1-\alpha$ (e.g. 95 %), there is a limited Type I error probability (α, the size of the test) that a significant result is obtained (i.e. a difference is found) whereas no difference exists in reality. However, these tests do not restrict the Type II error probability (β) of finding no significance whereas in reality there is a difference. So the absence of significant results is not a proof for equivalence of the GMO and the conventional counterpart, or ''absence of evidence is not evidence of absence'' (Altman and Bland, 1995, 2004).

Schuirmann (1987) proposed a new approach to this problem. The objectives of assessment of equivalence may be formulated into the following statistical hypotheses:

$$H_0 : \mu_G - \mu_C \leq \theta_L \quad \text{or} \quad \mu_G - \mu_C \geq \theta_U \tag{2.2}$$

$$Ha : \theta_L < \mu_G - \mu_C < \theta_U,$$

where θ_L is the lower bound for equivalence; θ_U is the upper bound for equivalence. It is the interval hypothesis, which can be decomposed into two sets of one-sided hypotheses:

$$H_{01} : \mu_G - \mu_C \leq \theta_L$$

$$H_{a1} : \mu_G - \mu_C > \theta_L$$

14

and (2.3)
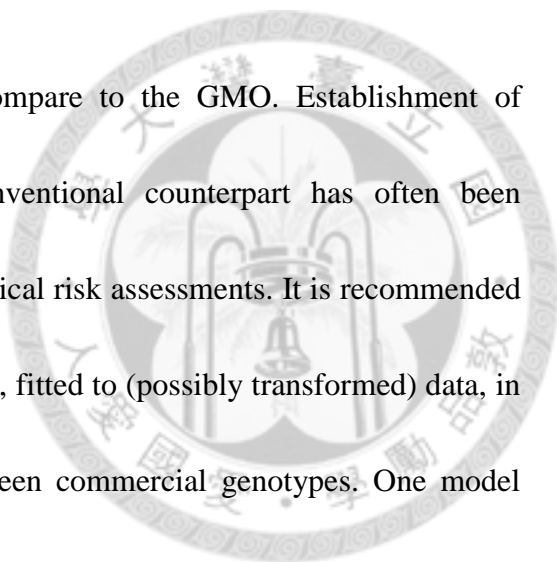
$$H_{02} : \mu_G - \mu_C \geq \theta_U$$

$$H_{a2} : \mu_G - \mu_C < \theta_U$$

The two one-sided tests procedure rejects the interval hypothesis H_0, and concludes equivalence between μ_G and μ_C, if and only if both H_{01} and H_{02} are rejected at a chosen nominal level of significance α. Equivalently, FDA (2001) and EMEA (2001) propose that the $(1-2\alpha)\%$ confidence interval for $\mu_G - \mu_C$ could be constructed for the bioequivalence test.

2.4 The EFSA's Approach
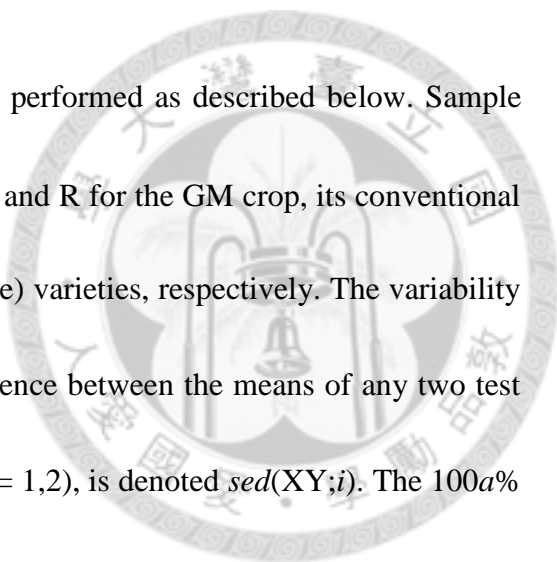
The approach which the EFSA recommended for establishment of equivalence limits is reviewed below (EFSA, 2010). In addition to the GMO and its conventional counterpart, the trials performed include several commercial crop varieties, which must represent non-GM varieties with a proven history of safe use, and these should be fully randomized as integral parts of the experiment. When commercial varieties are included in the same experiment where the GMO is tested against the conventional counterpart(s) then data on commercial varieties are obtained in identical conditions to that of the GM and its conventional counterpart. It is sensible to derive equivalence limits by
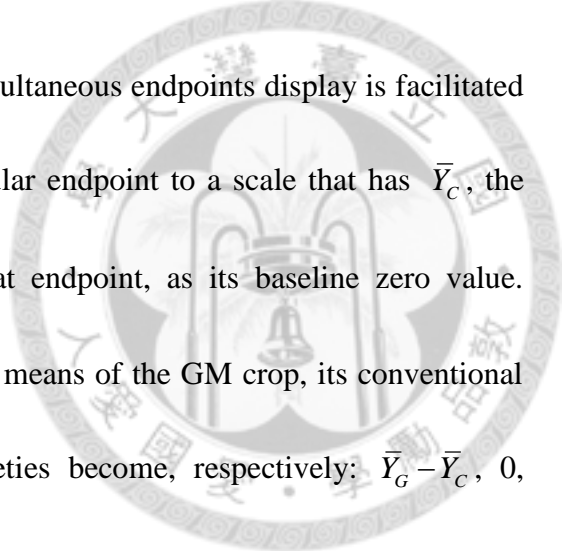
considering how the commercial varieties compare to the GMO. Establishment of equivalence between the GMO and its conventional counterpart has often been interpreted as relevant for subsequent toxicological risk assessments. It is recommended to apply a linear mixed-effects statistical model, fitted to (possibly transformed) data, in order to derive an estimate of variation between commercial genotypes. One model should be used for calculation of the confidence limits for both tests (difference and equivalence); a slightly different model should be used to estimate the equivalence limits to be used in the equivalence test.

Denote by *I* an indicator variable (uncentered in the mixed model) such that *I*=1 for a field plot having any of the commercial varieties, and *I*=0 otherwise. Then the random factors for model 1 should include, but not necessarily be restricted to, those representing the variation: (i) between the test materials (a set that includes the GM crop, its conventional counterpart, each of the commercial varieties and any additional comparators); (ii) in the interaction between the test materials and *I;* (iii) between sites; and (iv) between blocks within sites. Model 2 should be identical to model 1 except that the random factor representing the interaction between the test materials and *I* is omitted.

For each endpoint, calculation of the confidence limits, estimation of equivalence

limits and associated statistical tests should be performed as described below. Sample means are denoted by $\bar{Y}$, with subscripts G, C and R for the GM crop, its conventional counterpart and the set of commercial (reference) varieties, respectively. The variability encompassed in the standard error of the difference between the means of any two test materials, X and Y, calculated using model i ($i = 1,2$), is denoted $sed(XY;i)$. The 100a% point of Student's t distribution is denoted as $t(df;i;a)$, where I denotes the model used and df is the appropriate number of degrees of freedom which is recommended to be calculated by the Kenward-Roger method (Kenward and Roger, 1997). The least significant difference between the means of any two test materials, X and Y, using model i, should be calculated as the product of $t(df;i;a)$ and $sed(XY;i)$, and is denoted $lsd(XY;i;a)$. For the difference test, the two-sided 90% confidence limits should be calculated about $\bar{Y}_G$, as $\bar{Y}_G \pm lsd(GC;1;95)$; the null hypothesis of equality between $\bar{Y}_G$ and $\bar{Y}_C$ should be rejected and the test deemed statistically significant if $\bar{Y}_C$ falls outside these limits. For the equivalence test, the two-sided 95% equivalence limits should be estimated as $\bar{Y}_R \pm lsd(GR;2;97.5)$ and two-sided 90% confidence limits should be calculated about $\bar{Y}_G$ as $\bar{Y}_G \pm lsd(GR;1;95)$; the null hypothesis of non-equivalence should be rejected and the test deemed statistically significant if and only if the confidence limits lie entirely inside the equivalence limits.

After the appropriate transformation, simultaneous endpoints display is facilitated by shifting all relevant values for each particular endpoint to a scale that has $\bar{Y}_C$, the mean of the conventional counterpart for that endpoint, as its baseline zero value. Therefore, on this new scale, the values of the means of the GM crop, its conventional counterpart and the set of commercial varieties become, respectively: $\bar{Y}_G - \bar{Y}_C$, 0, $\bar{Y}_R - \bar{Y}_C$. And the confidence limits for the difference test on this new scale become: $\bar{Y}_G - \bar{Y}_C \pm lsd(GC;1;95)$, the equivalence limits $\bar{Y}_R - \bar{Y}_C \pm lsd(GR;2;97.5)$, and the confidence limits for the equivalence test $\bar{Y}_G - \bar{Y}_C \pm lsd(GR;1;95)$.

To facilitate visual interpretation, instead of using the two sets of confidence limits in the graphs, it is recommended for convenience that only one be displayed, that for the difference test. Without some adjustment, the confidence limits for the difference test would not give a valid visual representation for the equivalence test on the graph. This problem is overcome by making an adjustment to the displayed equivalence limits. After this adjustment the displayed confidence limits for the difference test may be used as a basis also for the visual representation of the equivalence test. In this way, one confidence limit may serve visually for assessing the outcome of both tests simultaneously. The adjustment of the equivalence limits consists of two steps: (1) scaling the basic equivalence limits, so that the confidence limits required for the

difference and equivalence tests have the same width; and (2) an appropriate shift to facilitate display of the adjusted limits, together with $\bar{Y}_G$, on the scale that has $\bar{Y}_C$ as its baseline zero value. The adjusted equivalence limits for visual display should be calculated by the formula:

$$(\bar{Y}_G - \bar{Y}_C) + \{[\bar{Y}_R - \bar{Y}_G) \pm lsd(GR;2;97.5)]lsd(GC;1;95) / lsd(GR;1;95)\} \qquad (2.4)$$

Figure 2.1 lists the possible outcomes if both approaches are considered simultaneously (EFSA, 2010). Stringent use of the concept of equivalence would require the necessity of proving equivalence for all endpoints simultaneously (global equivalence). Figure 2.1 Simplified version of a graph for comparative assessment showing the 7 outcome types possible for a single endpoint. After adjustment of the equivalence limits, a single confidence limit (for the difference) serves visually for assessing the outcome of both tests (difference and equivalence). Here, only the upper adjusted equivalence limit is considered. Shown are: the mean of the GM crop on an appropriate scale (square), the confidence limits (whiskers) for the difference between the GM crop and its conventional counterpart (bar shows confidence interval), a vertical line indicating zero difference (for proof of difference), and vertical lines indicating adjusted equivalence limits (for proof of equivalence). For outcome types 1, 3 and 5 the

null hypothesis of no difference cannot be rejected: for outcomes 2, 4, 6 and 7 the GM crop is different from its conventional counterpart. Regarding interpretation of equivalence, four categories are identified: in category (i) the null hypothesis of non-equivalence is rejected in favor of equivalence; in categories (ii), (iii) and (iv) non-equivalence cannot be rejected. More detailed explanations about (ii), (iii) and (iv) can be found in EFSA (2010).
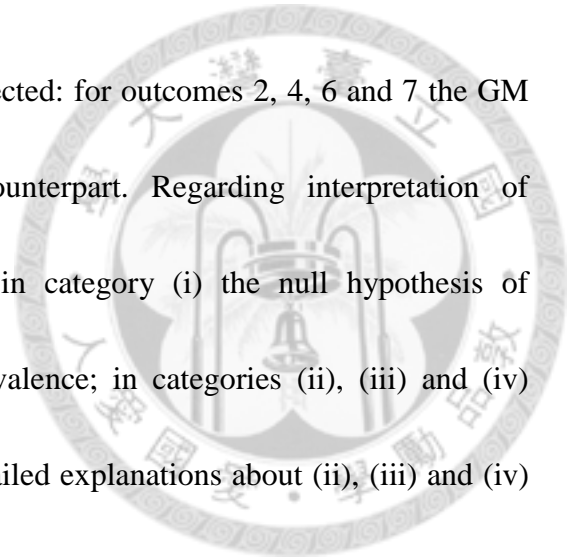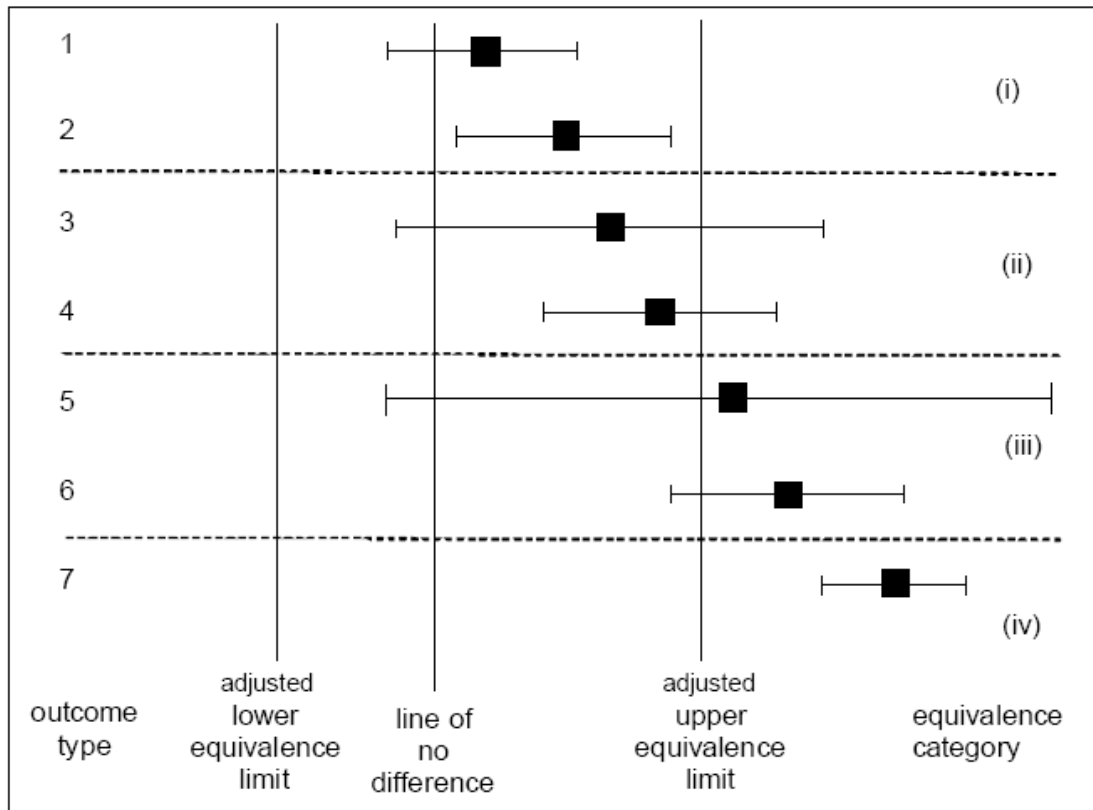
Figure 2.1 Seven possible outcomes for a single endpoint of difference testing and equivalence testing.

Source: EFSA (2010)

Table 2.1 Design of experiment of GM crops, non-GM crops, and commercial varieties

Block	Plot									
	1	2	3	4	5	6	7	8	9	10
1	GM2	CV2	CV1	GM3	NIC3	NIC1	CV3	GM1	NIC2	CV4
2	CV2	GM2	CV3	NIC3	NIC2	GM1	NIC1	CV4	CV1	GM3
3	NIC1	NIC3	GM1	CV1	GM3	NIC2	CV2	CV4	CV3	GM2
4	GM3	GM2	CV1	NIC1	CV2	NIC2	NIC3	CV3	CV4	GM1

Source: EFSA (2010)

Chapter 3 Proposed Methods
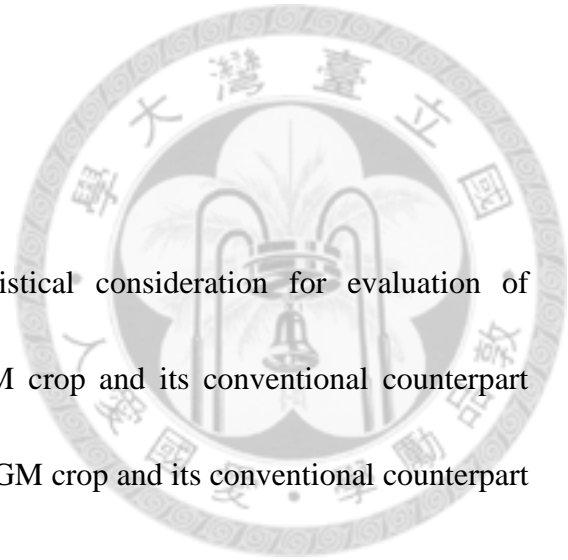
The EFSA scientific opinion on statistical consideration for evaluation of equivalence in safety profile between the GM crop and its conventional counterpart suggests that the mean difference between the GM crop and its conventional counterpart lie within the natural variability of the commercial varieties.

In this chapter, based on the concept of scaled average bioequivalence (Patterson and Jones, 2012), we propose a new criterion for assessment of equivalence between the GM crop and its conventional counterpart which is the scaled square mean difference between the GM crop and its conventional crop with the variance of commercial crops as the scale factor. Then the modified large sample (MLS) method (Lee, et al., 2004) is applied to obtain the $((1-\alpha)\%$ upper confidence limit for the linearized criterion which is used as the testing statistics for the hypothesis based on the proposed criterion.

3.1 The Linear Statistical Model for Mixed-Effects Models

We use the mixed-effects model proposed by EFSA for the data. Let Y_{ijkl} denote the response, possibly logarithmically transformed, of replicate j within site i for treatment k and genotype l ; $j = 1, 2..., p$, $i = 1, 2, ..., n$, $k = 1, 2, 3$ (1:conventional

counterpart; 2: GM crops; 3: commercial variety), $l = 1, 2, ..., q$. The following nested

mixed-effects model is suggested by the EFSA (2010).

$$y_{ijkl} = \mu + e_i + r_{(i)j} + t_k + g_l + \varepsilon_{ijkl}$$

where

μ: mean of observations, e: environment (site), $e_i \sim N(0, \sigma_e^2)$

r: replication per site, $r_{(i)j} \sim N(0, \sigma_r^2)$, t: treatment, $\sum_{k=1}^{3} t_k = 0$

k=1: conventional variety, k=2: GMO, k=3: commercial varieties

g: genotype of commercial varieties, $g_l \sim N(0, \sigma_g^2)$, ε: residuals, $\varepsilon_{ijkl} \sim N(0, \sigma_\varepsilon^2)$

Define

$$\mu = \bar{y}_{....}; \quad \hat{e}_i = \bar{y}_{i...} - \bar{y}_{....}; \quad \hat{r}_{(i)j} = \bar{y}_{ij..} - \bar{y}_{i...}; \quad \hat{t}_k = \bar{y}_{..k.} - \bar{y}_{....}; \quad \hat{g}_l = \bar{y}_{..3l} - \bar{y}_{..3.}; \quad \hat{\varepsilon}_{ijkl} = \bar{y}_{ijkl} - \bar{y}_{i \cdot kl}$$

Where

$$\bar{y}_{ij..} = \frac{\sum_k^3 \sum_l^q y_{ijkl}}{s} \quad \text{, mean of the observations in } i \text{ th site and } j \text{ th replicate,}$$

$$\bar{y}_{i...} = \frac{\sum_j^p \sum_k^3 \sum_l^q y_{ijkl}}{ps}, \text{ mean of the observations in } i \text{ th site,}$$

$$\bar{y}_{..3l} = \frac{\sum_i^n \sum_j^p y_{ijkl}}{npq}, \text{ mean of the observations of commercial variety with } l \text{ th genotype,}$$

$$\bar{y}_{..k.} = \frac{\sum_i^n \sum_j^p \sum_l^q y_{ijkl}}{np} \quad \text{mean of the observations of } k \text{ th treatment, and}$$

$$\bar{y}_{....} = \frac{\sum_i^n \sum_j^p \sum_k^3 \sum_l^q y_{ijkl}}{nps}, \text{ overall mean, and}$$

$s=q+(k-1)$.
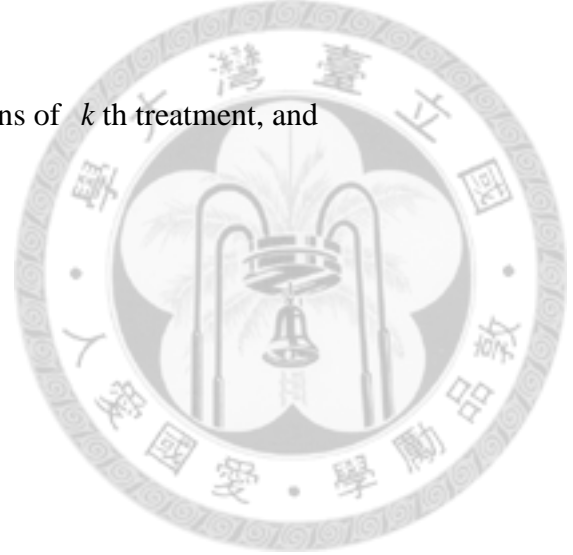
The sums of squares, degree of freedom, and the expected value of its corresponding mean square for environment, replication, treatment, genotype, and errors are given respectively as follows:

$$SSe = ps \sum_i^n (\bar{y}_{i...} - \bar{y}_{....})^2 \text{, Degree of freedom : } n-1, \quad E(MSe) = ps\sigma_e^2 + s\sigma_r^2 + \sigma_\varepsilon^2;$$

$$SSr = s \sum_i^n \sum_j^p (\bar{y}_{ij..} - \bar{y}_{i...})^2 \quad \text{, Degree of freedom : } n(p-1), \quad E(MSr) = s\sigma_r^2 + \sigma_\varepsilon^2;$$

$$SSt = np \sum_k^3 (\bar{y}_{..k.} - \bar{y}_{....})^2 \quad \text{, Degree of freedom : } k-1, \quad E(MSt) = npQ(t) + np\sigma_g^2 + \sigma_\varepsilon^2;$$

$$SSg = np3 \sum_l^q (\bar{y}_{..3l} - \bar{y}_{..3.})^2 \text{, Degree of freedom : } q-1, \quad E(MSg) = np\sigma_g^2 + \sigma_\varepsilon^2;$$

$$SSE = \sum_k^3 \sum_i^n \sum_j^p \sum_l^q (\bar{y}_{ijkl} - \bar{y}_{i\cdot kl})^2 \quad \text{, Degree of freedom: } nps-np-k-q+2,$$

$$E(MSE) = \sigma_\varepsilon^2$$

The corresponding ANOVA table for the nested mixed-effects model is provided in Table 3.1.

3.2 The Criteria and Hypotheses

Since the EFSA scientific opinion for safety assessment for GM crop requires the average difference between the GM crop and its conventional counterpart should not exceed the natural variation of commercial variation, based on the concept of scaled average bioequivalence (Patterson and Jones, 2012), we propose the following scaled average criterion:

$$\theta = \frac{(\mu_G - \mu_C)^2}{\sigma_g^2} \tag{3.1}$$

Where μ_G (μ_C) is the average of the GM (conventional) crop and σ_g^2 is the variance of the commercial varieties. It follows that the hypothesis for evaluation of equivalence between the GM crop and its conventional counterpart is given as

$$H_0 : \theta \geq \theta_0 \quad \text{vs.} \quad H_a : \theta < \theta_0, \tag{3.2}$$

where θ_0 is the equivalence limit which that the maximally allowable natural variability of the commercial varieties to claim equivalence between the GM crop and its conventional counterpart.

Following the linearized criterion for individual bioequivalence (IBE) proposal by Hyslop, et al. (2000), we also can formulate θ into the following linearized criterion

$$\eta = (\mu_G - \mu_C)^2 - \theta_0 \sigma_g^2 \tag{3.3}$$

As a result, hypothesis in Equation (3.2) can be re-formulated as

$$H_0 : \eta \geq 0 \quad \text{vs.} \quad H_a : \eta < 0 \tag{3.4}$$

Hyslop et al. (2000) also suggest using the $(1-\alpha)\%$ upper confidence bound derived

by the MLS method as a test statistic for hypothesis (3.4)
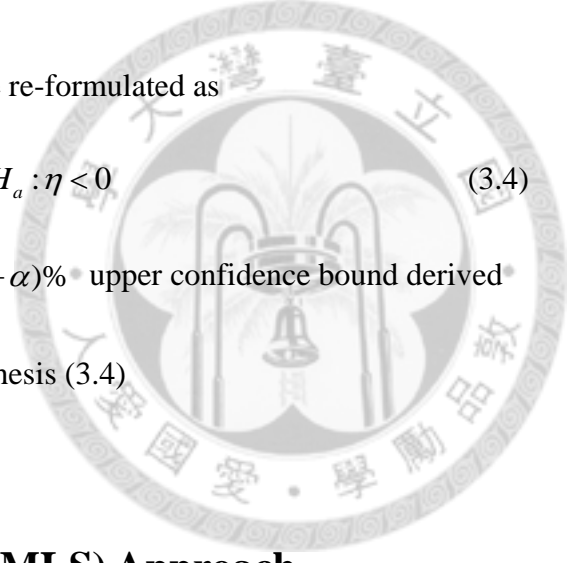
3.3 The Modified Large-Sample (MLS) Approach

The MLS method from Lee et al. (2004) and Hyslop, et al. (2000) is briefly

reviewed as follows:

Consider the problem of setting a $(1-\alpha)100\%$ confidence bound for

$\eta = c_1 \sigma_1^2 + \cdots + c_p \sigma_p^2$, where $\sigma_1^2, \cdots, \sigma_p^2$ are unknown variance components and

$c_1, \cdots, c_p$ are known nonzero constants. Let $\hat{\sigma}_i^2$ be an unbiased estimator of σ_i^2 such

that $L(\hat{\sigma}_i^2) = L(\sigma_i^2 n_i^{-1} X_i)$, where $L(X_i) = \chi_{n_i}^2$, $i = 1, \cdots, p$. $\hat{\sigma}_i^2$ is a quadratic forms of

an observed normal random vector. The extension builds on the fact that the quadratic

form $c_1 \hat{\sigma}_1^2 + \cdots + c_p \hat{\sigma}_p^2$ may have a chi-squared representation in the sense that

$$L(c_1 \hat{\sigma}_1^2 + \cdots + c_p \hat{\sigma}_p^2) = L(\lambda_1 v_1^{-1} X_1 + \cdots + \lambda_p v_p^{-1} X_p)$$

where $X_1, \cdots, X_P$ are independent chi-squared random variables with v_i degrees of

freedom and $\lambda_i's$ are unknown parameters. In addition, the number of independent

chi-squared random variables is the same as the number of $\hat{\sigma}_i^2$. The results can be

extended to the situation where $c_1\hat{\sigma}_1^2 + \cdots + c_p\hat{\sigma}_p^2$ can be represented in distribution by a linear combination of $q(\neq p)$ independent chi-squared random variables. Applying the independent linear combination $\lambda_1 n_1^{-1}X_1 + \cdots + \lambda_p n_p^{-1}X_p$, the MLS upper and lower confidence bounds could be obtained as below

$$c_1\hat{\sigma}_1^2 + \cdots + c_p\hat{\sigma}_p^2 + \sqrt{\hat{\lambda}_1^2(\frac{v_1}{u_1}-1)^2 + \cdots + \hat{\lambda}_p^2(\frac{v_p}{u_p}-1)^2}$$

, and

$$c_1\hat{\sigma}_1^2 + \cdots + c_p\hat{\sigma}_p^2 - \sqrt{\hat{\lambda}_1^2(\frac{v_1}{l_1}-1)^2 + \cdots + \hat{\lambda}_p^2(\frac{v_p}{l_p}-1)^2}$$

, where $\hat{\lambda}_i$ is a consistent estimator of λ_i, $i=1,\cdots,p$,

$$u_i = \begin{cases} \chi^2_{\alpha,v_i} & \text{, if } \lambda_i > 0 \\ \chi^2_{1-\alpha,v_i} & \text{, if } \lambda_i < 0 \end{cases}$$

, and

$$l_i = \begin{cases} \chi^2_{\alpha,v_i} & \text{, if } \lambda_i < 0 \\ \chi^2_{1-\alpha,v_i} & \text{, if } \lambda_i > 0 \end{cases}$$

χ^2_{α,v_i} is the $(1-\alpha)\%$ upper percentile of chi-square random variable with v_i degrees of freedom.

3.4 The MLS Upper Confidence Bound

Let $\bar{Y}_G$ and $\bar{Y}_C$ denote the average of the GM crop and its conventional counterpart, respectively.

Since

$$E(Y_{ijkl}) = E(\mu + e_i + r_{(i)j} + t_k + g_l + \varepsilon_{ijkl}) = \mu + t_k$$

$$\begin{aligned} Var(Y_{ijkl}) &= Var(\mu + e_i + r_{(i)j} + t_k + g_l + \varepsilon_{ijkl}) \\ &= Var(e_i) + Var(r_{(i)j}) + Var(t_k) + Var(g_l) + Var(\varepsilon_{ijkl}) \\ &= \sigma_e^2 + \sigma_r^2 + \sigma_g^2 + \sigma_\varepsilon^2 \end{aligned}$$

Define

$$\bar{Y}_G = \frac{1}{np} \sum_i^n \sum_j^p y_{ij2l} \text{ , and } \bar{Y}_C = \frac{1}{np} \sum_i^n \sum_j^p y_{ij1l} \text{ .}$$

It follows that

$$\begin{aligned} Var(\bar{Y}_G) &= Var(\frac{\sum_i^n \sum_j^p y_{ij2l}}{np}) = \frac{np}{n^2 p^2} Var(y_{ij2l}) \\ &= \frac{1}{np} Var(e_i + r_{(i)j} + \varepsilon_{ij2l}) \\ &= \frac{1}{np} (\sigma_e^2 + \sigma_r^2 + \sigma_\varepsilon^2), \end{aligned}$$

$$\begin{aligned} Var(\bar{Y}_C) &= Var(\frac{\sum_i^n \sum_j^p y_{ij1l}}{np}) = \frac{np}{n^2 p^2} Var(y_{ij1l}) \\ &= \frac{1}{np} Var(e_i + r_{(i)j} + \varepsilon_{ij1l}) \\ &= \frac{1}{np} (\sigma_e^2 + \sigma_r^2 + \sigma_\varepsilon^2), \end{aligned}$$

$$Var(\bar{Y}_G - \bar{Y}_C) = Var(\bar{Y}_G) + Var(\bar{Y}_C) = \frac{2}{np} (\sigma_e^2 + \sigma_r^2 + \sigma_\varepsilon^2), \text{ and}$$

$$\begin{aligned} E[Var(\bar{Y}_G - \bar{Y}_C)] &= E[\frac{2}{np^2 s}(MSe + (p-1)MSr + p(s-1)MSE)] \\ &= \frac{2}{np}(\sigma_e^2 + \sigma_r^2 + \sigma_\varepsilon^2). \end{aligned}$$

Then $\bar{Y}_G - \bar{Y}_C$ follows a normal distribution with mean $t_G - t_C$ and variance $\frac{2}{np}\sigma_t^2$,

where $t_G - t_C$ is denoted as $\mu_G - \mu_C$ and $\sigma_t^2 = \sigma_e^2 + \sigma_r^2 + \sigma_\varepsilon^2$

From Table 3.1, the ANOVA unbiased estimator of σ_e^2, σ_r^2, and σ_ε^2 are given as

$$\hat{\sigma}_e^2 = \frac{MSe - MSr}{ps}, \quad \hat{\sigma}_r^2 = \frac{MSr - MSE}{s}, \text{ and } \hat{\sigma}_\varepsilon^2 = MSE.$$

It follows that an unbiased estimator of σ_t^2 is given as

$$
\begin{aligned}
\hat{\sigma}_t^2 &= \hat{\sigma}_e^2 + \hat{\sigma}_r^2 + \hat{\sigma}_\varepsilon^2 \\
&= \frac{MSe - MSr}{ps} + \frac{MSr - MSE}{s} + MSE \\
&= \frac{1}{ps}[MSe + (p-1)MSr + p(s-1)MSE].
\end{aligned}
\tag{3.5}
$$

In addition, MSe, MSr, and MSE are mutually independent with the following

distributions

$$MSe \sim \frac{\sigma_\varepsilon^2 + ps\sigma_e^2 + s\sigma_r^2}{n-1}\chi_{(n-1)}^2,$$

$$MSr \sim \frac{\sigma_\varepsilon^2 + s\sigma_r^2}{p(n-1)}\chi_{(n(p-1))}^2, \text{ and}$$

$$MSE \sim \frac{\sigma_\varepsilon^2}{df_\varepsilon}\chi_{(df_\varepsilon)}^2.$$

On the other hand, an ANOVA unbiased estimator of σ_g^2 is given as

$$\hat{\sigma}_g^2 = \frac{MSg - MSE}{np}
\tag{3.6}$$

where

$$MSg \sim \frac{\sigma_\varepsilon^2 + np\sigma_g^2}{q-1}\chi_{(q-1)}^2.$$

The deviation of $100(1-\alpha)\%$ MLS upper confidence bound for $\eta = (\mu_G - \mu_C)^2 - \theta\sigma_g^2$

by the MLS method is given as follows:

Let $A = (\bar{Y}_G - \bar{Y}_C)^2$, and B is the unbiased estimator of $-\theta_0\sigma_g^2 = \frac{(-\theta_0)}{np}[MSg - MSE]$

Denote the A_U as the $100(1-\alpha)\%$ upper bound of $\bar{Y}_G - \bar{Y}_C$, then

$$A_U^2 = [|\bar{Y}_G - \bar{Y}_C| + t_{1-\alpha/2,2(np-1)}\sqrt{Var(\bar{Y}_G - \bar{Y}_C)}]^2.$$

Define C as

$$C = [A_U^2 - A]^2$$
$$= \{[|\bar{Y}_G - \bar{Y}_C| + t_{1-\alpha/2,2(np-1)}\sqrt{\frac{2}{np^2s}(MSe + (p-1)MSr + p(s-1)MSE)}]^2 - (\bar{Y}_G - \bar{Y}_C)^2\}^2.$$

Furthermore

$$L(-\theta_0\hat{\sigma}_g^2) = L[(\frac{-\theta_0}{np})MSg + (\frac{\theta_0}{np})MSE]$$
$$= L[\lambda_1\chi_{(q-1)}^2 + \lambda_2\chi_{df_\varepsilon}^2],$$

where $\lambda_1 = (-\theta_0)[\frac{\sigma_\varepsilon^2}{np} + \sigma_g^2]$, and $\lambda_2 = \frac{\theta_0}{np}\sigma_\varepsilon^2$.

It follows that

$$D = \hat{\lambda}_1[\frac{q-1}{\chi_{1-\alpha,q-1}^2} - 1]^2 + \hat{\lambda}_2[\frac{df_\varepsilon}{\chi_{\alpha,df_\varepsilon}^2} - 1]^2,$$

where $\hat{\lambda}_1 = (\dfrac{-\theta_0}{np})MSg$ and $\hat{\lambda}_2 = \dfrac{\theta_0}{np}MSE.$

Consequently, the $100(1-\alpha)\%$ MLS upper confidence for η is given as

$$\hat{\tau} = A + B + \sqrt{C + D}. \tag{3.7}$$

Then the null hypothesis of (3.4) is rejected at the α significance level and the safety profile of the GM crop can be claimed to be equivalent to that of its conventional counterpart if the $100(1-\alpha)\%$ MLS upper confidence bound $\hat{\tau} = A + B + \sqrt{C + D}$ is smaller than 0.

Table 3.1: ANOVA table for the nested mixed-effects model

Source	D.F	EMS
Environment	$n-1$	$\sigma_\varepsilon^2 + ps\sigma_e^2 + s\sigma_r^2$
Replication	$n(p-1)$	$\sigma_\varepsilon^2 + s\sigma_r^2$
Treatment	$k-1$	$\sigma_\varepsilon^2 + npQ(t) + np\sigma_g^2$
Genotype	$q-1$	$\sigma_\varepsilon^2 + np\sigma_g^2$
Error	df_ε	σ_ε^2
Total	$nps-1$	

Chapter 4 Numeric Examples
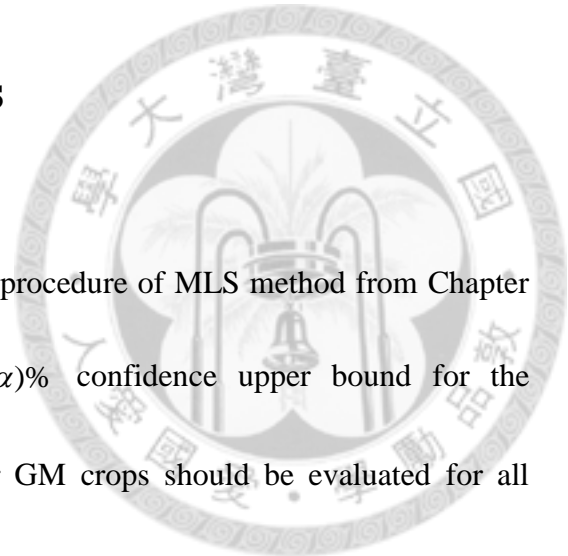
In this chapter, we will demonstrate the procedure of MLS method from Chapter 3 with examples to construct the $100(1-\alpha)\%$ confidence upper bound for the linearized criterion. Assessment of safety for GM crops should be evaluated for all phenotypes, chemical components, or agricultural traits. However, we would only focus on one agricultural trait or one chemical component in order to demonstrate the MLS method and to compare the results from the approaches recommended by the EFSA scientific opinion on statistical considerations in 2010.

4.1 Design of the Example

This example considers an field trial which is repeated at eight sites. For each site, there are one GM crop, one conventional crop and six commercial varieties. the number of replicates is 4 for each site. We consider the log-transformed responses. The equivalence limit θ_0 is set at 0.2 while the variance of the genotype from commercial varieties σ_g^2 is 2. In addition, the specifications of the parameter for the nested mixed-effects model in Section 3.1 are given below:

$\mu = 0$

$e_i \sim N(0,3)$, $i = 1,...,8$,

$r_j \sim N(0,1)$, $j = 1,...,4$,

$$t_k = \begin{cases} 0.05 & , k = 1 \text{ (conventional counterpart)} \\ 0.05 & , k = 2 \text{ (GM crops)} \\ -0.1 & , \text{others (commercial varieties),} \end{cases}$$

$g_l \sim N(0,2)$, $l = 1,...,6$, and

$\varepsilon_{ijkl} \sim N(0,0.6)$.

Hence, $\mu_G - \mu_C$ is $0.05 - 0.05 = 0$. A total of 256 responses were generated according to the rested mixed-effects model with the specifications of the distribution assumptions and parameters as described above. The raw logarithmic responses are provided in Table 4.1.

4.2 The 95% Upper Confidence Bound

The estimates of μ_G and μ_C are given as

$\bar{Y}_C = 0.4278$, $\bar{Y}_G = 0.5027$

The ANOVA table is given in Table 4.2. From Table 4.2, the mean squares for environment (site), replicate, variety, genotype, and error are given respectively as

$MSe = 55.7653$, $MSr = 7.9580$, $MSt = 6.4432$, $MSg = 131.7064$, $MSE = 0.5843$

It follows that

$$A = (\bar{Y}_G - \bar{Y}_C)^2 = (0.5027 - 0.4278)^2 = 0.00566$$

$$B = -\theta_0 \hat{\sigma}_g^2 = (-\theta_0)\frac{MSg - MSE}{np} = (-0.2)\frac{131.7064 - 0.5843}{(8)(4)} = -0.8200$$

Then

$$\hat{\eta} = (\hat{\mu}_G - \hat{\mu}_C)^2 - \hat{\theta}_0 \hat{\sigma}_g^2 = (\bar{Y}_G - \bar{Y}_C)^2 - \hat{\theta}_0 \hat{\sigma}_g^2$$
$$= A + B = 0.00566 + (-0.8200) = -0.8143$$

On the other hand,

$$\hat{A}_U^2 = [|\bar{Y}_G - \bar{Y}_C| + t_{1-\alpha, 2(np-1)}\sqrt{Var(\bar{Y}_G - \bar{Y}_C)}]^2$$
$$= [|\bar{Y}_G - \bar{Y}_C| + t_{1-\alpha, 2(np-1)}\sqrt{\frac{2}{np^2 s}(MSe + (p-1)MSr + p(s-1)MSE)}]^2$$
$$= [|0.5027 - 0.4278| + 1.67 \times \sqrt{\frac{2}{(8)(4^2)(8)}(55.7653 + (3)(7.9580) + 4(7)(0.5843))}]^2$$
$$= 0.6373,$$

and hence

$$C = (\hat{A}_U^2 - A)^2 = (0.6373 - 0.00566)^2 = 0.3990$$

In addition,

$$\hat{\lambda}_1' = (\frac{-\theta}{np})MSg = (\frac{-0.2}{8 \times 4})(131.7064) = -0.8232 ,$$

$$\hat{\lambda}_2' = (\frac{\theta}{np})MSE = (\frac{0.2}{8 \times 4})(0.5843) = 0.00365 ,$$

$q = 6 - 1$, and

$$df_\varepsilon = df_{Total} - df_e - df_r - df_t - df_g$$
$$= (nps - 1) - (n - 1) - n(p - 1) + (k - 1) + (q - 1)$$
$$= 255 - 7 - 24 - 2 - 5 = 217,$$

$\chi^2_{1-0.5,6-1} = 11.0705$, and $\chi^2_{0.05,217} = 183.907$.

It follows that

$$D = (-0.8232)^2 (\frac{6-1}{11.0705} - 1)^2 + (0.00365)^2 (\frac{217}{183.907} - 1)^2 = 0.2038$$
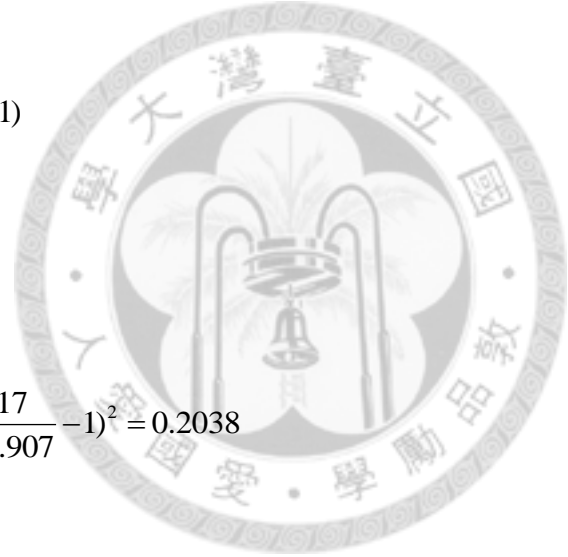
Then the 95% upper confidence bound for the linearized criterion is

$$U = A + B + \sqrt{C + D}$$
$$= 0.00566 + (-0.8200) + \sqrt{0.3990 + 0.2038}$$
$$= -0.0379.$$

Since the 95% upper confidence bound is smaller than 0, the null hypothesis in Equation (3.4) is rejected and the GM crop is claimed to be equivalent to its conventional counterpart at the 5% significance level.

4.3 The EFSA's Approach

We applied the EFSA approach to the same data set using the SAS codes given by the GMO Panel of the EFSA (EFSA, 2010). Recall that the adjusted equivalence limits suggested by the EFSA scientific opinion on statistical considerations is

$$(\bar{Y}_G - \bar{Y}_C) + \{[\bar{Y}_R - \bar{Y}_G) \pm lsd(GR; 2; 97.5)]lsd(GC; 1; 95) / lsd(GR; 1; 95)\}$$

From the ANOVA table in Table 4.1, it can be verified that

$lsd(GC;1;95) = 0.3157$,

$lsd(GR;1;95) = 1.6721$,

$lsd(GR;2;97.5) = 5.6329$.

$\bar{Y}_R - \bar{Y}_G = 0.4789$, and $\bar{Y}_G - \bar{Y}_C = 0.0752$.

The lower and upper adjusted equivalence limits are given respectively

$$0.0752 + \{[0.4789 - 5.6329] \times (\frac{0.3157}{1.6721})\} = -0.8979$$

and

$$0.0752 + \{[0.4789 + 5.6329] \times (\frac{0.3157}{1.6721})\} = 1.2291.$$

Their corresponding adjusted equivalence limits on the original scale are

$\exp(-0.8979) = 0.4074$, and $\exp(1.2291) = 3.4182$.

The 90% confidence interval for the average difference between GM product and conventional counterparts are respectively given as

$\exp(0.0752 - 0.3157) = 0.7862$, and $\exp(0.0752 + 0.3157) = 1.4783$.

Since the 90% confidence interval on the original scale for the average different

between the GM crop and its conventional counterpart $(0.7862, 1.4783)$ is completely

contained within the 95% adjusted equivalence limits $(0.4047, 3.4182)$, the GM crop

is claimed to be equivalent to its conventional counterpart at the 5% significance level.

In addition, the 90% confidence interval on the original scale contains 1. This is an

example of outcome (i) illustrated in Figure 2.1.
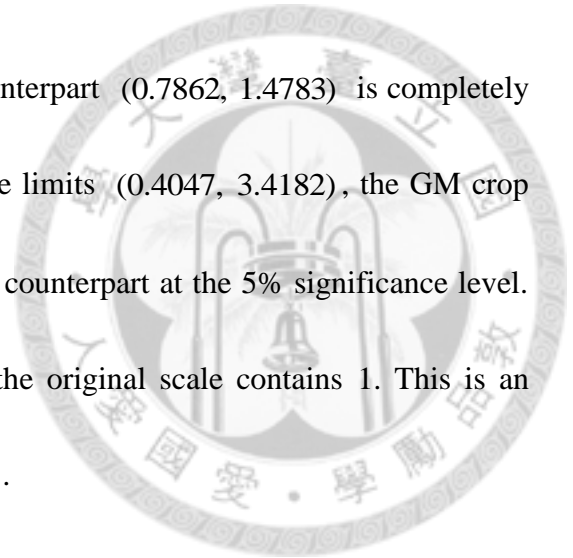
Table 4.1 Raw data of the numeric example

site	rep	k	genotype	y
1	1	1	0	-0.34348
1	1	2	0	-0.32463
1	1	3	1	-1.96276
1	1	4	2	-0.52594
1	1	5	3	-1.45369
1	1	6	4	0.136575
1	1	7	5	1.659405
1	1	8	6	3.659744
1	2	1	0	-2.39409
1	2	2	0	-0.77102
1	2	3	1	-1.32512
1	2	4	2	-2.32695
1	2	5	3	-2.04263
1	2	6	4	-2.57934
1	2	7	5	-2.11314
1	2	8	6	2.116118
1	3	1	0	-0.87054
1	3	2	0	-0.40614
1	3	3	1	-1.5087
1	3	4	2	-0.96802
1	3	5	3	-0.75545
1	3	6	4	-1.67232
1	3	7	5	-0.40345
1	3	8	6	3.004807
1	4	1	0	-2.68067
1	4	2	0	-4.15327
1	4	3	1	-3.88039
1	4	4	2	-3.2397
1	4	5	3	-4.83718
1	4	6	4	-2.47065
1	4	7	5	-2.02452
1	4	8	6	2.143112
2	1	1	0	0.651309

Table 4.1 Raw data of the numeric example (continued)

2	1	2	0	0.76004
2	1	3	1	1.311668
2	1	4	2	0.364564
2	1	5	3	0.066609
2	1	6	4	0.545993
2	1	7	5	2.910414
2	1	8	6	5.681826
2	2	1	0	1.030785
2	2	2	0	0.014076
2	2	3	1	0.577099
2	2	4	2	1.850748
2	2	5	3	1.153508
2	2	6	4	1.309675
2	2	7	5	3.138552
2	2	8	6	5.687911
2	3	1	0	0.933935
2	3	2	0	0.526651
2	3	3	1	-0.33949
2	3	4	2	2.031823
2	3	5	3	0.416347
2	3	6	4	0.395261
2	3	7	5	0.320967
2	3	8	6	6.074536
2	4	1	0	2.143204
2	4	2	0	2.700565
2	4	3	1	2.42217
2	4	4	2	3.520955
2	4	5	3	2.373655
2	4	6	4	3.363933
2	4	7	5	4.449046
2	4	8	6	5.862582
3	1	1	0	1.702914
3	1	2	0	4.238634
3	1	3	1	1.830989

Table 4.1 Raw data of the numeric example (continued)

3	1	4	2	3.35311
3	1	5	3	2.124325
3	1	6	4	2.561769
3	1	7	5	4.601721
3	1	8	6	8.298073
3	2	1	0	1.472821
3	2	2	0	1.114468
3	2	3	1	0.134322
3	2	4	2	1.198024
3	2	5	3	1.96735
3	2	6	4	0.538514
3	2	7	5	1.837776
3	2	8	6	6.26766
3	3	1	0	2.305447
3	3	2	0	2.345976
3	3	3	1	1.536964
3	3	4	2	2.135358
3	3	5	3	0.675815
3	3	6	4	1.454244
3	3	7	5	3.292453
3	3	8	6	7.001886
3	4	1	0	0.333695
3	4	2	0	0.995395
3	4	3	1	-1.16256
3	4	4	2	0.795557
3	4	5	3	1.654473
3	4	6	4	-0.08482
3	4	7	5	1.011555
3	4	8	6	4.764279
4	1	1	0	1.224277
4	1	2	0	1.412181
4	1	3	1	0.37732
4	1	4	2	1.359954
4	1	5	3	-0.10124

Table 4.1 Raw data of the numeric example (continued)

4	1	6	4	2.872095
4	1	7	5	2.297184
4	1	8	6	8.162462
4	2	1	0	-1.50315
4	2	2	0	0.204321
4	2	3	1	-1.78738
4	2	4	2	-0.81204
4	2	5	3	-1.66929
4	2	6	4	-1.17106
4	2	7	5	-0.60846
4	2	8	6	3.4998
4	3	1	0	0.102504
4	3	2	0	0.200354
4	3	3	1	-1.06567
4	3	4	2	2.251522
4	3	5	3	0.730513
4	3	6	4	0.829408
4	3	7	5	1.280226
4	3	8	6	4.621634
4	4	1	0	2.073987
4	4	2	0	-0.64327
4	4	3	1	-0.20676
4	4	4	2	0.988715
4	4	5	3	1.081729
4	4	6	4	0.651922
4	4	7	5	1.962178
4	4	8	6	5.685683
5	1	1	0	-2.61195
5	1	2	0	-1.22974
5	1	3	1	-1.67622
5	1	4	2	-1.65568
5	1	5	3	-1.4933
5	1	6	4	-1.3228
5	1	7	5	-1.31845

Table 4.1 Raw data of the numeric example (continued)

5	1	8	6	3.552251
5	2	1	0	2.419733
5	2	2	0	1.20675
5	2	3	1	-0.50543
5	2	4	2	1.510456
5	2	5	3	1.517052
5	2	6	4	0.294995
5	2	7	5	0.770145
5	2	8	6	5.16371
5	3	1	0	0.121968
5	3	2	0	-1.06157
5	3	3	1	-0.56196
5	3	4	2	0.123338
5	3	5	3	-0.6671
5	3	6	4	-0.47342
5	3	7	5	0.408198
5	3	8	6	3.442576
5	4	1	0	-0.71528
5	4	2	0	0.389971
5	4	3	1	-2.34624
5	4	4	2	-1.28451
5	4	5	3	-2.59937
5	4	6	4	0.54254
5	4	7	5	0.191222
5	4	8	6	4.988336
6	1	1	0	0.066613
6	1	2	0	1.444586
6	1	3	1	-3.03758
6	1	4	2	-0.09468
6	1	5	3	-0.33689
6	1	6	4	-1.11077
6	1	7	5	1.536383
6	1	8	6	3.757927
6	2	1	0	-0.63235

Table 4.1 Raw data of the numeric example (continued)

6	2	2	0	-1.06199
6	2	3	1	-3.78191
6	2	4	2	-0.41561
6	2	5	3	-2.76705
6	2	6	4	-1.17362
6	2	7	5	0.028566
6	2	8	6	2.93356
6	3	1	0	0.523345
6	3	2	0	1.303743
6	3	3	1	-2.32899
6	3	4	2	0.014612
6	3	5	3	-1.43351
6	3	6	4	-0.75155
6	3	7	5	-0.76545
6	3	8	6	4.599361
6	4	1	0	0.187322
6	4	2	0	-0.42691
6	4	3	1	-0.78645
6	4	4	2	0.808006
6	4	5	3	-2.19655
6	4	6	4	-0.96966
6	4	7	5	2.316466
6	4	8	6	3.914212
7	1	1	0	1.233668
7	1	2	0	1.244922
7	1	3	1	-0.64528
7	1	4	2	0.897506
7	1	5	3	0.703572
7	1	6	4	0.483171
7	1	7	5	1.767374
7	1	8	6	4.910563
7	2	1	0	0.56995
7	2	2	0	1.496337
7	2	3	1	0.338331

Table 4.1 Raw data of the numeric example (continued)

7	2	4	2	1.686811
7	2	5	3	1.369549
7	2	6	4	1.312299
7	2	7	5	3.071882
7	2	8	6	6.873333
7	3	1	0	2.462185
7	3	2	0	3.40637
7	3	3	1	0.645643
7	3	4	2	2.119472
7	3	5	3	1.591258
7	3	6	4	3.077188
7	3	7	5	3.004698
7	3	8	6	5.517664
7	4	1	0	4.400463
7	4	2	0	3.055759
7	4	3	1	1.620709
7	4	4	2	4.049734
7	4	5	3	2.073217
7	4	6	4	4.361169
7	4	7	5	4.570438
7	4	8	6	8.148809
8	1	1	0	-0.71274
8	1	2	0	-1.607
8	1	3	1	-2.81046
8	1	4	2	-1.17613
8	1	5	3	-1.33326
8	1	6	4	-1.54134
8	1	7	5	-0.69468
8	1	8	6	3.384182
8	2	1	0	0.894211
8	2	2	0	0.739134
8	2	3	1	-0.0863
8	2	4	2	-0.96981
8	2	5	3	0.257691

Table 4.1 Raw data of the numeric example (continued)

8	2	6	4	0.544606
8	2	7	5	0.144456
8	2	8	6	4.566771
8	3	1	0	0.172843
8	3	2	0	-0.96036
8	3	3	1	-2.35915
8	3	4	2	0.548608
8	3	5	3	-2.25037
8	3	6	4	0.215677
8	3	7	5	0.408113
8	3	8	6	4.59947
8	4	1	0	-0.87281
8	4	2	0	-0.06941
8	4	3	1	-0.86424
8	4	4	2	-0.78934
8	4	5	3	-1.94417
8	4	6	4	-2.11236
8	4	7	5	0.00038
8	4	8	6	3.474789

Table 4.2 Analysis of Variance table

| Source	DF	Sum of Squares	Mean Square
Treatment	2	12.8864	6.4432
Environment	7	390.3570	55.7653
Replication	24	190.9926	7.9580
Genotype	5	658.5320	131.7064
Error	217	126.7970	0.5843
Corrected Total	255	1379.5650	

Chapter 5 Simulations
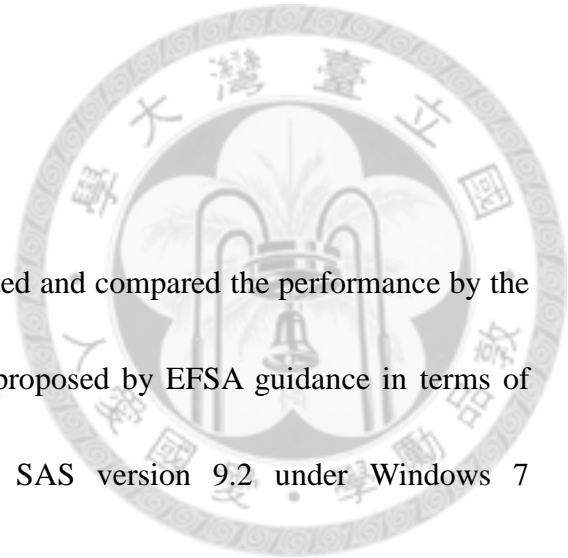
In this chapter, we empirically investigated and compared the performance by the proposed MLS procedure with the approach proposed by EFSA guidance in terms of size and power by simulation study. The SAS version 9.2 under Windows 7 environment and Aspire 4290 of acer was employed to conduct the simulation studies.

5.1 Specifications of Combinations of Parameters for Simulation Designs

Since the EFSA suggests that at least six commercial varieties should be used in the same field design in order to make sure that it is under sufficient condition to provide natural variation for establishing safety threshold. In addition, we assume that no interaction exists between environment and genotype, EFSA also recommends that field trials be conducted at eight sites. As a results, the number of commercial varieties ranges from 6 to 24 by 6. The number of sites is from 10 to 50 by 10 and the number of replicates per site is from 10 to 50 by 10.

The specifications of the parameters for size and power are provided in Table 5.1 and Table 5.2 respectively. The number of simulation samples for each combination is

10000. If the proposed method can adequately control the type Ⅰ error rate at the 0.05

nominal level, 95% of the empirical sizes should be within the range of

$(0.00457, \ 0.0543)$.
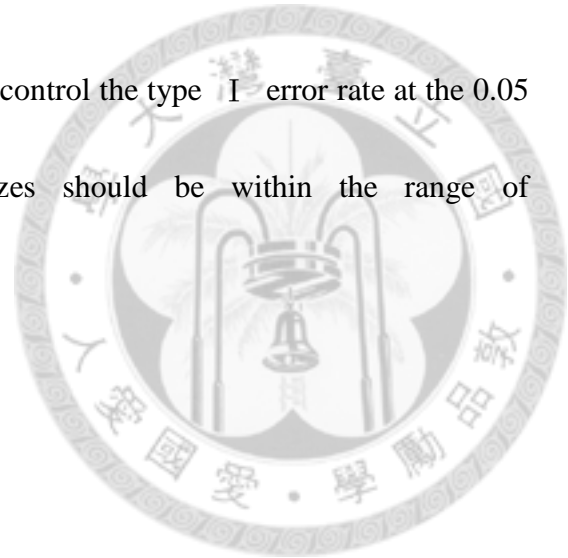
5.2 Simulation Results

5.2.1 Empirical Size

Table 5.3 provides the results on size for the proposed MLS procedure. Table 5.3

reveals that empirical size increases as the number of sites or the number of replicates

per site increases. However, it decreases as the number of commercial varieties.

Furthermore, the empirical size converges to the nominal level of 0.05 when the number

of sites or number of replicates per site approaches 40 or above.

To further explore the impact of interrelationship among the number of site, the

number of replicates, and number of commercial varieties on the size, we conducted

additional simulation studies. The results depicted in Figure 5.1 and Figure 5.2 confirm

those presented in Table 5.3 and Table 5.4 that when the number of sites or the number

of replicates per site is at least 40, the empirical size converges to the nominal level of

0.05. However, the proposed method is conservative when the number of sites or the
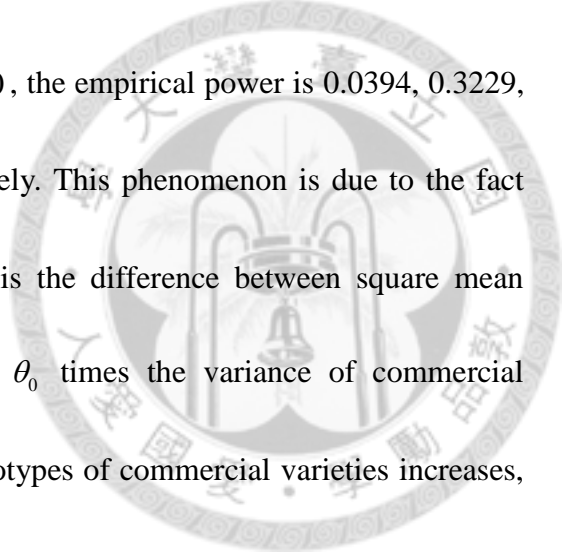
number of replicates is fewer than 40.

Table 5.4 provides the results of size for the EFSA method under the same experimental setting and the specification of parameters as shown in Table 5.1. The results in Table 5.4 reveal that the empirical sizes of the EFSA method are generally smaller than those of the proposed method given in Table 5.3. Therefore the EFSA method is more conservative than our proposed method.

In addition, the empirical size of the EFSA method decreases as either the number of sites or the number of replicates per site increases. This phenomenon is probably due to the fact that the adjusted equivalence limits become narrower as the number of sites or the number of replicates increases. Consequently, the EFSA method is more difficult to reject the null hypothesis of inequivalence as the adjusted equivalence limit narrow due to increase of the number of sites or of the number of replicates per site.

5.2.2 Empirical Power

Table 5.5, Table 5.6 and Table 5.7 provide the empirical power for the variance of the genotype of commercial varieties being 0.2, 0.5, and 3 respectively. The results of empirical power given in these three tables indicate that the empirical power is an increasing function of the variance of genotypes of commercial varieties.

For example, for $q = 6$, $n = 20$, $p = 20$, the empirical power is 0.0394, 0.3229, and 0.9348, for $\sigma_g^2 = 0.2, 0.5$, and 2 respectively. This phenomenon is due to the fact the linearized criterion given Equation (3.3) is the difference between square mean difference $(\mu_G - \mu_C)^2$ and equivalence limit θ_0 times the variance of commercial varieties. Hence, when the variance of the genotypes of commercial varieties increases, the linearized criterion becomes more negative, and hence it is easier to reject the null hypothesis in Equation (3.4). In addition, the empirical power increases as the number of sites or the number of replicates per site increases.

Table 5.5 shows the power is close to 80% when either number of sites or replicates approaches 50 when variance of the commercial varieties is 0.2 and number of commercial varieties is 6. Table 5.6 shows that when either number of sites or number of replicates is 30 above and the variance of the commercial varieties is 0.5, the power approaches 0.8. When the variance of commercial varieties is 3, the empirical power becomes close to 0.8 when the number of sites is at least 10 or above. Figure 5.3 to Figure 5.5 also demonstrates that the empirical power increases as the number of commercial varieties increases. Figure 5.6 to Figure 5.8 also show that the empirical power increases as the number of sites, the number of replicates, and number of commercial varieties increase.

Figure 5.6 indicates that when the number of replicates is 20, it is difficult to achieve power of 80% no matter how large the number of sites or the number of commercial varieties is for $\sigma_g^2 = 0.2$. The power increases rapidly when the number of replicates is 20 or more and the number of sites ranges from 20 to 40.

Figure 5.7 shows that for $\sigma_g^2 = 0.5$ except for the replicates of 20, the powers under the other three replicates are close to each other, and not only meet power of 80% but also approach 100% when the number of sites is 30. When $\sigma_g^2 = 3$, the empirical power can reach 80% when the number of commercial varieties is at least 6, the number of replicates is at least 10 and the number of sites is at least 20 as demonstrated in Figure 5.8.

Figure 5.9 to Figure 5.11 display the response surfaces of empirical power when the number of commercial varieties is 6 for $\sigma_g^2 = 0.2$, 0.5, and 3 respectively. The response surfaces of the empirical power confirm the observations of the relationship between the power and the number of sites, replicates, and commercial varieties and the magnitude of the natural variability of commercial varieties.

On the other hand, Table 5.8 provides the results of power for the EFSA method under the specification of parameters as shown in Table 5.2.

Table 5.1 Specifications of parameters of simulation for size between EFSA method and MLS method

EFSA	MLS
$d = \lvert \mu_G - \mu_C \rvert = 0.223$	$\eta = (\mu_G - \mu_C)^2 - \theta\sigma_g^2 = 0$
$e_i \sim N(0,2)$	$e_i \sim N(0,2)$
$r_{i(j)} \sim N(0,1)$	$r_{i(j)} \sim N(0,1)$
$\sigma_g^2 = 0.248645$	$\sigma_g^2 = 0.248645$
$\theta_0 = 0.2$	$\theta_0 = 0.2$

Table 5.2 Specifications of parameters of simulation for power between EFSA method and MLS method

EFSA	MLS
$d = \lvert \mu_G - \mu_C \rvert = 0$	$\eta = (\mu_G - \mu_C)^2 - \theta \sigma_g^2 < 0$
$e_i \sim N(0,2)$	$e_i \sim N(0,2)$
$r_{i(j)} \sim N(0,1)$	$r_{i(j)} \sim N(0,1)$
$\sigma_g^2 = 0.2$	$\sigma_g^2 = 0.2, 0.5,$ and 3
$\theta_0 = 0.2$	$\theta_0 = 0.2$

Table 5.3 Empirical size of the MLS method with 10000 simulation samples
Definition for the number of commercial varieties as q, the number of sites as n, and the number of replicates per site as p.

q=6	n=10	n=20	n=30	n=40	n=50
p=10	0.0002	0.0023	0.0104	0.0226	0.0323
p=20	0.0018	0.0214	0.0394	0.0467	0.0491
p=30	0.0129	0.0398	0.0476	0.0517	0.053
p=40	0.023	0.0472	0.0497	0.0535	0.0545
p=50	0.0313	0.0484	0.0512	0.0533	0.0569
q=12	**n=10**	**n=20**	**n=30**	**n=40**	**n=50**
p=10	0	0.0009	0.0075	0.0185	0.0338
p=20	0.0011	0.017	0.0365	0.0463	0.0482
p=30	0.007	0.0393	0.0469	0.0502	0.051
p=40	0.0171	0.0474	0.049	0.0526	0.0524
p=50	0.0302	0.0494	0.052	0.0517	0.0549
q=18	**n=10**	**n=20**	**n=30**	**n=40**	**n=50**
p=10	0	0.0001	0.0053	0.0156	0.0274
p=20	0.0001	0.0173	0.0373	0.0464	0.0498
p=30	0.003	0.0386	0.0466	0.051	0.0497
p=40	0.0186	0.0478	0.0497	0.0493	0.0515
p=50	0.0296	0.0485	0.0513	0.0511	0.0514
q=24	**n=10**	**n=20**	**n=30**	**n=40**	**n=50**
p=10	0	0	0.0031	0.0145	0.0244
p=20	0	0.0142	0.0385	0.0457	0.0467
p=30	0.0039	0.04	0.0465	0.0473	0.0473
p=40	0.0151	0.0472	0.0482	0.0491	0.0509
p=50	0.0285	0.0482	0.0473	0.0526	0.0533

Table 5.4 Empirical size of the EFSA method with 10000 simulation samples
Definition for the number of commercial varieties as q, the number of sites as n, and
the number of replicates per site as p.

q=6	n=10	n=20	n=30	n=40	n=50
p=10	0.052303	0.033502	0.030302	0.026501	0.022001
p=20	0.037602	0.023901	0.021601	0.015801	0.012301
p=30	0.028001	0.018201	0.014201	0.011901	0.012401
p=40	0.025401	0.017401	0.010101	0.010301	0.009801
p=50	0.019401	0.014801	0.012201	0.009301	0.0087
q=12	n=10	n=20	n=30	n=40	n=50
p=10	0.040302	0.022701	0.019601	0.015701	0.012401
p=20	0.021601	0.014601	0.014701	0.012201	0.010201
p=30	0.018601	0.014301	0.010401	0.0076	0.0076
p=40	0.017701	0.011801	0.009001	0.0086	0.0068
p=50	0.014001	0.009601	0.0074	0.0085	0.0049

Table 5.5 Empirical power of the MLS method with 10000 samples with $\sigma_g^2 = 0.2$
Definition for the number of commercial varieties as q, the number of sites as n, and the number of replicates per site as p.

q=6	n=20	n=30	n=40	n=50	n=60	n=70
p=20	0.0394	0.1291	0.2201	0.3166	0.403	0.4804
p=30	0.1291	0.2785	0.4139	0.5158	0.5986	0.6599
p=40	0.2311	0.4009	0.5449	0.652	0.7192	0.7665
p=50	0.3215	0.5229	0.6479	0.7254	0.7912	0.8323
p=60	0.4098	0.6011	0.7104	0.7908	0.8416	0.8771
p=70	0.4871	0.6646	0.767	0.8282	0.8739	0.9016
q=12	n=20	n=30	n=40	n=50	n=60	n=70
p=20	0.0294	0.1317	0.2594	0.388	0.5012	0.587
p=30	0.1313	0.3205	0.4923	0.6255	0.7178	0.7803
p=40	0.2623	0.4929	0.6589	0.7716	0.8309	0.8823
p=50	0.3859	0.6177	0.7544	0.8505	0.8992	0.9322
p=60	0.4952	0.7142	0.8351	0.8983	0.9307	0.955
p=70	0.5851	0.7796	0.8816	0.9261	0.9602	0.9736
q=18	n=20	n=30	n=40	n=50	n=60	n=70
p=20	0.0195	0.1259	0.2683	0.425	0.5366	0.6273
p=30	0.128	0.3479	0.5349	0.6688	0.7612	0.8309
p=40	0.2774	0.5324	0.6968	0.8092	0.8712	0.917
p=50	0.4151	0.6645	0.8044	0.885	0.9322	0.959
p=60	0.5332	0.7691	0.8716	0.9308	0.9623	0.9751
p=70	0.6248	0.829	0.9175	0.9582	0.9776	0.9878
q=24	n=20	n=30	n=40	n=50	n=60	n=70
p=20	0.0137	0.126	0.288	0.4368	0.5589	0.6511
p=30	0.1233	0.3733	0.5486	0.6988	0.7853	0.8529
p=40	0.2913	0.5523	0.7292	0.8307	0.8967	0.9402
p=50	0.4421	0.7029	0.834	0.9063	0.9487	0.9707
p=60	0.5675	0.7851	0.8885	0.9464	0.9713	0.9845
p=70	0.6451	0.8506	0.9329	0.9707	0.9837	0.9911

Table 5.6 Empirical power of the MLS method with 10000 samples with $\sigma_g^2 = 0.5$
Definition for the number of commercial varieties as q, the number of sites as n, and the number of replicates per site as p.

q=6	n=10	n=20	n=30	n=40	n=50
p=10	0.005	0.0832	0.2042	0.3225	0.4229
p=20	0.0764	0.3229	0.504	0.6489	0.7264
p=30	0.1963	0.5147	0.6879	0.7878	0.8495
p=40	0.3309	0.6418	0.7934	0.8612	0.9033
p=50	0.428	0.728	0.8531	0.9085	0.9373
q=12	n=10	n=20	n=30	n=40	n=50
p=10	0.0005	0.0674	0.229	0.3889	0.5315
p=20	0.0654	0.3813	0.6205	0.7636	0.8479
p=30	0.225	0.6108	0.817	0.9032	0.9436
p=40	0.3751	0.7598	0.9002	0.9528	0.976
p=50	0.5149	0.8443	0.9429	0.9747	0.9896
q=18	n=10	n=20	n=30	n=40	n=50
p=10	0.0003	0.061	0.2444	0.416	0.5546
p=20	0.0634	0.4188	0.6636	0.8132	0.89
p=30	0.2403	0.6701	0.8486	0.9306	0.9629
p=40	0.4172	0.8065	0.9331	0.973	0.9858
p=50	0.5478	0.8876	0.9662	0.9885	0.9963
q=24	n=10	n=20	n=30	n=40	n=50
p=10	0	0.0584	0.2488	0.4348	0.584
p=20	0.0584	0.4379	0.6934	0.8326	0.908
p=30	0.2515	0.6955	0.8792	0.9464	0.9769
p=40	0.4329	0.8271	0.9465	0.9808	0.9928
p=50	0.5745	0.908	0.9764	0.9929	0.9972

Table 5.7 Empirical power of the MLS method with 10000 samples with $\sigma_g^2 = 3$
Definition for the number of commercial varieties as q, the number of sites as n, and the number of replicates per site as p.

q=6	n=5	n=10	n=15	n=20	n=25
p=5	0.0303	0.1974	0.3684	0.5093	0.6173
p=10	0.2017	0.5103	0.6953	0.7883	0.8482
p=15	0.3812	0.6903	0.8226	0.8849	0.9305
p=20	0.5148	0.7871	0.8907	0.9348	0.9603
p=25	0.6156	0.8466	0.9266	0.9589	0.9716
q=12	**n=5**	**n=10**	**n=15**	**n=20**	**n=25**
p=5	0.015	0.215	0.4535	0.611	0.7336
p=10	0.224	0.6235	0.8105	0.8988	0.9421
p=15	0.4453	0.807	0.9263	0.9637	0.9848
p=20	0.6207	0.8944	0.9642	0.9886	0.9937
p=25	0.7265	0.9419	0.9825	0.9934	0.998
q=18	**n=5**	**n=10**	**n=15**	**n=20**	**n=25**
p=5	0.011	0.2239	0.482	0.6625	0.7794
p=10	0.2273	0.6728	0.8465	0.9319	0.9635
p=15	0.4839	0.8525	0.9531	0.9824	0.9927
p=20	0.6711	0.9269	0.9832	0.9946	0.9985
p=25	0.7803	0.9653	0.9914	0.9981	0.9994
q=24	**n=5**	**n=10**	**n=15**	**n=20**	**n=25**
p=5	0.0062	0.2386	0.5036	0.6861	0.7964
p=10	0.2403	0.6872	0.8705	0.9488	0.9769
p=15	0.5053	0.8754	0.9654	0.9871	0.996
p=20	0.694	0.949	0.9901	0.998	0.9993
p=25	0.8042	0.9762	0.9954	0.9991	0.9998

Table 5.8 Empirical power of the EFSA method with 10000 samples with $\sigma_g^2 = 0.2$
Definition for the number of commercial varieties as q, the number of sites as n, and the number of replicates per site as p.

q=6	n=4	n=6	n=8	n=10
p=4	0.743	0.743	0.765	0.81
p=6	0.781	0.822	0.82	0.812
p=8	0.82	0.816	0.83	0.822
p=10	0.802	0.818	0.823	0.808
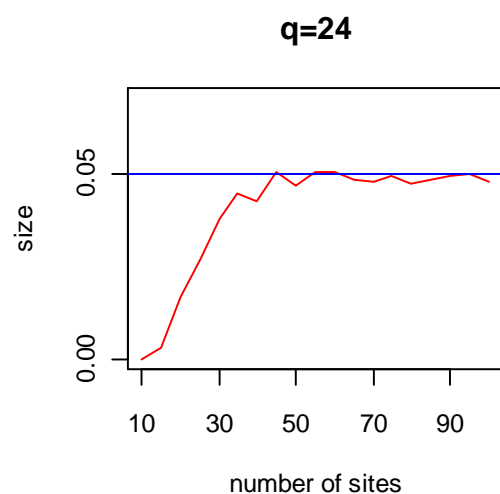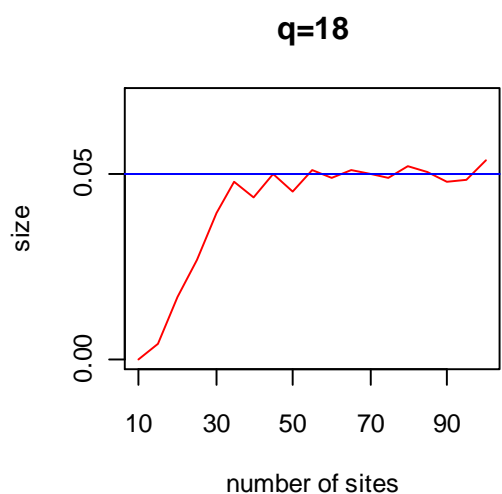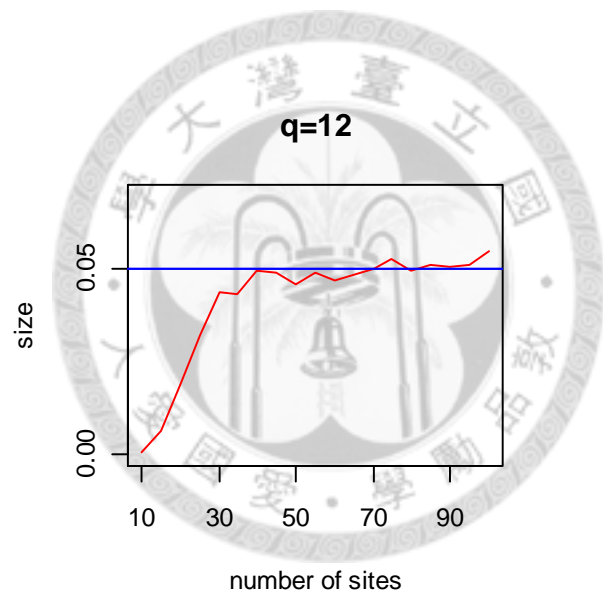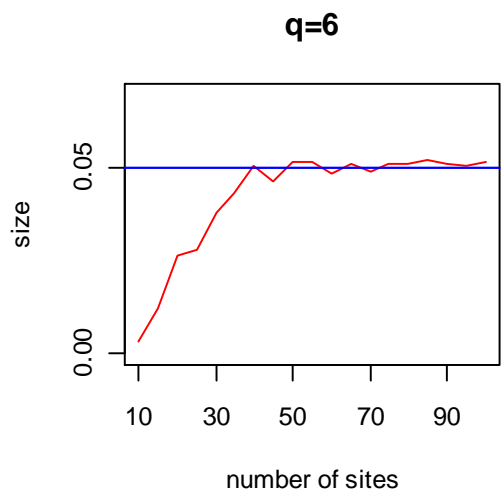
Figure 5.1 Results of convergence about the empirical size under different numbers of

sites and four different numbers of commercial varieties. And q is definite as the

number of commercial varieties.
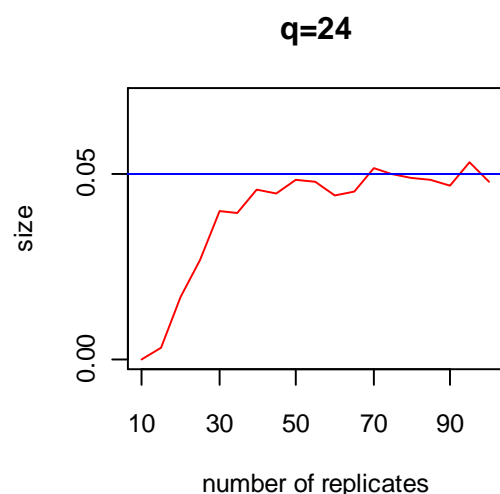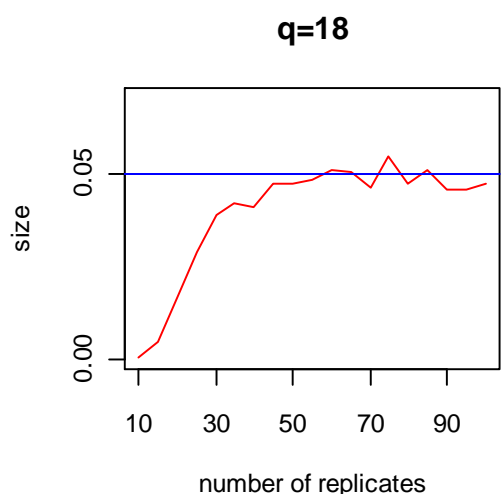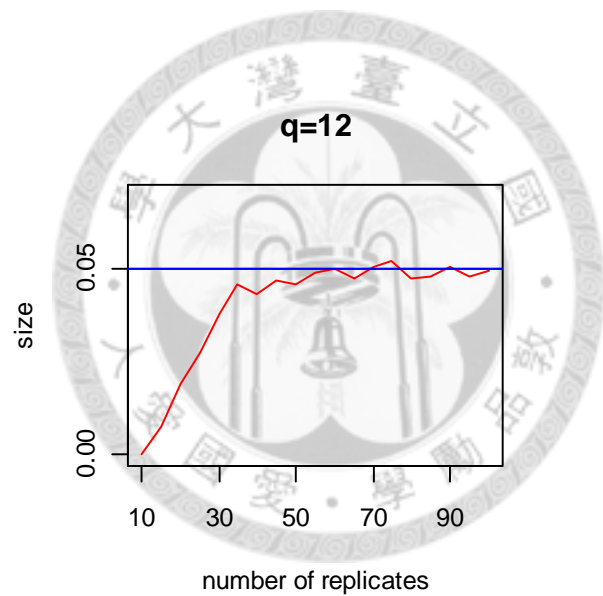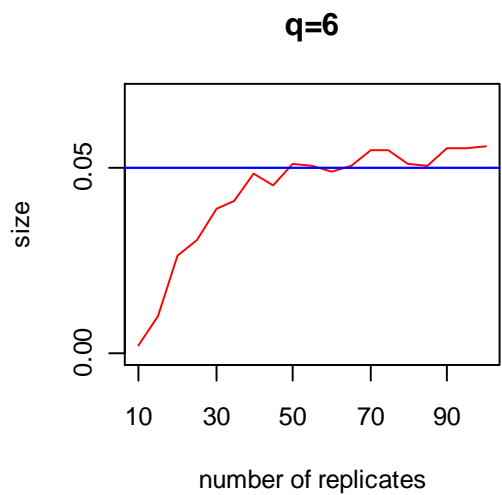
Figure 5.2 Results of convergence about the empirical size under different numbers of replicates per site and four different numbers of commercial varieties. And q is definite as the number of commercial varieties.
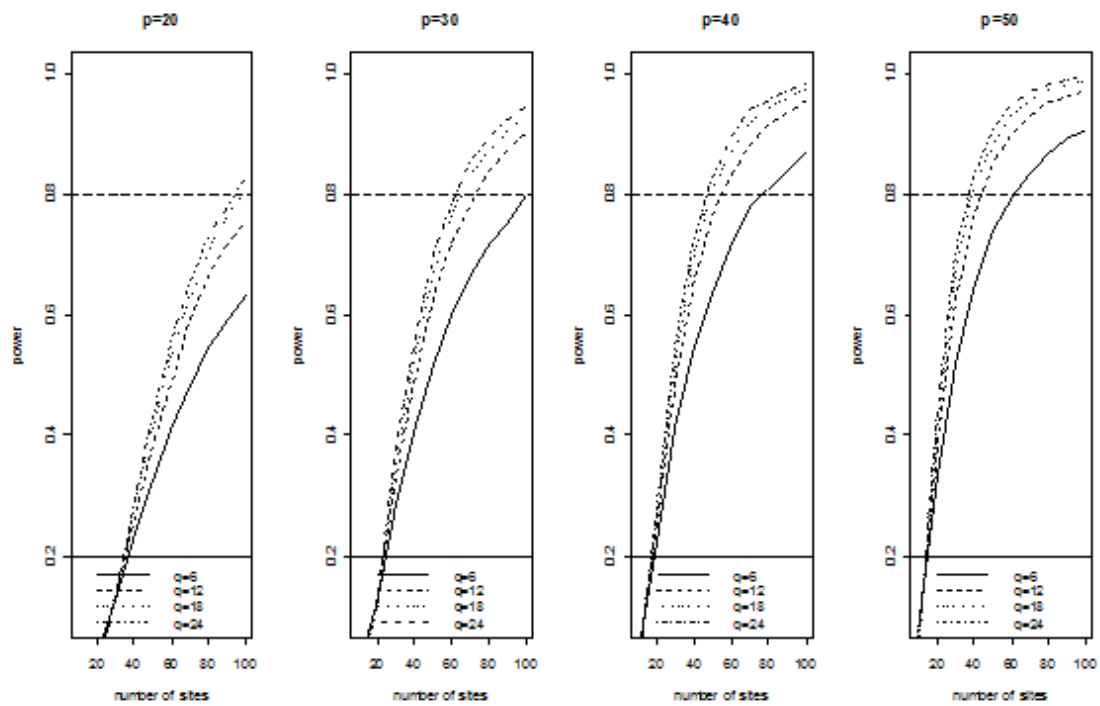
Figure 5.3 Results of empirical power with $\sigma_g^2 = 0.2$ and different numbers of commercial varieties under different replicates. And q is definite as the number of commercial varieties; p is definite as the number of replicates per site.
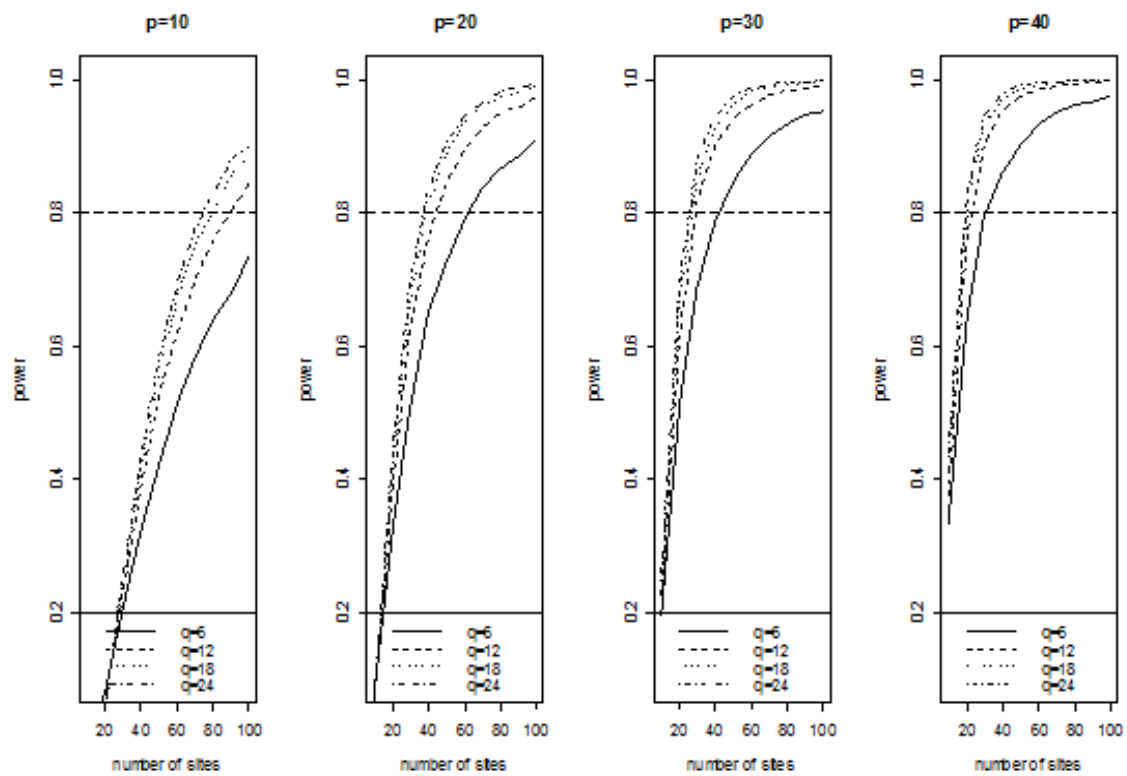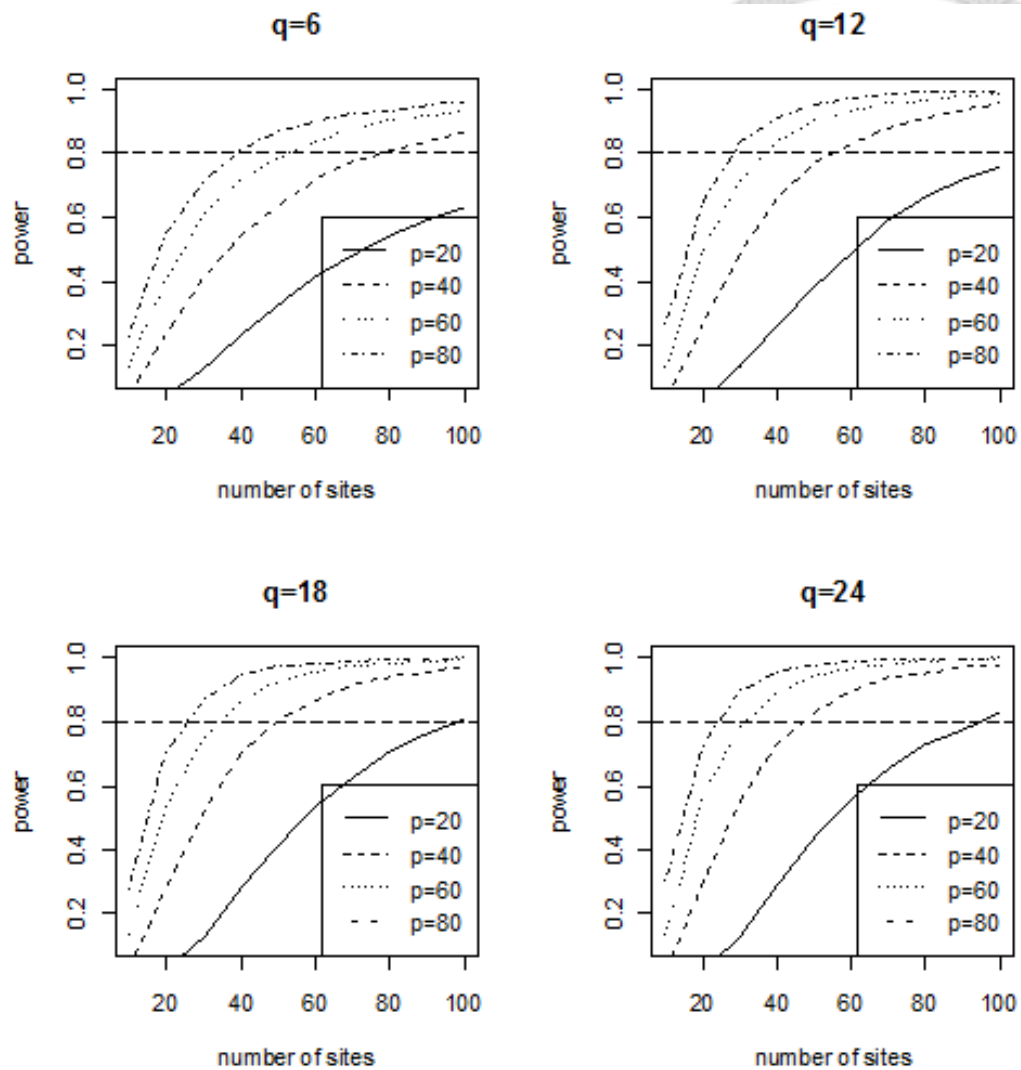
Figure 5.4 Results of empirical power with $\sigma_g^2 = 0.5$ and different numbers of commercial varieties under different replicates. And q is definite as the number of commercial varieties; p is definite as the number of replicates per site.
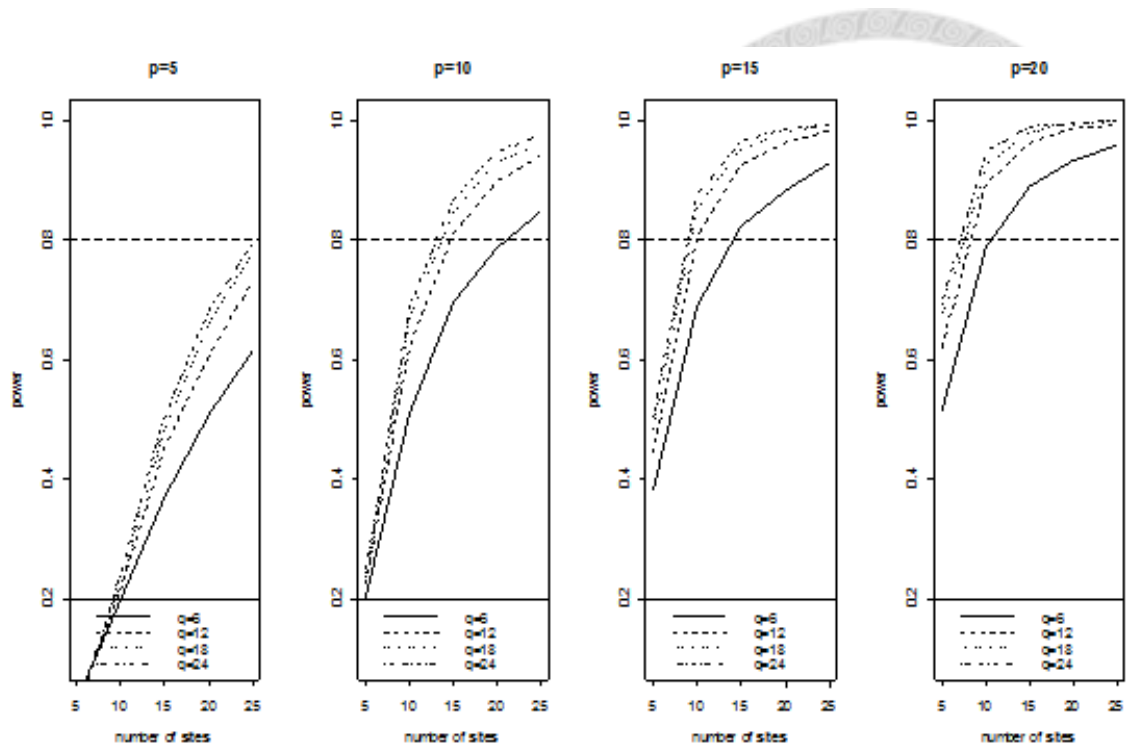
Figure 5.5 Results of empirical power with $\sigma_g^2 = 3$ and different numbers of commercial

varieties under different replicates. And q is definite as the number of commercial

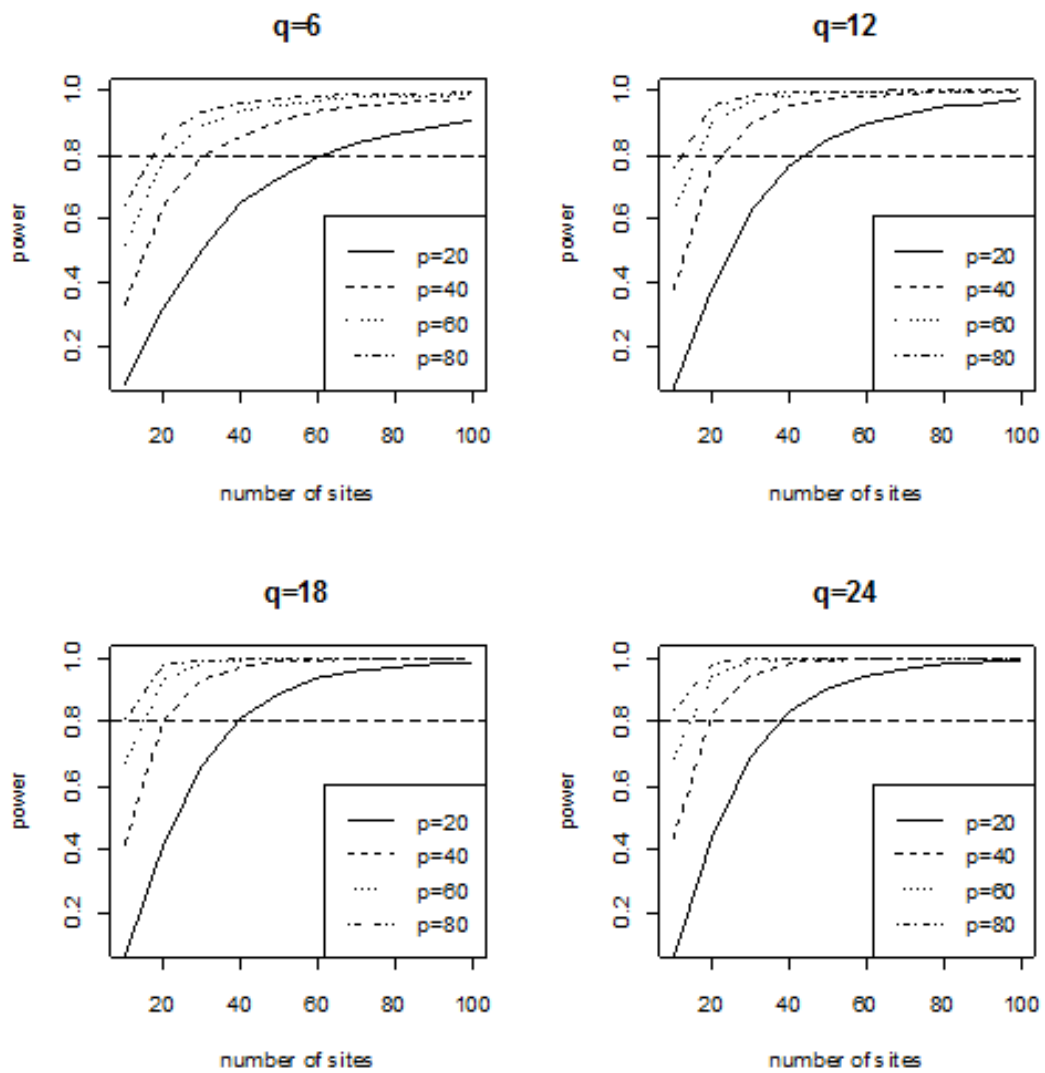varieties; p is definite as the number of replicates per site.

Figure 5.6 Results of simulation for power with $\sigma_g^2 = 0.2$ and different replicates per sites under four different numbers of commercial varieties, respectively. And q is definite as the number of commercial varieties; p is definite as the number of replicates per site.
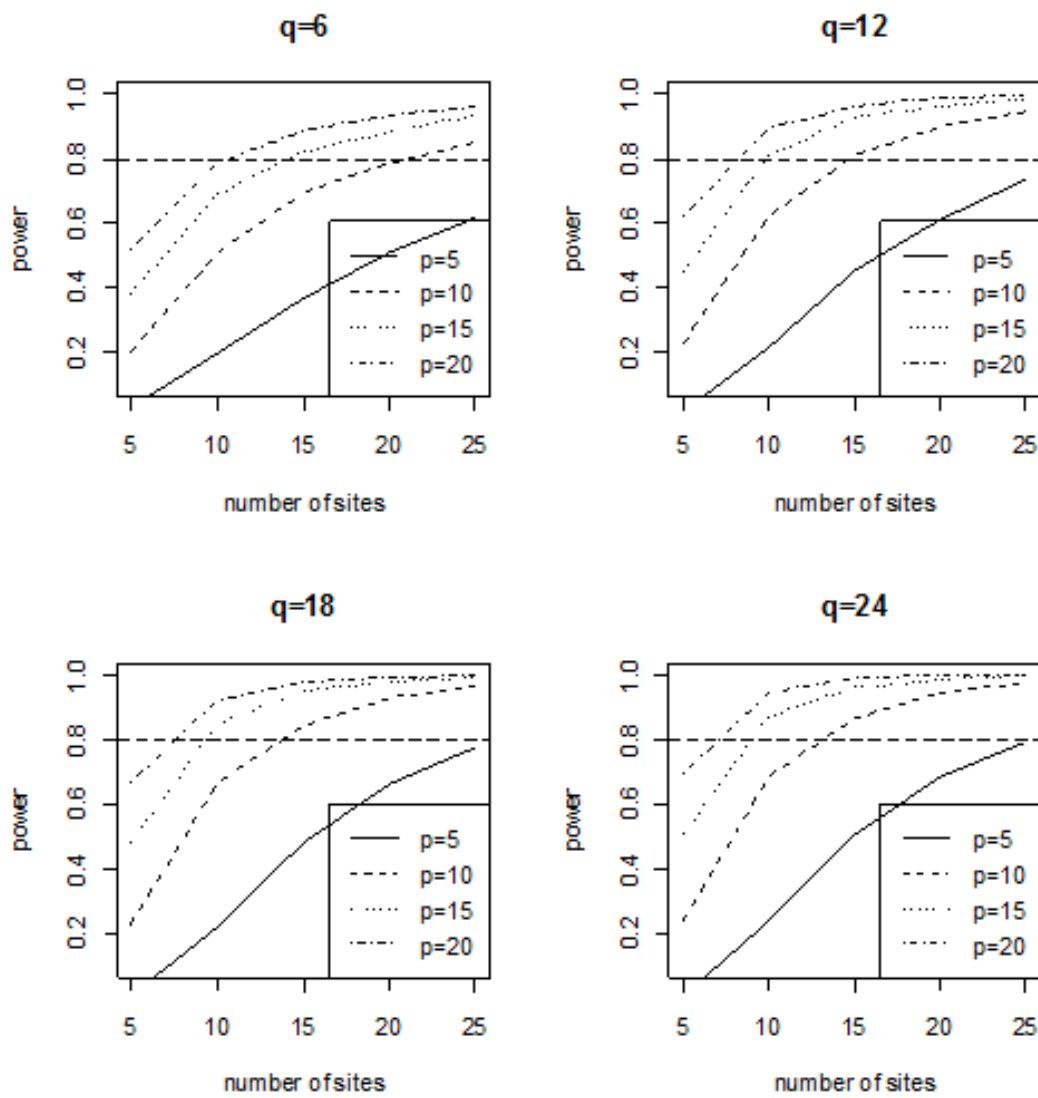
Figure 5.7 Results of simulation for power with $\sigma_g^2 = 0.5$ and different replicates per

site under four different numbers of commercial varieties, respectively. And q is definite

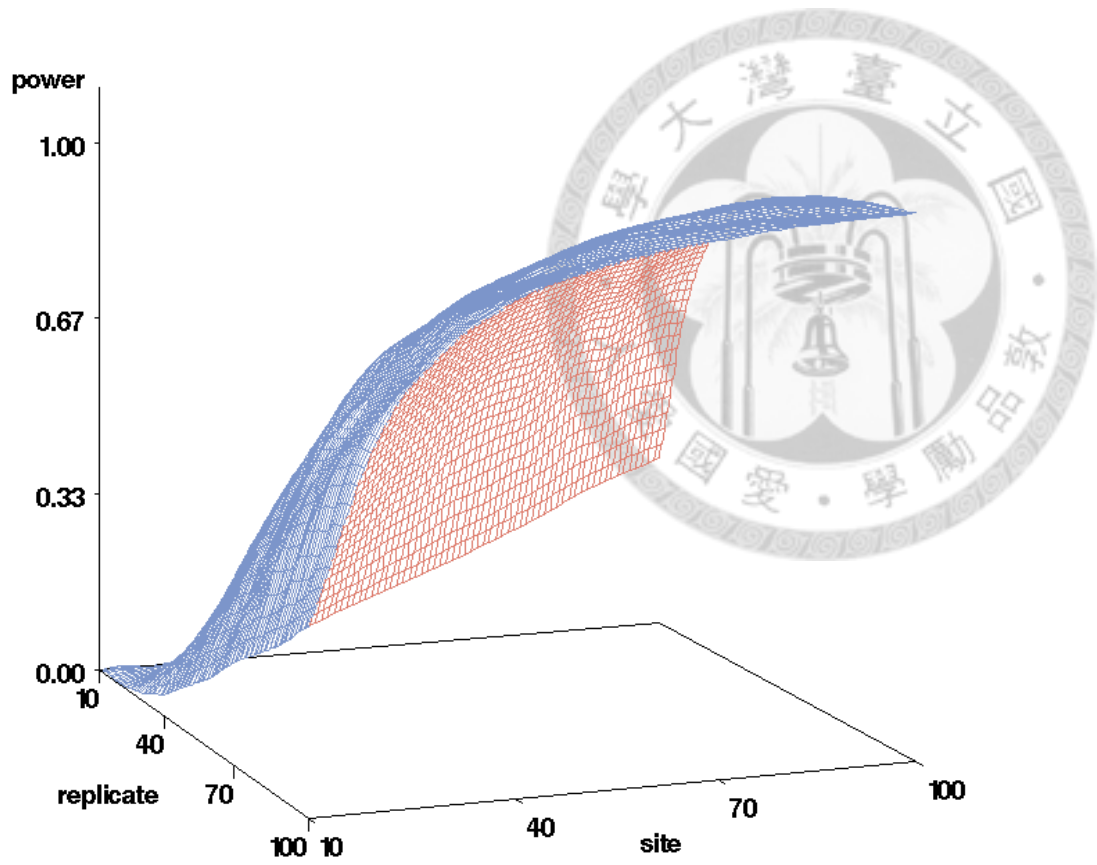as the number of commercial varieties; p is definite as the number of replicates per site.

Figure 5.8 Results of simulation for power with $\sigma_g^2 = 3$ and different replicates per site

under four different numbers of commercial varieties, respectively. And q is definite as

the number of commercial varieties; p is definite as the number of replicates per site.

Figure 5.9 Response surface of power with $\sigma_g^2 = 0.2$ and q=6 under 10000 simulation

samples.
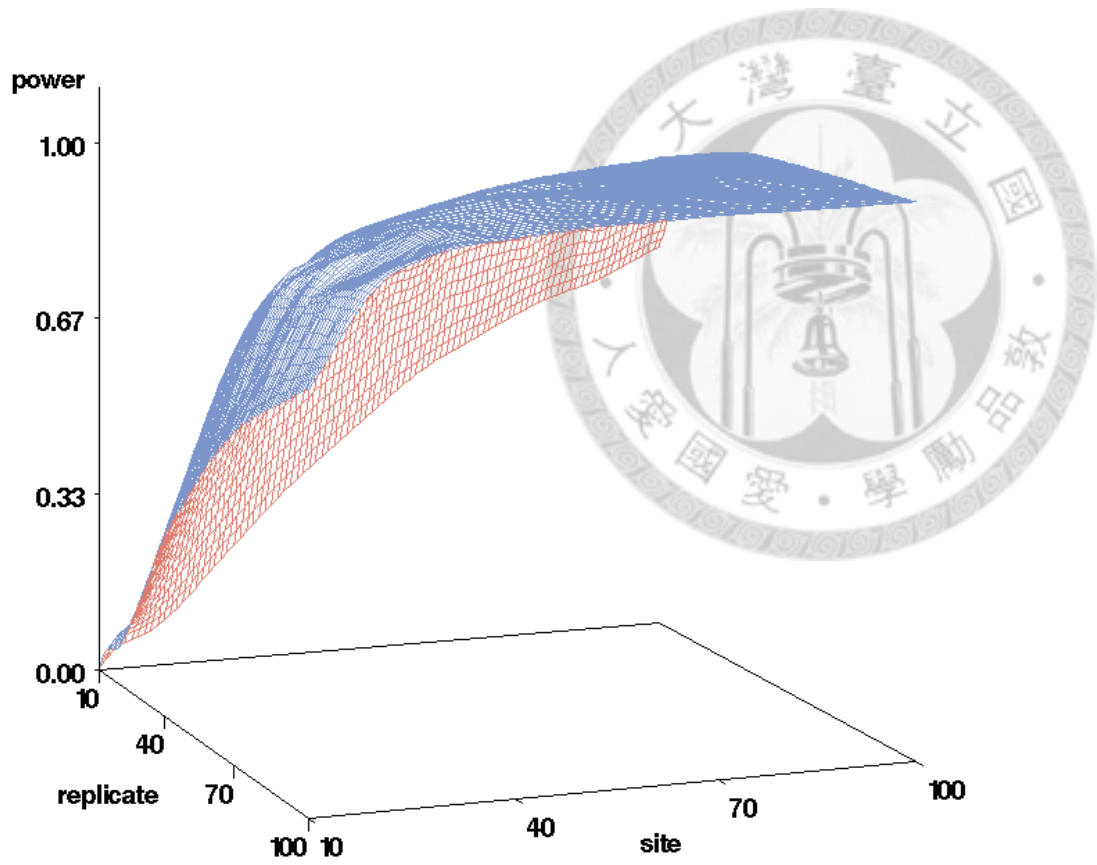
Figure 5.10 Response surface of power with $\sigma_g^2 = 0.5$ and q=6 under 10000 simulation
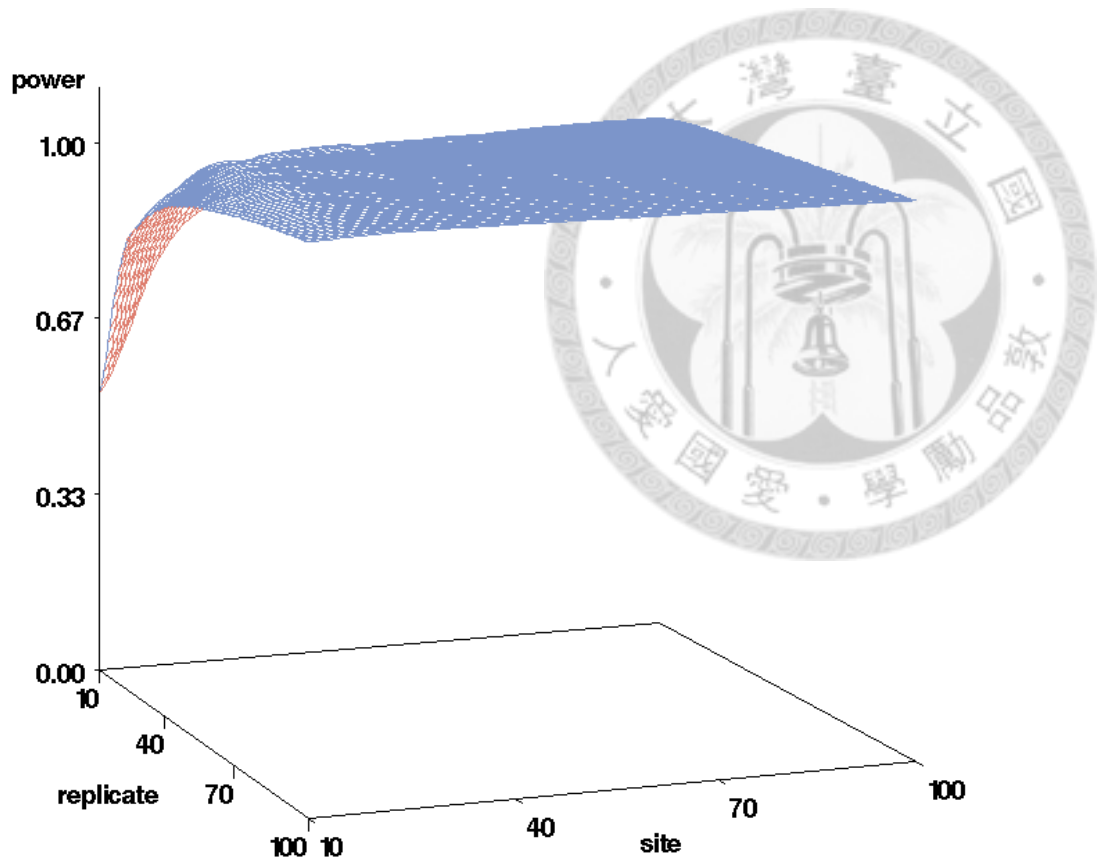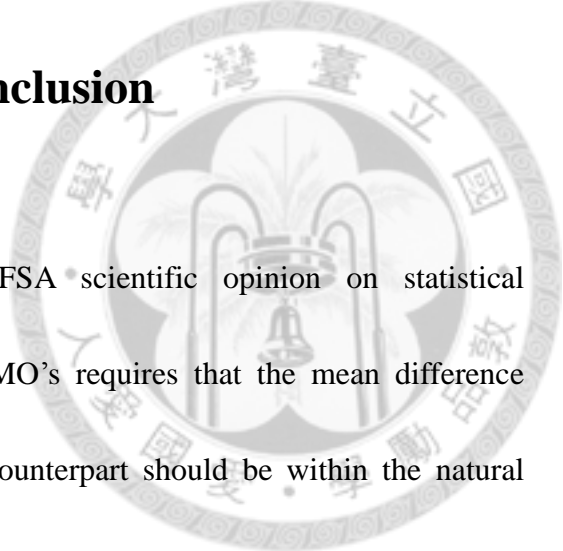
samples.

Figure 5.11 Response surface of power with $\sigma_g^2 = 3$ and q=6 under 10000 simulation

samples.

Chapter 6 Discussion and Conclusion

Because the GMO Panel of the EFSA scientific opinion on statistical consideration for the safety evaluation of GMO's requires that the mean difference between the GM crop and its conventional counterpart should be within the natural variability of available commercially varieties. We proposed a new criterion to achieve this goal. In addition, equivalence hypothesis was formulated based on the linearization of proposed criterion. A statistical inference procedure was proposed to test the equivalence hypothesis. This method rejects the null hypothesis in Equation (3.4) at the α significance level if the $100(1-\alpha)\%$ upper confidence bound constructed by the MLS method is less than 0.

Although the simulation studies show that our proposed method can adequately control the size for moderate to large number of sites and replicates, the power increases as the variance of the commercial varieties increases. In other words, equivalence can easily be established when the variance of commercial varieties becomes large. This shortcoming of the proposed criterion can be resolved if we can restrict the variance of commercial varieties to an upper limit $\sigma_{g_0}^2$. In other words, the scaled criterion can be revised as

$$\theta^{'} = \frac{(\mu_G - \mu_C)^2}{\min(\sigma_{g_0}^2, \sigma_g^2)} \quad\quad (6.1)$$

The performance of the revised criterion required further research in terms of the choice
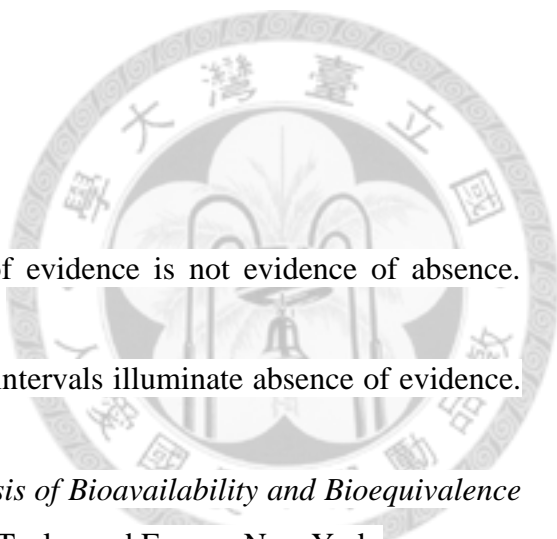
of $\sigma_{g_0}^2$, size and power.

On the other hand, one of the drawbacks of the EFSA method is that both the

confidence interval for the mean difference between the GM crop and its conventional

counterpart and the adjusted equivalence limits given in Equation (2.4) are random

intervals. Therefore, the probability of a random interval within another random interval

is difficult to evaluate. In addition, the adjusted equivalence interval becomes narrow as

the number of sites or the number of replicates increases. It turns out, as shown by the

simulation results, that the empirical size decreases as the number of site or the number

of replicate increases. This is a counter-intuitive phenomenon which requires further

research.

The EFSA statistical considerations also suggest that at least 6 commercial

varieties should be compared concurrently in the same field trial using a randomized

completely block design. The EFSA statistical consideration also provide an example of

design for multiple crops with 4 commercial varieties. Since the block size increases as

the numbers of GM crops, their conventional counterparts, and commercial varieties

increase, then heterogeneity within each block also increases. Therefore, we suggest that

the experiment should be restricted to one GM crop, its conventional counterpart, and 6 commercial varieties for a total of 8 plots per block. For the situation where multiple GM crops with more than 6 commercial varieties are compared in a field trial at a site, a balanced incomplete block design should be used.

As mentioned before, our research is confined only to one variable. However for assessment of equivalence between the GM crop and its counterparts consists of many variables such as agronomic, phenotypes and compositional characteristics, usually exceeding 25. If the requirement for approved of a GM crop is that all characteristics must meet equivalence criterion at the 5% significance level, then its corresponding overall type Ⅰ error rate is 3.0×10^{-33} which is extremely low assuming that all endpoints are independent. This is a problem of multiple correlated must-win comparisons (Julious and Mclntyre, 2012). Further research in this area for evaluation of equivalence for multiple endpoints between the GM crop and its counterpart is required.

References

Altman D. and Bland J.M. (1995) Absence of evidence is not evidence of absence. *British Medical Journal*, 311:485

Altman D. and Bland J.M. (2004) Confidence intervals illuminate absence of evidence. *British Medical Journal*, 328:1016-1017

Chow S. C., Liu J. P. (2010) *Design and Analysis of Bioavailability and Bioequivalence Studies,* 3th Ed., CRC/Chapman and Hall, Taylor and France, New York.

Chow S. C., Shao J., Wang H. (2003) Individual bioequivalence testing under 2x3 designs. *Statistics in Medicine* 2002; 21:629-648

European Food Safety Authority (EFSA) (2010). EFSA GMO Panel opinion on Statistical considerations for the safety evaluation of GMOs. European Food Safety Authority, Parma, Italy

European Food Safety Authority (EFSA) (2001) Directive 2001/18/EC. European Food Safety Authority, Parma, Italy

European Food Safety Authority (EFSA) (2010) EFSA GMO Panel opinion on Statistical considerations for the Safety Evaluation of GMOs. Presentation of the full statistical code used for the example.   European Food Safety Authority, Parma, Italy

European Agency for the Evaluation of Medicinal Products (EMEA). 2001. *Note for guidance on the investigation of bioavailability and bioequivalence.* Document CPMP/EWP/QWP/1401/98, European Agency for the Evaluation of Medicinal Products, Committee for Proprietary Medicinal Products, London, UK.

Food and Drug Administration (FDA). (2001). *Guidance for Industry - Statistical approaches to establishing bioequivalence.* U.S. Department of Health and Human Services, Center for Drug Evaluation and Research. Food and Drug Administration, Rockville, Maryland, 2001.

Hyslop T., Hsuan F., Holder D.J. (2000) Asmall sample confidence interval approach to assess individual bioequivalence. *Statistics in Medicine*; 19:2885-2897

Hsu, J.C., Hwang, J.T.G., Liu, H.K., and Ruberg, S.J. (1994) Confidence intervals

associated with tests for bioequivalence. *Biometrika*, 81, 103-114.

James, C. (2012) Global Status of Commercialized Biotech/GM Crops: 2012. ISAAA Brief No. 44. ISAAA: Ithaca, NY

Julious, S. A., McIntyre, N. E.. (2012) Sample sizes for trials involving multiple correlated must-win comparisons. *Pharmaceutical Statistics,* 11 (2):177-85

Kenward M.G. and Roger J.H. 1997. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 53: 983-997.

Lee, A.F.S. and Fineberg, N.S. 1991. A fitted test for the Behrens-Fisher problem. *Communications in Statistics*-Theory and Methods, 20, 653-666.

Lee Y.H, Shao J. and Chow S.C. (2004). Modified large-sample confidence intervals for linear combinations of variance components: extension, theory, and application. *Journal of the American Statistical Association*, 99, 467-478

Patterson S.D., Jones B. (2012) Observations on scaled average bioequivalence. *Pharmaceutical Statistics*. 11,1-7

Schuirmann D.J. (1987) A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, 15: 657-680.

Tang F.W. (2009) A study on application of Hotelling T-squared to statistical evaluation of genetically modified products. Master thesis, National Taiwan University

Tsai P.Y. (2009) A study on statistical evaluation of in vitro bioequivalence with consideration of variance components. Master thesis, National Taiwan University

World Health Organization (WHO) (2005) *World Health Organization Draft Revision on Multisource (Generic) Pharmaceutical Products: Guidelines on Registration Requirements to Establish Interchangeability*, Geneva, Switzerland.

World Health Organization (WHO) (2002) 20 questions on genetically modified foods, Geneva, Switzerland.

Appendix

SAS Codes for Simulation Program

```
/*Dataset simulation*/

data aaa;

do site=1 to 4;

site_value=rand("normal",0,sqrt(3));

do rep=1 to 4;

rep_value=rand("normal",0,sqrt(1));

do k=1 to 8;

if k=1 then treat= -0.073;

else if k=2 then treat=0.15;

else treat=-0.077;

output;

end;

end;

end;run;
```

```
data bbb;

do k=1 to 8;

if k=1 then genotype=0;

else if k=2 then genotype=0;

else genotype=rand("normal",0,sqrt(2));

do rep=1 to 4;

output;

end;

end;

proc sort data=aaa; by rep k;

proc sort data=bbb; by rep k;

data ccc; merge aaa bbb; by rep k;

e=rand("normal",0,sqrt(0.6));

m=20;

y=site_value+rep_value+treat+genotype+e+m;

proc sort data=ccc; by site rep;

run;

proc sql;
```

```
create table ln as

select site, site_value, rep, rep_value, k, treat, genotype, e, m, log(y) as y

from ccc;

run;

/*End of dataset simulation*/



/*Calculation for mean of conventional counterparts*/

proc sql;

create table ccon as

select mean(y) as y1mean

from ln

where k eq 1;

run;



/* Calculation for mean of GMOs */

proc sql;

create table gmo as

select mean(y) as y2mean
```

```
from ln

where k eq 2;

run;
```

/*Calculation for variance of between GMO and its conventional counterpart*/

```
proc sql;

  create table ccon_y as

select y as y1

from ln

where k eq 1;

run;

proc sql;

create table gmo_y as

select y as y2

from ln

where k eq 2;

run;

data y12;
```

```
merge ccon_y gmo_y;

run;

proc sql;

create table y12var as

select var(y1-y2) as var_y12

from y12;

run;



/* Logarithmic transformation of dataset*/

data ss;

set ln;

if k eq 1 then genotypegroup="comp";

else if k eq 2 then genotypegroup="gmo";

else genotypegroup="ref";

run;

/*Definition of variance components*/

data sss;

set ss;
```

```
if (k eq 1) or (k eq 2) then gmo_genotype=0;

if k eq 3 then gmo_genotype=1;

if k eq 4 then gmo_genotype=2;

if k eq 5 then gmo_genotype=3;

if k eq 6 then gmo_genotype=4;

if k eq 7 then gmo_genotype=5;

else if k eq 8 then gmo_genotype=6;

run;



/*Analysis of varieties for fixed effects*/

proc varcomp method=type1 data=sss;

    class site rep genotypegroup gmo_genotype;

    model y=genotypegroup site rep(site) gmo_genotype     / fixed=1;

    ods output    Estimates=est_varcomp ANOVA=ddd;

run;

data comvarcomp;

set est_varcomp(where=(varcomp="Var(gmo_genotype)"));

comvarcomp=Estimate;
```

```
keep comvarcomp;

run;

/*Capture for required data for MLS method*/

data eee (keep= source MS);

set ddd ;

run;

proc transpose data=eee

                    out=eee_transposed name=MS prefix=MS;

                    id source;

run;



data fff (keep= source DF);

set ddd;

run;

proc transpose data=fff

                    out=fff_transposed name=DF prefix=DF;

                    id source;

run;
```

```
data ggg;

merge fff_transposed eee_transposed    ccon gmo comvarcomp y12var;

run;


/*Construction of 95% upper bound by MLS method*/

data hhh;

set ggg;

n=4;

p=4;

t=3;

q=6;

s=(t-1)+q;

diff=abs(y2mean-y1mean);

A=diff**2;

tinv=tinv(0.95,2*(n*p-1));

/*var_diff=2*(MSsite+(p-1)*MSrep_site_+p*(s-1)*MSError)/(n*p*p*s);*/

ucl=(diff+tinv*sqrt((var_y12)/(n*p)))**2;

C=(ucl-A)**2;
```

```
theta=0.2;

lamda1=-theta*MSgmo_genotype/(n*p);

lamda2=theta*MSError/(n*p);

B=-theta*comvarcomp;

chi95=cinv(0.95,q-1);

chi5=cinv(0.05,DFError);

D=(lamda1**2)*((q-1)/chi95-1)**2+(lamda2**2)*(DFError/chi5-1)**2;

CL=A+B+sqrt(C+D);

proc print data=hhh;run;

---------------------------EFSA guidance of statistical consideration---------------------------

data s;

set ln;

if k eq 1 then genotypegroup="comp";

if k eq 2 then genotypegroup="gmo";

if k ge 3 then genotypegroup="ref";

if genotypegroup eq "ref" then indref=1;

                                    else indref=0;

/*Mixed model analysis for differences*/
```

```
proc mixed data=s CL=WALD;

class site site_value rep rep_value k treat genotype e m genotypegroup y;

model y=genotypegroup/s covb outp=out ddfm=kenwardroger;

random site_value site_value*rep_value indref*genotype;

estimate "gmo_comp" genotypegroup -1 1 0/CL alpha=0.1;

lsmeans genotypegroup;

ods output lsmeans=lsmeans_g estimates=estdif covparms=covparms;

run;

/*Geometric means*/

data lsmeans_g2;

set lsmeans_g;

gm=exp(estimate);

keep y genotypegroup gm;

run;

/*Assessment of differences*/

data estdif1;

set estdif;

ratiod=exp(estimate);lowd=exp(lower);uppd=exp(upper);
```
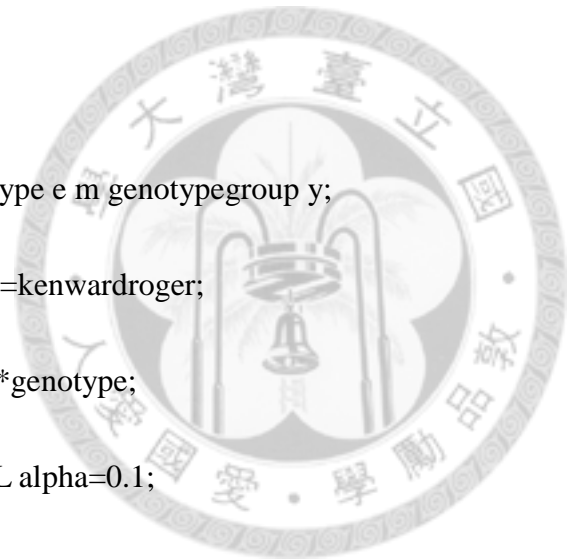
```
dgc=estimate;sed=stderr;dfd=df;

lsddif=sed*tinv(0.95,dfd);

keep y dgc ratiod lowd uppd sed dfd lsddif;

run;
```

```
/*Mixed model for setting equivalence limits*/

proc mixed data=s CL=WALD;

class site_value rep_value genotype genotypegroup;

model y=genotypegroup /s covb outp=out ddfm=kenwardroger;

random site_value site_value*rep_value genotype;

estimate "gmo_ref" genotypegroup 0 1 -1/CL alpha=0.05;

ods output estimates=esteq covparms=covparms;

run;
```

```
/*Keep estimated variance components for genotype*/

data varg;

set covparms(where=(covparm="genotype"));

varg=estimate;
```

```
keep varg;

run;



/*95% equivalence limits ratio GMO/ref*/

data esteq1;

merge esteq varg;

ratioe=exp(estimate);lowe=exp(lower);uppe=exp(upper);

dgr=estimate;see=stderr;dfe=df;

lsdeq=see*tinv(0.975,dfe);

keep varg dgr ratioe lowe uppe see dfe lsdeq;

run;



/*Mixed model analysis for testing non-equivalence*/

proc mixed data=s CL=WALD;

class site_value rep_value genotype genotypegroup;

model y=genotypegroup/s covb outp=out ddfm=kenwardroger;

random site_value site_value*rep_value indref*genotype;

estimate "gmo_ref" genotypegroup 0 1 -1/CL alpha=0.1;
```

```
ods output estimates=estteq;

run;



/*Assessment of equivalences(adjusted scale)*/

data estteq1;

merge estdif1 esteq1 estteq;

lsdteq=stderr*tinv(0.95,df);

fac=lsddif/lsdteq;

lower=-dgr-lsdeq;upper=-dgr+lsdeq;

adjlow=dgc+fac*lower;adjupp=dgc+fac*upper;

adjlow=exp(adjlow);adjupp=exp(adjupp);

keep dgc stderr df lsdteq fac adjlow lowd uppd adjupp varg;

run;



/*Classification of results*/

data classification;

set estteq1;

tol=1e-6;
```

```
noequivtest=(varg<=tol);

sigdif=((lowd>1) or (uppd<1));

equiv=((lowd>=adjlow) and (uppd<=adjupp));

warning=((exp(dgc)<adjlow) or (lowd>adjupp));

nonequiv=((uppd<adjlow) or (lowd>adjupp));

if ((not sigdif) and equiv) then type=1;

if ( sigdif and equiv)              then type=2;

if ((not sigdif) and (not equiv) and (not noequivtest) and (not warning)) then type=3;

if ( sigdif and (not equiv) and (not noequivtest) and (not warning)) then type=4;

if ((not sigdif) and (not equiv) and (not noequivtest) and warning) then type=5;

if ( sigdif and    (not noequivtest) and warning and nonequiv) then type=6;

if (    noequivtest) then type=7;

keep noequivtest sigdif equiv warning nonequiv type;

proc sort data=classification;

by type;

proc print data=classification;

run;
```

------------------------End of EFSA Guidance of Statistical Consideration-------------------

-----------------------------------Size/Power Simulation-----------------------------------

```
data aaa;

do n=10 to 100;

do p=10 to 100;

do q=6,12,18,24;

t=3;

s=(t-1)+q;

dfe=(n*p*s-1)-(n-1)-(n*p-n)-(t-1)-(q-1);

delta=0; /*let delta=0.223 for size simulation*/

sigma_site=2;

sigma_rep=1;

sigma_genotype=0.2; /*let sigma_genotype=0.248645 for size simulation*/

sigma_e=0.6;

theta=0.2;

t=quantile("t",0.95,2*(n*p-1));

q1=((q-1)/quantile("chisq",0.95,(q-1))-1)**2;

q2=(dfe/quantile("chisq",0.05,dfe)-1)**2;

do j=1 to 10000;
```

```
diff=rand("normal",delta,sqrt(2*(sigma_site+sigma_rep+sigma_e)/(n*p)));

MSgenotype=rand("chisq",q-1)*(sigma_e+n*p*sigma_genotype)/(q-1);

MSe=rand("chisq",dfe)*sigma_e/dfe;

A=diff**2;

u=(abs(diff)+t*sqrt(2*(sigma_site+sigma_rep+sigma_e)/(n*p)))**2;

C=(u-A)**2;

B=-theta*(MSgenotype-MSe)/(n*p);

lamda1=-theta*MSgenotype/(n*p);

lamda2=theta*MSe/(n*p);

D=lamda1**2*q1+lamda2**2*q2;

CL=A+B+sqrt(C+D);

if CL<0 then type=1; else type=0;

output;

end;end;end;end;

proc means data=aaa mean;    by p;

output out=result mean(type)=;

run;
```

------------------------------------End of Power Simulation------------------------------------

--Response Surface---

```
goptions reset=global gunit=pct border

                 ftext=swissb htitle=6 htext=3;

data one;

input site replicate power ;

cards;

(data input)

;

run;

proc g3grid data=one out=a2;
   grid site*replicate = power /
      axis1=10 to 100 by 1
      axis2=10 to 100 by 1
        ;
run;

proc g3d data=a2;
   plot site*replicate = power/
     rotate=290 tilt=80
     zmin=0
     zmax=1;

run;
```

--End of Response Surface----------------------------------