國立臺灣大學生物資源暨農學院農藝學研究所

生物統計組博士論文

Division of Biometry, Graduate Institute of Agronomy

College of Bioresources and Agriculture

National Taiwan University

Doctoral Dissertation

定量體外檢驗試劑線性確校統計評估方法之研究

Statistical Evaluation of the Linearity for Quantitative in Vitro Diagnostic Devices

謝宗成

Tsung-Cheng Hsieh

指導教授：劉仁沛 博士，蕭金福 博士

Advisor: Jen-Pei Liu, Ph.D., Chin-Fu Hsiao, Ph.D.

中華民國 98 年 6 月

June, 2009

\

誌謝

猶記十二年前方走出校門踏入職場，而四年前則緣於由職場中的磨練知悉在知識的根底上仍有不足，而再回到校園，重拾學生生活。謝謝四年來指導教授 劉仁沛教授的仔細指導，也謝謝研究室學長、弟妹的多方協助，而能順利拿到學位。最後，則希望把獲得文憑的喜悅與家人共同分享。有了家人的支持，正是能堅持下去的最大動力!

謝宗成 謹誌

中華民國九十八年六月

摘要

在檢定確效性的評估中，線性是最重要的特性之一。目前，評估線性的統計方法是由 Clinical Laboratory Standard Institute (CLSI) EP6-A 準則 所提出。這個方法直接比較點估計值和允許區間並且完全忽略點估計值的抽樣誤差。另一個評估線性的方法是由 Kroll, et al. (2000) 所提出，他使用了線性平均離散程度 (ADL) 當作統計檢定方法，但是卻使用了不正確的統計假設與對應之統計檢定方法。因此，現有兩個方法的型一誤差可能會因而變大而無法做出正確評估。我們提出了雙尾檢定方法與 corrected Kroll's 方法來改善現有方法之缺點。另一方面，我們亦建議了一個以廣義樞紐量(Generalized Pivotal Quantity, GPQ) 為基礎的 ADL 方法來克服由於 ADL 的機率分布存在未知之參數 (nuisance parameter)，而使得型一誤差受到未知之參數干擾的問題。

此外，我們亦建議了兩個新的用來評估線性程度的聚合型測度 (aggregate measure)。其中 SSDL 代表線性離散程度平方和。另一方面，CVDL 則同時考量了變異程度的影響，而定義為相對於變異之線性平均離散程度平方和。經由模擬研究結果顯示，我們所提出各個方法皆比現有由 CLSI EP6-A 準則 與 Kroll et al. 所提出之方法不僅能有效控制型一誤差並且達到一定水準的檢定力。最後，針對我們提出的方法，也利用了數個例子進行資料分析與方法間之比較。

關鍵字：允許區間，線性，量化分析的實驗方法、廣義樞紐量。

Abstract

Linearity is one of the most important characteristics for evaluation of the accuracy in assay validation. The current estimation method for evaluation of the linearity recommended by the Clinical Laboratory Standard Institute (CLSI) guideline EP6-A (Tholen et al., 2003) directly compares the point estimates with the pre-specified allowable limit and completely ignores the sampling error of the point estimates. An alternative method for evaluation of linearity proposed by Kroll, et al. (Kroll, 2000) considers the statistical testing procedure based on the average deviation from linearity (ADL). However this procedure is based on the inappropriate formulation of hypothesis for evaluation of the linearity. Consequently, the type I error rates of both current methods may be inflated for inference of linearity. Therefore, we propose a two one-sided test (TOST) procedure and a corrected Kroll's procedure as the more appropriate procedure for assessment of linearity. On the other hand, for the purpose to overcome the issue raised by the unknown nuisance parameters of the distribution of ADL, the GPQ-based ADL procedure is also proposed.

In addition, we introduced two new alternative measures SSDL and CVDL which are defined as the sum of square of deviations from linearity and the deviations scaled by the variability, respectively, as the aggregate criteria for assessment of linearity. Unlike ADL and SSDL, CVDL can consider linearity and repeatability of an assay method simultaneously. The relationship among the dofferent aggregate criteria is discussed. The simulation studies are conducted to empirically investigate the size and power among the current and proposed methods. The simulation results show that all proposed methods can adequately control size better than the current methods. Numerical

examples are also used to illustrate the application of the proposed methods.

Keyword: Allowable Limit, Linearity, Quantitative analytical laboratory methods, Generalized Pivotal Quantity.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

In validation of quantitative analytical laboratory procedures, one of the most important characteristics of the accuracy is the linearity. The ICH Q2A guideline (ICH Expert Working Group, 1995), defines the linearity of an analytical method as its ability (within a given range) to obtain the test results, which are directly proportional to the concentration (amount) of the analyte in the test sample. The objective for evaluation of linearity is to validate existence of a mathematically verified straight-line relationship between the observed values and the true concentrations or activities of the analyte. Linearity represents the simplest mathematical relationship and it also permits simple and easy interpolations of results for clinical practitioners. The approved Clinical Laboratory Standard Institute (CLSI) guideline EP6-A (Tholen et al., 2003) recommends that at least five solutions of different concentration levels across the anticipated range be included in an experiment for evaluation of linearity. At each concentration level, two to 4 replicates should be run. With respect to EP6-A, if the difference between the best-fit nonlinear polynomial curve and simple linear regression equation at each concentration is smaller than some pre-defined allowable bias δ_0, the linearity then can be claimed. For instance, in Figure 1.1, it shows that the linearity is claimed because the magnitude of deviation from linear regression at all concentrations for the best-fitted model, i.e., the quadratic model in the figure, are less than δ_0, while the linearity cannot be claimed in Figure 1.2 since the magnitude of deviation from

linear regression at concentration S2, S3 and S4 for the best-fitted model are greater than δ_0. On the other hand, Kroll, et al. (Kroll, 2000) proposes a statistical testing procedure based on the average deviation from linearity (ADL) which is defined as the square root of the average squared distances between the fitted concentrations of the best fit polynomial curve and the simple regression equation at each solution level, standardized by mean concentration. The linearity is concluded at the α nominal level if the observed value of the ADL is smaller than the upper α quantile of the sampling distribution of the observed ADL.

However, the procedure for assessment of linearity based on ADL proposed by Kroll, et al. (Kroll, 2000) is derived from formulating of the hypothesis for proving the linearity as the null hypothesis. On the other hand, the method for evaluation of linearity recommended by EP6-A directly compares the point estimates with the pre-specified allowable limit and completely ignores the sampling error of the point estimates. As results, the type I error rate may be inflated and the probability of the incorrect claims of linearity is not

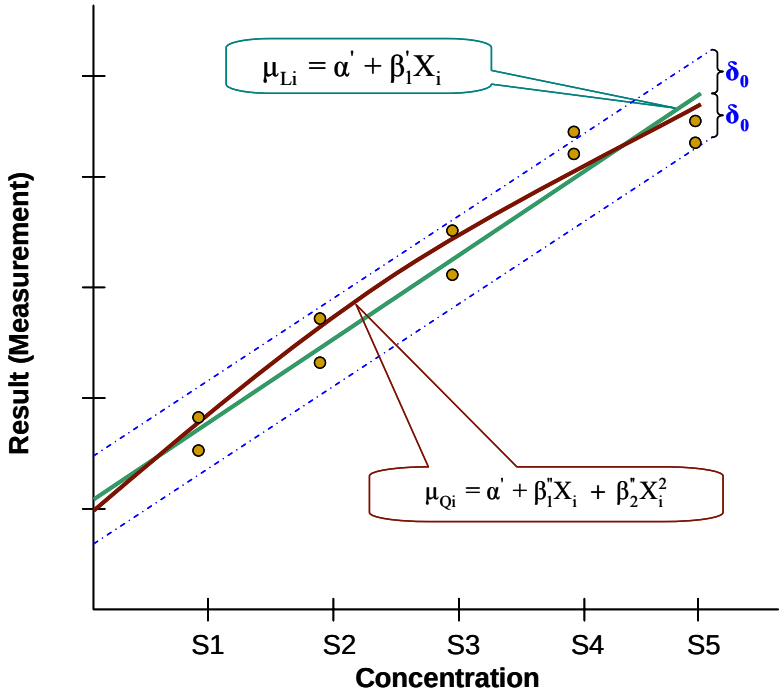


Figure 1.1 Acceptance of linearity by CLSI EP6-A guideline

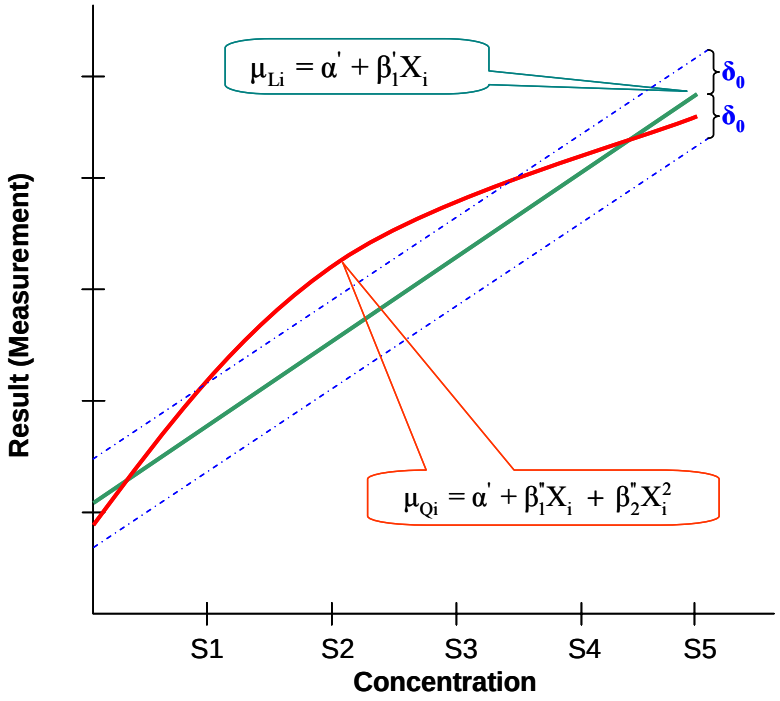


Figure 1.2 Un-acceptance of linearity by CLSI EP6-A guideline

adequately controlled at the nominal level. Therefore, in our research, we propose a two one-sided test (TOST) procedure and a corrected Kroll's procedure for assessing the linearity in validation of quantitative analytical laboratory procedures for solving these shortcomings of assessment procedures proposed by CLSI guideline EP6-A and Kroll's ADL method, respectively.

In addition to the inappropriate statistical hypothesis, the sampling distribution of the observed ADL of the Kroll's method is a function of a non-central chi-square distribution. It follows that the Kroll's method suggests using the estimate of the unknown non-centrality parameter as the true parameter. Hence, the variability associated with the estimated non-centrality parameter is completely ignored in the Kroll's procedure. Tsui and Weerahandi (Weerahandi, 1993) propose the generalized confidence interval based on the generalized pivotal quantity (GPQ) for the exact statistical inference. The method proposed by Tsui and Weerahandi (Tsui and Weerahandi, 1989; Weerahandi, 1993) can eliminate the unknown parameters by replacing them using the appropriate random variables. As a result, we propose to apply the concept of generalized confidence interval based on the generalized pivotal quantity to overcome the issue of unknown non-centrality parameter of the distribution of ADL. The linearity in assay validation can be concluded if the $100(1 - \alpha)$ % upper generalized confidence limit of ADL is less than pre-specified limit.

In addition, we also propose two new measures for assessment of linearity in assay validation. As mentioned above, the approved CLSI EP6-A recommends that for proving the linearity, the deviations from linearity, defined as the difference between the best-fitted nonlinear polynomial curve and simple linear regression equation, be smaller than some pre-defined allowable bias, say δ_0, at all concentrations. Therefore, we

propose the sum of squares of deviations from linearity (SSDL) which is formulated based on the nature of the criterion proposed by CLSI EP6-A guideline as an alternative metric for evaluation of the linearity in assay validation. On the other hand, the repeatability is also a important characteristic which stands for reliability of an assay method which is defined as the ability of a measuring system/instrument to provide closely similar indications for repeated applications of the same measurand under the same conditions of measurement. However, both ADL and SSDL do not take the experimental variability into consideration. Therefore, we proposed the coefficient of variation of the deviations from linearity (CVDL) which is the scaled deviation scaled by the variability of the best-fitted model as an alternative measure for assessment of linearity.

In the next chapter, the experiment designs for evaluation of linearity and assessment procedure recommended by the approved CLSI guideline EP6-A (Tholen et al., 2003) is introduced first. The assessment procedures proposed by the EP6-A guideline (Tholen et al., 2003) and Kroll et al. (Kroll, 2000) are then reviewed, respectively. The shortcomings of these two methods are highlighted. Various measures for assessing linearity based on disaggregate criterion and aggregate criterion are introduced in Chapter 3. Their corresponding statistical hypotheses are also provided. In Chapter 4, the proposed TOST and corrected Kroll's method to overcome the shortcomings of CLSI EP6-A and the current Kroll's method are introduced. The concept of GPQ and generalized confidence interval developed by Tsui and Weerahandi (Tsui and Weerahandi, 1989; Weerahandi, 1993) is introduced in Chapter 5. The GPQ-based ADL approach derived from the method of Tsui and Weerahandi (Tsui and Weerahandi, 1989; Weerahandi, 1993) for overcoming the issue of the unknown parameter of the distribution of ADL is then proposed. In Chapter 6 and 7, we propose two new

measures of SSDL and CVDL for assessment of linearity. The relationship and comparison among the introduced aggregate criteria are addressed in Chapter 8.

The results of the simulation studies to compare the empirical size and power between the current methods and proposed methods are summarized in each chapter. All simulation programs were written by Compaq Visual Fortran Professional Edition 6.6.0 under Microsoft Window operation system of the IBM compatible personal computer. The numerical examples are also provided to introduce the implementation of each proposed method. The final concluding remarks are provided in Chapter 9.

Chapter 2

Literature Review

2.1 Experiment Design

The approved CLSI guideline EP6-A (Tholen et al., 2003) recommends that the experiment for linearity assessment should be conducted at least five solutions of different concentrations run at least in duplicates. Let Y_{ij} be the test result of replicate j at concentration X_i, where j = 1,...,R; i=1,…,L. The approved CLSI guideline EP6-A considers the following linear, quadratic, and cubic models fitting the data obtained from the experiment:

Linear (First order) $\mu_{Li} = \alpha' + \beta_1' X_i$

Quadratic (Second-order polynomial) $\mu_{Qi} = \alpha'' + \beta_1'' X_i + \beta_2'' X_i^2$

or (2.1.1)

Cubic (Third-order polynomial) $\mu_{Ci} = \alpha''' + \beta_1''' X_i + \beta_2''' X_i^2 + \beta_3''' X_i^3$

where μ_{Li}, μ_{Qi}, and μ_{Ci} are the predicted mean of the corresponding models and $\alpha', \alpha'', \alpha'''; \beta_1', \beta_1'', \beta_1'''; \beta_2'', \beta_2'''$, and β_3''' are the intercepts, regression coefficients for the corresponding models in (2.1.1). In what follows all assumptions for fitting the best-fitted model specified in the approved CLSI guideline EP6-A are satisfied and we use the definition of the best-fitted model recommended by the EP6-A (Tholen et al., 2003). The best-fit model is the model such that lack-of-fit is not statistically significant

and the repeatability meets the manufacturer's claim. In addition, for the purpose of illustration, we also assume the random error is assumed to be approximately constant rather proportional in the range of concentrations considered in the experiment.

The spirit of the EP6-A guideline (Tholen et al., 2003) is to determine the concentrations(s) where an assay method is not linear and the extent of the nonlinearity at that level. As addressed in the guideline "The guideline emphasizes the necessity that each user establishes his or her requirements for linearity, or the allowable error due to nonlinearity. It also places less importance on global tests such as lack-of-fit test for linearity across the tested range. Global tests merely indicate that statistically significant nonlinearity exists; they do not show where that nonlinearity is, nor do they show the magnitude of the error.", therefore, even if the best-fitted model is a nonlinear model, it does not necessarily imply that the assay can not be concluded linear. Based on the above concept, the guideline proposes the following rule in instead of global test for assessing linearity: if the best-fitted model is the linear model over the some range of concentrations employed in the experiment, then the assay method can be concluded to be linear over the some range of concentrations. However, if the best-fitted model is not linear, the linearity of the analytical procedure can still be claimed if the magnitude of deviations from the linearity at each concentration is within some pre-specified allowable limit of δ_0 as showed in Figure 1.1.

2.2 Evaluation Procedure of CLSI Guideline EP6-A

Based on the suggested experiment design in Section 2.1, CLSI guideline EP6-A (Tholen et al., 2003) proposes the following procedure for assessment of the linearity in assay validation. Let the difference in predicted means between the best-fit nonlinear

and linear model $\mu_{Pi} - \mu_{Li}$ which represents a measure for the degree of the deviation from linearity at each concentration level. The hypothesis for evaluation of linearity can be formulated as

$$H_{0i}: |\mu_{Pi} - \mu_{Li}| \geq \delta_0 \text{ vs. } H_{ai}: |\mu_{Pi} - \mu_{Li}| < \delta_0, \text{ for all i=1 ,..., L .} \quad (2.2.1)$$

Let $\hat{Y}_{Pi}$ and $\hat{Y}_{Li}$ be the least squared (LS) estimators of the predicted mean of the best-fit and linear models, respectively, where

$\hat{Y}_{Li} = a' + b_1' X_i$, and

$$\hat{Y}_{Pi} = \begin{cases} a'' + b_1'' X_i + b_2'' X_i^2, \text{ if the best-fitted model is quadratic,} \\ a''' + b_1''' X_i + b_2''' X_i^2 + b_3''' X_i^3, \text{ if the best-fitted model is cubic;} \end{cases}$$

and a', a'', a''' ; b_1', b_1'', b_1''' ; b_2'', b_2''' , and b_3''' are the LS estimators of the intercepts, regression coefficients for the corresponding models in (2.1.1).

According to the approved CLSI EP6-A guideline (Tholen et al., 2003), the linearity of the proposed analytical method can be concluded if

$$\left|\hat{Y}_{Pi} - \hat{Y}_{Li}\right| < \delta_0 \text{ , for i = 1,…, L.} \quad (2.2.2)$$

This method is referred to as the estimation method because it only considers the estimators for evaluation of linearity. The estimation method completely ignores the variability and distribution associated with the estimators. Therefore, it also may inflate the type I error rate in assessment of linearity.

2.3 Uncorrected Kroll's Procedure

Kroll, et al. (Kroll, 2000) considers the average deviation from linearity (ADL) for assessment of linearity. The ADL is defined as

$$\theta = ADL = \frac{\sqrt{\sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2/L}}{\mu}, \qquad (2.3.1)$$

where μ is the population mean concentration for all solutions of the assay.

Since ADL is a function of standardized sum of squares of the differences in the predicted means between the best-fitted and linear models, it is an aggregate criterion. Therefore, the hypothesis for evaluation of linearity based on ADL proposed by Kroll, et al. (Kroll, 2000) is given as

$$H_0: \theta \le \theta_0 \text{ vs. } H_a: \theta > \theta_0, \qquad (2.3.2)$$

where θ_0 is the maximum allowable average deviation from linearity. θ_0 is usually presented as percent. Kroll, et al. (Kroll, 2000) refer it to as percent bound and suggest 5% as a reasonable cutoff for most relevant clinical applications.

An estimator of ADL suggested by Kroll, et al. (Kroll, 2000) is given as

$$\hat{\theta} = \frac{\sqrt{\sum_{i=1}^{L}(\hat{Y}_{Pi} - \hat{Y}_{Li})^2 / L}}{\overline{Y}}, \qquad (2.3.3)$$

where $\overline{X}$ is the observed mean concentration for all solutions of the assay.

At $\theta = \theta_0$, $\sum_{i=1}^{L}(\hat{Y}_{Pi} - \hat{Y}_{Li})^2$ follows a non-central chi-square distribution with degrees of freedom d -1, and non-centrality parameter $LJ\theta_0^2/(\sigma/\mu)^2$, where d is the degrees of freedom for regression of the best-fitted model and σ^2 is the variance of residuals under the best-fitted model. The decision rule for concluding that the assay is linear at the 5% significance level if

$$\hat{\theta} < \frac{\sigma}{\mu}\sqrt{\frac{q_{0.95}}{LJ}}, \qquad (2.3.4)$$

where $q_{0.95}$ is the 95th percentile of a non-central chi-square distribution with degrees of

freedom d -1 and non-centrality parameter $LJ\theta_0^2/(\sigma/\mu)^2$. We refer this method to as the uncorrected Kroll method.

The method proposed by Kroll, et al. (Kroll, 2000) has two shortcomings. The hypothesis for proving linearity is formulated as the null hypothesis. When Equation (2.3.4) is satisfied, the only conclusion is that the null hypothesis is not rejected and this does not imply that the linearity of the assay is proved. On the other hand, the critical value in (2.3.4) contains the unknown parameters μ and σ that need to be estimated from the data. Kroll, et al. (Kroll, 2000) suggested to estimate μ by $\overline{X}$, the observed mean concentration for all solutions of the assay and σ by the square root of residual mean square obtained the best-fitted model. Consequently, the variability associated with residual mean square is not considered in evaluation of linearity by Eq. (2.3.4). Because of these two shortcomings, the method based on ADL proposed by Kroll, et al. may not adequately control the type I error rate at the nominal level for evaluation of linearity.

2.4 Summary

As we introduced as above, both the current estimation method of CLSI EP6-A guideline and uncorrected Kroll's method for linearity assessment in assay validation will inflate the type I error. In particular, the uncorrected Kroll method will also conclude the linearity incorrectly because of the formulation of the incorrect hypothesis and corresponding rejection rule. In Chapter 3, we will introduce various measures for assessing linearity based on the aggregate criterion and disaggregate criterion which will be discussed and compared in our research. In addition to the statistical testing procedures corresponded to the new proposed measures, the two new methods for

improving the shortcoming of the current methods in this chapter will also be proposed in Chapter 3. The comparison of their performances in empirical sizes and powers are made by the simulation study.

Chapter 3

Criterion for Assessing Linearity

In this chapter, we summarize the measures for assessing linearity based on the disaggregate criterion and aggregate criterion which are reviewed and proposed in our research. Their corresponding statistical hypotheses are also introduced. In addition, the discussion for difference of the disaggregate criterion and aggregate criterion on the impact of their performance of assessment linearity are also addressed .

3.1 Disaggregate Criterion

As we introduced in Section 2.2 of Chapter 2, following the experiment recommended by EP6-A (Tholen et al., 2003), the guideline proposes that even though the best-fitted model is not linear, the linearity of the analytical procedure can be claimed if the magnitude of deviations from the linearity at each concentration is within some pre-specified allowable limit of δ_0. The hypothesis corresponded to the proposed evaluation rule can be formulated as

$$H_{0i}: |\mu_{Pi} - \mu_{Li}| \geq \delta_0 \text{ vs. } H_{ai}: |\mu_{Pi} - \mu_{Li}| < \delta_0, \text{ for all } i=1,..., L. \quad (3.1.1)$$

where the difference in predicted means between the best-fit nonlinear and linear model $\mu_{Pi} - \mu_{Li}$ represents a measure for the degree of the deviation from linearity at each concentration level. Since hypothesis (3.1.1) requires all differences in the predicted means between the best-fitting and linear models be within the pre-specified allowable limit, it is a disaggregate criterion.

3.2 Aggregate Criterion

3.2.1 Average Deviation from Linearity (ADL)

Recall the definition of ADL proposed by Kroll et al. (Kroll, 2000) defined as the following:

$$\theta = \mathrm{ADL} = \frac{\sqrt{\sum_{i=1}^{L} (\mu_{Pi} - \mu_{Li})^2 / L}}{\mu}, \qquad (3.2.1.1)$$

where μ is the population mean concentration for all solutions of the assay.

ADL is a scaled deviation defined as the square root of sum of squares of the difference in predicted means between the best-fitted and linear models. The correct hypothesis for evaluation of linearity based on ADL proposed by Kroll, et al. (Kroll, 2000) is given as

$$H_0: \theta \geq \theta_0 \text{ vs. } H_a: \theta < \theta_0, \qquad (3.2.1.2)$$

where θ_0 is the maximum allowable average deviation from linearity.

Unlike the evaluation rule of EP6-A which requires $\mu_{Pi} - \mu_{Li}$ be within some pre-specified allowable limit of δ_0 at all concentration levels, Hypothesis (3.2.1.2) only requires an summarized measure ADL be less than δ_0. Since ADL is a function of standardized sum of squares of the differences in the predicted means between the

best-fitted and linear models, it is an aggregate criterion.

3.2.2 Sum of Squares of Deviations from Linearity (SSDL)

According to the approved CLSI guideline EP6-A (Tholen et al., 2003), the linearity of the proposed analytical method can be concluded if the deviation from linearity is smaller than some pre-specified limit δ_0 at all concentrations:

$\left|\mu_{Pi} - \mu_{Li}\right| < \delta_0$, for i = 1,…, L.

As a result, a natural aggregate metric for assessment of assay linearity is the sum of squares of deviations from linearity (SSDL) denoted by τ defined as

$$\tau = \sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2. \qquad (3.2.2.1)$$

It follows that the hypotheses for proving the assay linearity can be formulated based on SSDL as follows:

$$H_0: \sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2 \geq L\delta_0^2 \text{ vs. } H_0: \sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2 < L\delta_0^2 \qquad (3.2.2.2)$$

or equivalently

$$H_0: \sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2/L \geq \delta_0^2 \text{ vs. } H_0: \sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2/L < \delta_0^2$$

Similar to ADL, SSDL is an aggregate criterion but formulated directly by the nature of disaggregate criterion proposed by CLSI guideline as the form of model-by-dilution interaction. However, the corresponding statistical hypothesis is not to detect existence of the model-by-dilution interaction but rather to verify whether the model-by-dilution interaction is within some pre-specified allowable upper limit.

3.2.3 Coefficient of Variation of the Deviations from Linearity (CVDL)

The CVDL is the scaled deviations scaled by σ, the variability or repeatability of the best-fitted model for assessment of linearity defined as the square root of the average sum of squares of the scaled deviations by σ:

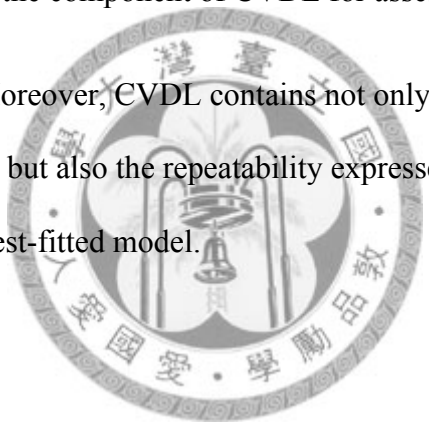
$$\eta = \text{CVDL} = \frac{\sqrt{\sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2/L}}{\sigma} = \frac{\sqrt{\sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2/L}}{\sqrt{\sigma^2}}. \quad (3.2.3.1)$$

The hypotheses for evaluation of linearity is given for CVDL as:

$H_0: \eta \geq \eta_0$ vs. $H_a: \eta < \eta_0$. (3.2.3.2)

where η_0 is the allowable limit of CVDL.

As $\sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2$ is also the component of CVDL for assessment of linearity, CVDL is an aggregate criterion. Moreover, CVDL contains not only the information of the deviation from the linearity but also the repeatability expressed by the residual mean square obtained from the best-fitted model.



3.3 Summary

As we introduced, the disaggregate criterion proposed by CLSI EP6-A guideline (Tholen et al., 2003) requires $\mu_{Pi} - \mu_{Li}$ to be within some pre-specified allowable limit at all concentration levels, while the aggregate criteria of ADL, SSDL and CVDL only require a summary measure of $\sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2$ either scaled by σ, μ or to be within their corresponding allowable limit. As a result, the evaluation based a disaggregate criterion is more conservative than an aggregate criterion since it requires an intersection-union test. In addition, $\sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2$ is actually the model-by-dilution

interaction. However, unlike the traditional hypothesis to test the existence of the interaction, our goal is to test if the magnitude of the interaction is within the allowable bound. In the next few chapters, the statistical testing procedures will be proposed for assessing linearity based on the disaggregate criterion and aggregate criterion introduced in this chapter. The comparison of the proposed methods and current methods will also be performed via the simulation studies and numerical examples.

Chapter 4

TOST Procedure and Corrected Kroll's Method

In this chapter, we propose the an one-sided tests procedure (TOST) and the corrected Kroll's method which are more suitable methods for assessment of linearity by improving the shortcomings of the current methods. The proposed TOST procedure is the method correspondeding to the estimation method of EP6-A (Tholen et al., 2003) which ignore the variability of the estimators, while the corrected Kroll's method is used to correct the inappropriate statistical hypothesis of the uncorrected Kroll method proposed by Kroll et al. (Kroll, 2000).

4.1 Two One-sided Test Procedure

With respect to the interval hypothesis in (2.2.1), it can also be decomposed into two sets of one-sided hypotheses as,

$$H_{0iL}: \mu_{Pi} - \mu_{Li} \leq -\delta_0 \text{ vs. } H_{aiL}: \mu_{Pi} - \mu_{Li} > -\delta_0, \text{ for all } i=1,\ldots, L,$$

and (4.1.1)

$$H_{0iU}: \mu_{Pi} - \mu_{Li} > \delta_0 \text{ vs. } H_{aiU}: \mu_{Pi} - \mu_{Li} < \delta_0, \text{ for all } i=1,\ldots, L.$$

An unbiased estimator of $\mu_{Pi} - \mu_{Li}$ is the LS estimator $\hat{Y}_{Pi} - \hat{Y}_{Li}$, i=1,..,L. Define

$$\mathbf{X}_L = (\mathbf{1}, \mathbf{X}),$$

$$\mathbf{X}_P = \begin{cases} (\mathbf{1}, \mathbf{X}, \mathbf{X}_2), & \text{if the best-fitted model is quadratic, and} \\ (\mathbf{1}, \mathbf{X}, \mathbf{X}_2, \mathbf{X}_3), & \text{if the best-fitted model is cubic,} \end{cases}$$

where **1** is LJx1 vector of 1s, $\mathbf{X} = (X_i)$, $\mathbf{X}_2 = (X_i^2)$, and $\mathbf{X}_3 = (X_i^3)$, and J is the number of replicates.

An unbiased estimator of the variance of $\hat{Y}_{Pi} - \hat{Y}_{Li}$ is given as

$$\hat{\sigma}_{di}^2 = w_{ii}\hat{\sigma}_e^2,$$

where w_{ii} is the ith diagonal element of matrix $\mathbf{WW}'$, $\hat{\sigma}_e^2$ is the residual mean square obtained from the best-fitted model with degrees of freedom LJ-d-1, $\mathbf{W} = \mathbf{W}_P - \mathbf{W}_L$, $\mathbf{W}_P$ and $\mathbf{W}_L$ are the projection matrices corresponding to the column spaces spanned by the design matrices of the best–fitted and linear models, respectively, i.e., $\mathbf{W}_P = \mathbf{X}_P(\mathbf{X}_P'\mathbf{X}_P)^{-1}\mathbf{X}_P'$ and $\mathbf{W}_L = \mathbf{X}_L(\mathbf{X}_L'\mathbf{X}_L)^{-1}\mathbf{X}_L'$.

It follows that the 100(1 - 2α)% confidence interval for $\mu_{Pi} - \mu_{Li}$ is given as

$$(\hat{Y}_{Pi} - \hat{Y}_{Li}) \pm t_{\alpha, LJ-d-1}\hat{\sigma}_{di}, \quad i=1,\ldots,L, \tag{4.1.2}$$

where $t_{\alpha,LJ-d-1}$ is the upper α percentile of a central distribution with degree of freedom of LJ-d-1.

The linearity of an analytical method can be concluded at the α significance level if the 100(1-2α)% confidence interval for $\mu_{Pi} - \mu_{Li}$ is completely contained within the pre-specified allowable limit of δ_0 at each concentration level, I=1,...,L. This method is referred to as the two one-sided tests (TOST) procedure which the statistical testing procedure is proposed instead f the estimation method of EP6-A.

4.2 Corrected Kroll's Method

The main drawback of the method for evaluation of linearity proposed by Kroll, et al. (Kroll, 2000) is the incorrect formulation of the hypotheses. We suggest the hypothesis for assessment of linearity based on ADL should be formulated as follows:

$H_0: \theta \geq \theta_0$ vs. $H_a: \theta < \theta_0$. (4.2.1)

where θ_0 is the allowable margin of ADL for linearity.

Consequently, the linearity of an analytical procedure is concluded at the 5% significance level if

$$\hat{\theta} < \frac{\sigma}{\mu}\sqrt{\frac{q_{0.05}}{LJ}}, \quad (4.2.2)$$

where $q_{0.05}$ is the 5th percentile of a non-central chi-square distribution with degrees of freedom d-1 and non-centrality parameter $LJ\theta_0^2/(\sigma/\overline{X})^2$. This method is referred to as the corrected Kroll method.

4.3 Simulation Study

We conduct a simulation study to compare the empirical sizes and powers of the uncorrected Kroll method, the corrected Kroll method, the estimation method of EP6-A and two one-sided tests procedures. Following the specification of the experiment designs for evaluation of linearity, the number of solutions (or dilutions) of different concentrations is set to be 5 or 7 and the number of replications at each concentration is 2, 3, or 4.

Throughout the simulation, the allowable margin of linearity based on ADL, θ_0, is specified as 0.05 while the margin for the estimation and TOST procedures, δ_0, is

specified as 0.2. There are two types of comparison of size. The first type is to compare the size between the uncorrected Kroll with corrected Kroll methods for which the data were generated at the value of 0.05 for ADL as recommended by Kroll, et al. (Kroll, 2000). The second type is to compare the size between the estimation method suggested in the approved CLSI guideline and the TOST procedure for which the data were generated with the true difference, $\mu_{Pi} - \mu_{Ci}$ at some solutions being either 0.2 or -0.2. In addition, standard deviation of normal random error was specified as 0.1 and 0.2. Table 4.3.1 provides the specifications of the values of parameters in the simulation for evaluation of size. For each of 12 combinations, five thousand (5,000) random samples are generated. For the 5% nominal significance level, a simulation study with 5,000 random samples implies that 95 percent of the empirical sizes evaluated at the equivalence limits will be within 0.04396 and 0.05604 if the proposed methods can adequately control the size at the nominal level of 0.05. In addition, the specifications of parameters for investigation of power are given Table 4.3.2.

Table 4.3.3 presents the results of the empirical sizes. For the comparison between the uncorrected Kroll and the correct Kroll methods, all empirical sizes of the uncorrected Kroll method are above 0.92. On the other hand, the empirical size of the corrected Kroll method ranges from 0.0516 to 0.0780. Only 8.33% (1/12) of the empirical sizes of the corrected Kroll method are within 0.04395 and 0.05604. The reason for the extremely high empirical size of the uncorrected Kroll method is from incorrect formulation of hypothesis for proving the linearity of the analytical methods. The type I error with respect to proving the linearity is the error that the analytical method is claimed to be linear but in fact it is not. Therefore, the empirical size of the uncorrected Kroll method at the 5% nominal level should be close to 95%. On the contrary, the

empirical size of the corrected Kroll method should be close to 5% at the 5% nominal level. However, one needs to estimate the non-central parameters for non-central χ^2 distribution of the observed ADL. In addition, the critical value in Eq. (4.2.2) also contains an estimator $\overline{X}$. Therefore, both the uncorrected and corrected Kroll methods ignore the variability of the estimators in the non-central parameters and critical value. As a result, although the empirical size of the corrected Kroll method is close to 0.05, it is still inflated. The empirical sizes of the estimation method and TOST procedure for the same specifications are also provided in Table 4.3.3. From Table 4.3.1, when the true ADL is 0.05, $\mu_{Pi} - \mu_{Ci}$ at some solutions is either greater than 0.2 or smaller -0.2. It follows that all empirical sizes of the TOST procedure are less than 0.02. However, on the contrary, the empirical size of estimation method suggested in the approved CLSI guideline EP6-A can reach as high as 0.30 even when the differences in means between the best fitted curve and the linear regression equation are outside the margin of (-0.2, 0.2) at three of the five solutions.

For the comparison between the estimation method in the approved CLSI guideline EP6-A and TOST procedure, the empirical sizes of the TOST procedure ranges from 0.0440 to 0.0564. Only 8.33% of the empirical sizes (1/12) are not included in (0.04395, 0.05604). The one outside (0.04396, 0.05604) has the empirical size of 0.0564, which is just 0.0036 above 0.05604. However, the range of the empirical sizes of the estimation method is from 0.4930 to 0.5066. Recall that the estimation method suggested in the approved CLSI guideline EP6-A (Tholen et al., 2003) ignores the variation of the estimates of $\mu_{Pi} - \mu_{Ci}$. When $\mu_{Pi} - \mu_{Ci}$ is equal to either 0.2 or -0.2 at some solutions, and then under the normal assumption, the size should be equal to 0.5 as confirmed by the empirical sizes of the simulation.

Table 4.3.1 Specifications of parameters for size (Uncorrected Kroll vs. Corrected Kroll and Estimation Method vs. TOST)

Type of Comparisons	No. of Solution Levels	True ADL	Solution Level	True $\mu_{Pi} - \mu_{Li}$
Uncorrected Kroll vs. Corrected Kroll				
	5	0.05	1	-0.23905
			2	0.11952
			3	0.23905
			4	0.11952
			5	-0.23905
	7	0.05	1	-0.28868
			2	0.00000
			3	0.17321
			4	0.23094
			5	0.17321
			6	-0.00000
			7	-0.28868
Estimation Method vs. TOST				
	5	0.01656	1	-0.20000
			2	0.10000
			3	0.20000
			4	0.10000
			5	-0.20000
	7	0.01017	1	-0.20000
			2	0.00000
			3	0.12000
			4	0.16000
			5	0.12000
			6	-0.00000
			7	-0.20000

Table 4.3.2 Specifications of parameters for power (Uncorrected Kroll vs. Corrected Kroll and Estimation Method vs. TOST)

No. of Solution Levels	True ADL	Solution Level	True $\mu_{Pi} - \mu_{Li}$
5	0.00151	1	-0.02
		2	0.01
		3	0.02
		4	0.01
		5	-0.02
7	0.00494	1	-0.10
		2	0.00
		3	0.06
		4	0.08
		5	0.06
		6	0.00
		7	-0.10

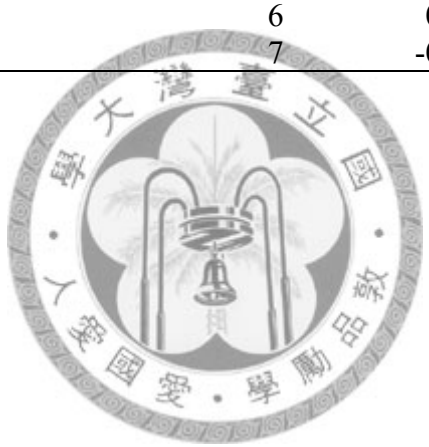


Table 4.3.4 presents the results of the empirical powers. For the simulation, the true ADL is specified as 0.00151 or 0.00494 when the number of solutions is 5 or 7, respectively. Therefore, with an allowable margin of 5%, the 91.67% of the empirical powers of the uncorrected and corrected Kroll methods reach 1. On the other hand, the empirical powers of the estimation method and TOST procedures are smaller than those of the uncorrected and corrected Kroll methods. In addition, the empirical powers of the estimation method suggested in the approved CLSI guideline EP6-A (Tholen et al., 2003) and TOST procedures increase as the number of replicates increases or the standard deviation decreases. The results in Table 4.3.4 show that the empirical power of the estimation method is greater than that of TOST procedure. However, from Table 4.3.3, the uncorrected and corrected Kroll methods, and the estimation procedure fail to control the size at the nominal level. Therefore, the advantage of power by these methods comes at the expense of inflation of type I error rate. From the results of the simulation in Table 4.3.4, the power of the TOST procedure is greater than 0.9 when the standard deviation is 0.1 and number of replicates is at least 3.

4.4 Numerical Example

We consider a hypothetical experiment for evaluation of the linearity of a new analytical procedure for determination of β-HCG (β-Human Chorionic Gonadotropic, mIU/mL). The design consists of 5 dilutions with two replicates at each dilution of concentrations. Table 4.4.1 presents a set of hypothetic measurements under the design described above. For the purpose $|\mu_{Pi} - \mu_{Li}|$ of the illustration, the allowable margin of percent bound for ADL is set as 0.05 for uncorrected and corrected Kroll's methods. On the other hand, the allowable limit of is set as 0.4 for the estimation method

Table 4.3.3 Results of empirical sizes (Uncorrected Kroll vs. Corrected Kroll and Estimation Method vs. TOST)

Type of Comparisons	No. of Sol.	No. of Rep.	SD	Kroll's Method		EP6-A	TOST
				Uncorr.	Corr.		
Kroll – uncorr. Vs. corr.	5	2	0.1	0.9232	0.0662	0.1590	0.0056
			0.2	0.9292	0.0780	0.3030	0.0166
		3	0.1	0.9328	0.0636	0.1050	0.0010
			0.2	0.9332	0.0646	0.2632	0.0114
		4	0.1	0.9420	0.0624	0.0768	0.0006
			0.2	0.9328	0.0608	0.2296	0.0074
	7	2	0.1	0.9328	0.0622	0.0110	0.0000
			0.2	0.9332	0.0606	0.1328	0.0028
		3	0.1	0.9418	0.0604	0.0024	0.0000
			0.2	0.9458	0.0576	0.0754	0.0012
		4	0.1	0.9438	0.0516	0.0002	0.0000
			0.2	0.9424	0.0586	0.0544	0.0008
Estimation vs. TOST.	5	2	0.1	1.0000	1.0000	0.4984	0.0522
			0.2	0.9998	1.0000	0.5050	0.0564
		3	0.1	1.0000	1.0000	0.4930	0.0440
			0.2	1.0000	1.0000	0.5050	0.0486
		4	0.1	1.0000	1.0000	0.5048	0.0490
			0.2	1.0000	1.0000	0.5066	0.0560
	7	2	0.1	0.9998	0.9998	0.4972	0.0512
			0.2	1.0000	1.0000	0.5024	0.0484
		3	0.1	0.9998	0.9998	0.4978	0.0478
			0.2	1.0000	1.0000	0.4946	0.0504
		4	0.1	1.0000	1.0000	0.5044	0.0504
			0.2	1.0000	1.0000	0.5066	0.0494

Sol.: Solution; Rep.: Replications; Uncorr.: Uncorrected; Corr.: Correction.

Table 4.3.4 Results of empirical powers (Uncorrected Kroll vs. Corrected Kroll and Estimation Method vs. TOST)

No. of Sol.	No. of Rep.	SD	Kroll's Method Uncorr.	Corr.	EP6-A	TOST
5	2	0.1	1.0000	1.0000	0.9954	0.7616
		0.2	1.0000	1.0000	0.9052	0.2976
	3	0.1	1.0000	1.0000	0.9998	0.9232
		0.2	1.0000	1.0000	0.9470	0.4452
	4	0.1	1.0000	1.0000	0.9998	0.9754
		0.2	1.0000	1.0000	0.9664	0.5518
7	2	0.1	1.0000	1.0000	0.9954	0.7754
		0.2	1.0000	1.0000	0.9014	0.3168
	3	0.1	0.9998	0.9998	0.9994	0.9164
		0.2	1.0000	1.0000	0.9450	0.4468
	4	0.1	1.0000	1.0000	1.0000	0.9704
		0.2	1.0000	1.0000	0.9660	0.5570

Sol.: Solution; Rep.: Replications; Uncorr. Uncorrected: Corr.: Correction.

suggested in the approved CLSI guideline EP6-A, and for the TOST procedure.

Table4.4.2 provides the results of regression analyses for the linear, quadratic and cubic linear regression models. The results of the regression analyses presented in Table 4.4.2 demonstrates that all estimates of the regression coefficients of the cubic model are significantly different from 0 at the 5% level. ($t_{0.025,\ 6}$ = 2.4469) In addition, the standard error of the residuals from the estimated cubic regression equation is 0.1799 that is at least 40% smaller than those from the linear or the quadratic models. Furthermore, the coefficient of determination, R^2, is also above 0.99. As a result, the cubic model is the best-fitted model among the three models recommended by the approved CLSI guideline EP6-A. Figure 4.4.1 presents the fitted the cubic, linear regression equations and the means at each of the five dilution. It clearly shows that the relationship between the dilutions of concentrations and the analytical results is nonlinear and the cubic model is a better fit than the simple linear regression model.

Table 4.4.3 gives the predicted means from the cubic and linear regression models at each of the five dilutions as well as their corresponding differences, while Table 4.4.4 present summarized results of linearity by the four methods. From these differences and observed mean concentration, the observed ADL yields a value of 0.0842. With respect to the hypothesis in Eq. (2.3.2) and a margin of percent bound of 5%, the critical value in Eq. (2.3.3) is 0.0851 which is greater than the observed ADL of 0.0842, According to the decision rule of the uncorrected Kroll method, the analytical method can be concluded linear at the 5% significance level. However, it should be noted that for this example, even though the observed ADL of 0.0842 is already greater than the allowable percent bound of 0.05, the linearity of the analytical method still can be claimed by the uncorrected Kroll method. On the other hand, with respect to hypothesis in Eq. (4.2.1) for the corrected Kroll method, the critical value with an allowable margin of 5% in Eq.

Table 4.4.1 Measurement of β-HCG (mIU/mL)

Dilution	Replicate1	Replicate 2
1	1.00	0.99
2	1.60	1.59
3	2.50	2.60
4	4.36	4.39
5	5.10	5.00

Table 4.4.2 Summary of results of regression analyses of β-HCG

Order	Coefficient	Value	SE	t-test	Std err $S_{y.x}$	Degrees freedom
Linear	α'	-0.354	0.234	-1.51		
	β_1'	1.089	0.071	**15.44**	0.3154	8
Quadratic	α''	0.156	0.461	0.34		
	β_1''	0.652	0.351	1.85		
	β_2''	0.073	0.058	1.27	0.3041	7
Cubic	α'''	2.263	0.626	**3.62**		
	β_1'''	-2.308	0.818	**-2.82**		
	β_2'''	1.202	0.304	**3.96**		
	β_3'''	-0.125	0.034	**-3.74**	0.1799	6

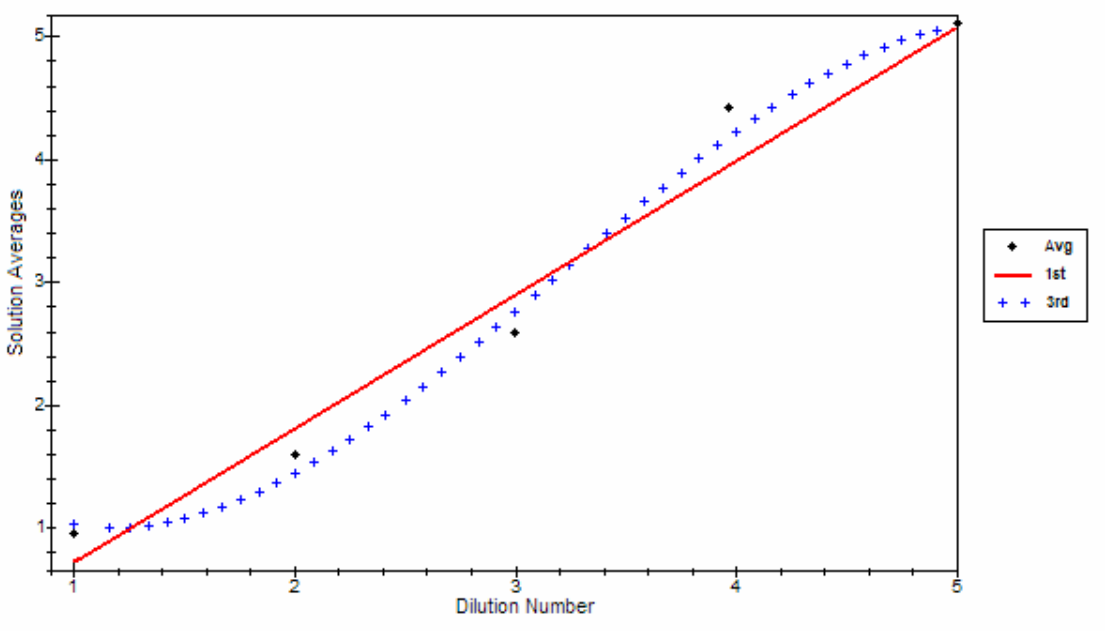


Figure 4.4.1 Regression curves for cubic versus linear models of β-HCG

(4.2.2) is 0.0237. Since the observed ADL of 0.0842 is greater than 0.0237, we cannot reject the null hypothesis and cannot concluded the linearity of the analytical method at the 5% significance level. Unlike the uncorrected Kroll method, the conclusion of the corrected Kroll method is consistent with the evidence for which the observed ADL is 0.0842, which is greater than the allowable percent bound of 0.05.

With respect to the estimation method suggested in the approved CLSI guideline EP6-A, the observed differences in the predicted means between the cubic and linear regression models at all dilutions are within the allowable margin of ±0.4. As a result, the linearity is claimed by the estimation method. On the other hand, the results of the TOST procedure show that the 95% confidence intervals for μ_{Pi} - μ_{Ci} at the first two dilutions are not contained within (-0.4, 0.4). With respect to hypotheses in Eq. (4.1.1), the analytical method cannot be concluded linear at the 5% significance level. Because the estimation method completely ignores the variability in the observed differences in the predicted means, its conclusion is made without any statement of the probability of type I error. However, in fact, as demonstrated by the simulation, the probability of type I error of the estimation method far exceeds its nominal significance level.

4.5 Summary

With respect to the disaggregate criterion, the estimation method suggested by the approved CLSI guideline ignores the variation of the estimates of the differences in the predicted means and is not a formal statistical inference procedure. On the other hand, the procedure based on the aggregate criterion of ADL proposed by Kroll et al. (Kroll, 2000) incorrectly formulated the hypothesis for proving linearity as the null hypothesis. As a result, the uncorrected Kroll method cannot control the type I error in

Table 4.4.3 Mean differences between the best-fitted curve and simple linear regression equation of β-HCG

Result Mean	Predicted (Linear)	Predicted (Cubic)	Difference	% Difference
0.995	0.735	1.031	0.296	28.7
1.595	1.824	1.450	-0.374	25.8
2.550	2.913	2.767	-0.146	5.3
4.375	4.002	4.230	0.228	5.4
5.050	5.091	5.086	-0.005	0.1

Table 4.4.4 Results of the linearity by four different methods of β-HCG

Kroll						TOST			Estimation Method		
	Un-corrected		Corrected			90% C.I. of	Result			Result	
Sample ADL	Critical Value	Result	Critical Value	Result	Dil. (i)	$\mu_{Pi} - \mu_{Li}$	Ind.	Ove-rall	$\hat{Y}_{Pi} - \hat{Y}_{Li}$	Ind.	Ove-rall
0.0842	0.0851	L	**0.0237**	**NL**	1	**(0.064,0.529)**	**NL**	**NL**	0.296	L	L
					2	**(-0.524,-0.223)**	**NL**		-0.374	L	
					3	(-0.318, 0.027)	L		-0.146	L	
					4	(0.078,0.379)	L		0.228	L	
					5	(-0.237, 0.228)	L		-0.005	L	

Dil. : Dilution level
Ind. : Individual dilutions
Overall : Overall conclusion
NL: Conclusion of nonlinearity at the 5% nominal level
L: Conclusion of linearity at the 5% nominal level

decision-making of conclusion for linearity. Therefore, we proposed the TOST procedure for the disaggregate criterion and the corrected Kroll method for the aggregate criterion based on ADL by formulating the hypothesis for proving linearity as the alternative hypothesis. Simulation results and the numerical example described above demonstrate that the proposed TOST and the correct Kroll method not only can adequately control the type I error rate but also reach the conclusion consistent with the data.

Since TOST procedure is constructed based on a disaggregate criterion which requires all differences in the predicted means between the best-fitting and linear models be within the pre-specified allowable limit, the method is more conservative than the corrected Kroll's method which is based on an aggregate criterion and only requires ADL, a function of standardized sum of squares of the differences in the predicted means between the best-fitted and linear models to be controlled within the pre-specified allowable percent bound, However, as mentioned before, the inference based on ADL involves the estimation of the unknown non-centrality parameter and the average population mean concentration. When these estimates are assumed fixed constants for the inference based on ADL, the simulation study shows that the empirical size can be inflated up to 0.078 at the 0.05 significance level. In the next chapter, we will propose GPQ-based ADL statistical testing procedure to overcome the issue of the unknown parameter of the distribution of ADL.

Chapter 5

General Pivotal Quantity Approach of ADL

In Chapter 4, we introduced the corrected Kroll's method which reformulates the inappropriate statistical hypothesis of the uncorrected Kroll's method. However, as we observed in the simulation results of the proposed corrected Kroll's method, the type I error still inflates up due to variability in estimation of unknown non-centrality parameter of the chi-square distribution. To solve this issue, in this chapter we propose an alternative statistical testing procedure based on ADL by applying the generalized pivotal quantity approach introduced by Tsui and Weerahandi (Weerahandi, 1993).

5.1 General Pivotal Quantity (GPQ)

Weerahandi (Tsui and Weerahandi, 1989) used a generalized p-value for comparing parameters of two regressions with unequal variances. Motivated by that application, Tsui and Weerahandi (Tsui and Weerahandi, 1989) gave the explicit definition of generalized p-values, and showed that it is an exact probability of a extreme region. Their proposed method has been successfully used to provide small sample solution for many hypothesis testing problems when nuisance parameters are present and frequentist testing procedures are difficult to obtain, even nonexistent. Furthermore, Weerahandi

(Weerahandi, 1993) extended the concept of generalized p-values, and presented the generalized confidence interval (GCI) to construct an exact interval estimation.

Suppose that V is a random variable whose distribution depends on a vector of unknown parameters $\zeta=(\theta, \boldsymbol{\eta})$, where θ is a parameter of interest and $\boldsymbol{\eta}$ is a vector of nuisance parameter. Let **V** be a random sample from V and **v** be the observed value of **V**. Also let $\boldsymbol{R}=\boldsymbol{R}(\mathbf{V}; \mathbf{v}, \zeta)$ be a function of **V**, **v** and ζ. The random quantity $\boldsymbol{R}$ is said to be a GPQ if satisfies the following two conditions:

(a) The distribution of $\boldsymbol{R}$ does not depend on any unknown parameters.

(b) The observed value of $\boldsymbol{R}$, say $r= \boldsymbol{R}(\mathbf{v}; \mathbf{v}, \zeta)$, is free of the vector of nuisance parameters $\boldsymbol{\eta}$. In other words, the value of R at **V** = **v** should be a function only of (**v**, θ).

Specifically, if the observed quantity $r = \theta$, then the GPQ is called the fiducial generalized pivotal quantity (FGPQ) and generalized confidence interval (GCI) based on FGPQ are proven to have asymptotically correct frequent coverage probability in Hanning et al (Hanning et al., 2006). In consequence, an upper 100(1-α)th percentile GCI for θ is given by $R_{1-\alpha}$, where $R_{1-\alpha}$ are the 100(1-α)th percentile of the distribution of $\boldsymbol{R}$. The percentile of $\boldsymbol{R}$ can be estimated using Monte-Carlo algorithms.

5.2 Generalized Pivotal Quantity of ADL

Following the regression models in Eq. (2.1.1), we adopt the same expression of its matrix form in Chapter 4 as follows:

$\mathbf{Y}$ = the LJx1 vector of observations,

$\mathbf{X_L} = (\mathbf{1}, \mathbf{X})$,

$$\mathbf{X_P} = \begin{cases} (\mathbf{1}, \mathbf{X}, \mathbf{X_2}), \text{if the best-fitted model is quadratic, and} \\ (\mathbf{1}, \mathbf{X}, \mathbf{X_2}, \mathbf{X_3}), \text{if the best-fitted model is cubic,} \end{cases}$$

$\boldsymbol{\mu}_\mathbf{P}$ = the LJx1 predicted mean vector of best-fitted polynomial model, and

$\boldsymbol{\mu}_\mathbf{L}$ = the LJx1 predicted mean vector of linear model,

where $\mathbf{1}$ is LJx1 vector of 1s, $\mathbf{X}$ =(X_i), $\mathbf{X}_2$=(X_i^2), and $\mathbf{X}_3$=(X_i^3). L and J are the number of concentrations and number of replicates, respectively.

We have $\hat{\mathbf{Y}}_\mathbf{P} = \mathbf{W}_\mathbf{P}\mathbf{Y}$ and $\hat{\mathbf{Y}}_\mathbf{L} = \mathbf{W}_\mathbf{L}\mathbf{Y}$ as the LS estimators of the predicted mean vectors of the best-fit and linear models, where $\mathbf{W}_\mathrm{P} = \mathbf{X}_\mathrm{P}(\mathbf{X}_\mathrm{P}'\mathbf{X}_\mathrm{P})^{-1}\mathbf{X}_\mathrm{P}'$ and $\mathbf{W}_\mathrm{L} = \mathbf{X}_\mathrm{L}(\mathbf{X}_\mathrm{L}'\mathbf{X}_\mathrm{L})^{-1}\mathbf{X}_\mathrm{L}'$ and $\mathbf{W} = (\mathbf{W}_\mathrm{P} - \mathbf{W}_\mathrm{L})$. As a result, the unbiased and sufficient estimator of $\boldsymbol{\mu}_\mathbf{P} - \boldsymbol{\mu}_\mathbf{L}$ and its covariance matrix, $\boldsymbol{\Sigma}$, are given as respectively:

$$\begin{aligned} \hat{\boldsymbol{\mu}}_\mathbf{P} - \hat{\boldsymbol{\mu}}_\mathbf{L} &= \hat{\mathbf{Y}}_\mathbf{P} - \hat{\mathbf{Y}}_\mathbf{L} = \mathbf{WY} \\ \hat{\boldsymbol{\Sigma}}\mathbf{Cov}\left(\hat{\mathbf{Y}}_\mathbf{P} - \hat{\mathbf{Y}}_\mathbf{L}\right) &= S^2\mathbf{WW}' \end{aligned} \qquad (5.2.1)$$

where S^2 is the residual mean square obtained from the best-fitted polynomial model with degree of freedom LJ-d-1. Under the assumption that random errors in the above regression model are identically and independently distributed as normal distribution with mean of zero and variance of σ^2, $\hat{\mathbf{Y}}_\mathbf{P} - \hat{\mathbf{Y}}_\mathbf{L}$ is distributed as a multinormal distribution with mean $\boldsymbol{\mu}_\mathbf{P} - \boldsymbol{\mu}_\mathbf{L}$ and variance $\boldsymbol{\Sigma}$ which is equal to $\sigma^2\mathbf{WW}'$. In addition, $\overline{\mathrm{Y}}$ can be expressed as $\mathbf{1}'\mathbf{Y}/\mathrm{LJ}$ which is distributed as an univariate normal distribution with mean μ and variance σ^2/LJ.

It is easy to verify that the estimators $\mathbf{WY}$, S^2 and $\overline{\mathrm{Y}}$ are associated with pivotal quantities $\mathbf{Z}$, U and $\mathbf{Z}_\mu$ which are independent with the following distributions:

$$\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}[\mathbf{WY} - (\boldsymbol{\mu}_{\mathbf{P}} - \boldsymbol{\mu}_{\mathbf{L}})] \sim \mathbf{N}_{\mathrm{LJ}}(\mathbf{0}, \mathbf{I})$$

$$U = \frac{(\mathrm{LJ\text{-}d\text{-}1})\mathrm{S}^2}{\sigma^2} \sim \chi^2_{\mathrm{LJ\text{-}d\text{-}1}} \tag{5.2.2}$$

$$Z_{\mu} = \frac{\overline{\mathrm{Y}} - \mu}{\sqrt{\frac{\sigma^2}{\mathrm{LJ}}}} \sim N(0,1)$$

where matrix $\boldsymbol{\Lambda}^{1/2}$ denotes the positive definite square root of a positive definite matrix $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}^{-1/2} = (\boldsymbol{\Lambda}^{1/2})^{-1}$. $\mathbf{N}_{\mathrm{LJ}}(\mathbf{0}, \mathbf{I})$, $\chi^2_{\mathrm{LJ\text{-}d\text{-}1}}$ and $N(0,1)$ denote the multivariate standard normal distribution with LJ×1 random vector, the chi-square random variable with LJ-d-1 degrees and univariate standard normal distribution, respectively.

Recall that the definition of ADL denoted by θ as the following:

$$\theta = \mathrm{ADL} = \frac{\sqrt{\sum_{i=1}^{\mathrm{L}} (\mu_{\mathrm{Pi}} - \mu_{\mathrm{Li}})^2 / \mathrm{L}}}{\mu}$$

To obtain a GPQ for ADL, we can start the work by deriving GPQs for $\boldsymbol{\mu}_{\mathbf{P}} - \boldsymbol{\mu}_{\mathbf{L}}$ and μ, respectively. According to the first equation in (5.2.2), $\boldsymbol{\mu}_{\mathbf{P}} - \boldsymbol{\mu}_{\mathbf{L}}$ can be expressed as:

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{P}} - \boldsymbol{\mu}_{\mathbf{L}} &= \mathbf{WY} - \boldsymbol{\Sigma}^{1/2}\mathbf{Z} \\ &= \mathbf{WY} - \left(\sigma^2 \mathbf{WW}'\right)^{1/2} \mathbf{Z} \\ &= \mathbf{WY} - \left(\frac{(\mathrm{LJ} - \mathrm{d} - 1) S^2}{U} \mathbf{WW}'\right)^{1/2} \mathbf{Z} \end{aligned} \tag{5.2.3}$$

Let $\mathbf{y}$ and s^2 be the observed values of $\mathbf{Y}$ and S^2, respectively, a GPQ for $\boldsymbol{\mu}_{\mathbf{P}} - \boldsymbol{\mu}_{\mathbf{L}}$ is given by

$$\boldsymbol{R}_{\mu_P-\mu_L} = \mathbf{W}\mathbf{y} - \left(\frac{(\text{LJ-d-1})s^2}{U}\mathbf{W}\mathbf{W}'\right)^{1/2}\mathbf{Z} \tag{5.2.4}$$

$$= \mathbf{W}\mathbf{y} - \left(\frac{s^2\sigma^2}{S^2}\mathbf{W}\mathbf{W}'\right)^{1/2}\sum^{-1/2}\left[\mathbf{W}\mathbf{Y} - (\boldsymbol{\mu}_P - \boldsymbol{\mu}_L)\right] \tag{5.2.5}$$

From (5.2.4), $\boldsymbol{R}_{\mu_P-\mu_L}$ has distribution that is free of parameters and thus does not depend on any unknown parameters. When $\mathbf{Y}$ and S^2 are substituted by their observed values $\mathbf{y}$ and s^2 in (5.2.5), the observed value of $\boldsymbol{R}_{\mu_P-\mu_L}$ denoted by $\boldsymbol{r}_{\mu_P-\mu_L}$ is obtained as:

$$\begin{aligned}\boldsymbol{r}_{\mu_P-\mu_L} &= \mathbf{W}\mathbf{y} - \left(\frac{s^2\sigma^2}{s^2}\mathbf{W}\mathbf{W}'\right)^{1/2}\sum^{-1/2}\left[\mathbf{W}\mathbf{y} - (\boldsymbol{\mu}_P - \boldsymbol{\mu}_L)\right] \\ &= \mathbf{W}\mathbf{y} - \left[\mathbf{W}\mathbf{y} - (\boldsymbol{\mu}_P - \boldsymbol{\mu}_L)\right] \\ &= \boldsymbol{\mu}_P - \boldsymbol{\mu}_L\end{aligned}$$

which is equal to $\boldsymbol{\mu}_P - \boldsymbol{\mu}_L$ and free of the nuisance parameters. Hence, it fulfills the requirements of (a) and (b) for being a GPQ for $\boldsymbol{\mu}_P - \boldsymbol{\mu}_L$. Moreover, since $\sqrt{\sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2 / L} = \sqrt{\left(\hat{\mathbf{Y}}_P - \hat{\mathbf{Y}}_L\right)'\left(\hat{\mathbf{Y}}_P - \hat{\mathbf{Y}}_L\right)/LJ}$, a GPQ of $\sqrt{\sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2 / L}$ denoted by ξ can then be obtained as :

$$R_\xi = \sqrt{\frac{1}{LJ}\left(\boldsymbol{R}_{\mu_P-\mu_L}\right)'\left(\boldsymbol{R}_{\mu_P-\mu_L}\right)} \tag{5.2.6}$$

where $\boldsymbol{R}_{\mu_P-\mu_L}$ is defined as (5.2.4).

In addition, a GPQ for μ can be obtained as:

$$R_{\mu} = \bar{y} - \frac{1}{\sqrt{n}} \sqrt{\left(\frac{(LJ\text{-}d\text{-}1)s^2}{U} \right)} Z_{\mu} \tag{5.2.7}$$

$$= \bar{y} - \frac{1}{\sqrt{n}} \sqrt{\left(\frac{s^2 \sigma^2}{S^2} \right)} \frac{\left(\bar{Y} - \mu \right)}{\sqrt{\frac{\sigma^2}{n}}} \tag{5.2.8}$$

From (5.2.7), R_{μ} has distribution that is free of parameters. In addition, when $\bar{Y}$ and S^2 are substituted by their observed values $\bar{y}$ and s^2 in (5.2.8), then the observed value of R_{μ} denoted by $\boldsymbol{r}_{\mu}$ is obtained as:

$$R_{\mu} = \bar{y} - \frac{1}{\sqrt{n}} \sqrt{\left(\frac{s^2 \sigma^2}{s^2} \right)} \frac{\left(\bar{y} - \mu \right)}{\sqrt{\frac{\sigma^2}{n}}} = \bar{y} - \left(\bar{y} - \mu \right) = \mu$$

which is equal to μ and free of the nuisance parameters. Hence, it fulfills the requirements of (a) and (b) for being a GPQ for μ.

A GPQ for ADL can be obtained:

$$R_{\theta} = \frac{R_{\xi}}{R_{\mu}} \tag{5.2.9}$$

where R_{ξ} and R_{μ} are defined as (5.2.4) and (5.2.7), respectively.

5.3 Generalized Confidence Interval of ADL

An upper 100(1-α)th percentile GCI for ADL can be obtained from the following

Monte-Carlo algorithm:

Step 1: Choose a large simulation sample size, say K=10,000. For k equal to 1 through K, carry out the following two steps.

Step 2: Generate LRx1 standard normal random vector $\mathbf{Z}$, univariate standard normal variable Z_{μ}, and central chi-square random variables U with degree of freedom LJ-d-1.

Step 3: For the realized values of $\mathbf{Y}$ and S^2, compute $R_{\theta,k}$ defined in (5.2.9).

The required upper 100(1-α)th percentiles of the distribution of GPQ for ADL, which is also the upper 100(1-α)th generalized confidence limit for ADL, is then estimated by the 100(1-α)th sample percentiles of the collection of K=10,000 realizations $R_{\theta,1}$, $R_{\theta,2}$ ……., $R_{\theta,10000}$.

5.4 Statistical Testing Procedure

With respect to the hypothesis of H_0: $\theta \geq \theta_0$ vs. H_a: $\theta < \theta_0$ based on the ADL, the upper 100(1-α)% generalized confidence limit for ADL based on GPQ can be used to test the statistical hypothesis for assessment of linearity. The null hypothesis is rejected and the linearity of a analytical method is concluded at the α significance level if the upper 100(1-α)% generalized confidence limit for ADL is less than θ_0.

5.5 Simulation Study

A simulation study is performed to compare the empirical sizes and powers of the corrected Kroll's and GPQ-based ADL methods. The specifications of the simulation study are given as follows: The number of solutions (or dilutions) of different

concentrations is set to be 5 or 7 and the number of replications at each concentration is 2, 3, or 4. Throughout the simulation, mean concentration μ is assumed to be 4. If follows that the allowable margin of linearity based on ADL, θ_0, is specified as 0.05 as recommended by Kroll et al. (Kroll, 2000). For each of 12 combinations, ten thousand (10,000) random samples are generated. For the 5% nominal significance level, a simulation study with 10,000 random samples implies that 95 percent of the empirical sizes evaluated at the allowable margins will be within 0.0457 and 0.0543 if the proposed methods can adequately control the size at the nominal level of 0.05.

The results of the empirical sizes are provided in Table 4.5.1. All empirical sizes of the corrected Kroll's method are larger than 0.0543. This indicates that the corrected Kroll's method inflates the size and is quite liberal in concluding the linearity of an analytical procedure. On the other hand, all of empirical sizes of the GPQ methods based on ADL are within the range between 0.0457 and 0.0543. The simulation results reveal that the GPQ-based methods for ADL can adequately control the size at the nominal level. The reason for a better performance of the GPQ-based methods for ADL may be that the distributions of GPQs are free of their respective nuisance parameters. On the other hand, the corrected Kroll's method fails to take into account the variability in estimator of the non-centrality parameter of the non-central chi-square distribution.

The results of the empirical powers are presented in Table 5.5.2. In Table 5.5.2, the true value of ADL is assumed to be 0.005 for both number of solutions of 5 and 7. The results in Table 5.5.2 also show that the empirical power is an increasing function of the number of replicates and number of solutions. Although the empirical power of the corrected Kroll's method is larger than the GPQ-based ADL methods, its better performance on the empirical power results from inflation of the size above the nominal

Table 5.5.1 Empirical sizes (Corrected Kroll's method vs. GPQ-based ADL method)

No. of Solutions	No. of Replicates	Standard Deviation	Corrected Kroll	GPQ-based ADL
5	2	0.1	0.0702	0.0467
		0.2	0.0763	0.0517
	3	0.1	0.0623	0.0502
		0.2	0.0655	0.0517
	4	0.1	0.0594	0.0505
		0.2	0.0595	0.0508
7	2	0.1	0.0655	0.0501
		0.2	0.0635	0.0494
	3	0.1	0.0592	0.0509
		0.2	0.0583	0.0498
	4	0.1	0.0562	0.0498
		0.2	0.0571	0.0510

Table 5.5.2 Empirical powers with the true ADL=0.005 (Corrected Kroll's method vs. GPQ-based ADL method)

No. of Solutions	No. of Replicates	Standard Deviation	Corrected Kroll	GPQ-based ADL
5	2	0.1	1.0000	1.0000
		0.2	0.9670	0.9331
	3	0.1	1.0000	1.0000
		0.2	0.9965	0.9942
	4	0.1	1.0000	1.0000
		0.2	0.9996	0.9995
7	2	0.1	1.0000	1.0000
		0.2	0.9923	0.9888
	3	0.1	1.0000	1.0000
		0.2	0.9996	0.9994
	4	0.1	1.0000	1.0000
		0.2	1.0000	1.0000

level. Figure 5.5.1 and 5.5.2 present the empirical powers when σ are 0.1 and 0.2, respectively with number of solutions is 5, number of replicates is 3. The true values of ADL are ranged from 0 to 0.08. A comparison of Figure 5.5.1 and Figure 5.5.2 reveals that the power of both methods is a deceasing function of σ. In Figure 5.5.1, when the ADL = 0.05, the empirical size for the corrected Kroll's and the GPQ-based methods are 0.0623 and 0.0502 for ADL respectively. Similar findings are observed in Figure 5.5.2. Again these results show that the corrected Kroll's method inflate the size above the 0.05 level while the GPQ-based procedure can adequately control the size at the nominal level of 5%.

5.6 Numerical Example

Table 5.6.1 presents the duplicate determinations at the first five concentrations given in Example 2 of CLSI guideline EP6-A (Tholen et al., 2003) to illustrate the proposed testing procedures in evaluation of linearity of an analytical procedure. Following EP6-A (Tholen et al., 2003), the criterion of $|\mu_{Pi}-\mu_{Li}|$ for linearity is set as 0.2 mg/dL for all 5 concentrations. In this example, the allowable margin of percent bound for ADL is set as 0.05 for all methods based on ADL as suggested by Kroll, et al. (Kroll, 2000). The results of regression analyses for the linear, quadratic and cubic linear regression models are given in Table 5.6.2. From Table 5.6.2, the estimates of the regression coefficient $\beta_2^{"}$ of the quadratic model are statistically significantly different from 0 at the 5% level ($t_{0.025,\,7}$ = 2.4469) while none of them is significantly different from 0 for the cubic model. In addition, the standard error of the residuals from the estimated quadratic regression equation is 0.124 which is smaller than the 0.2 set by the

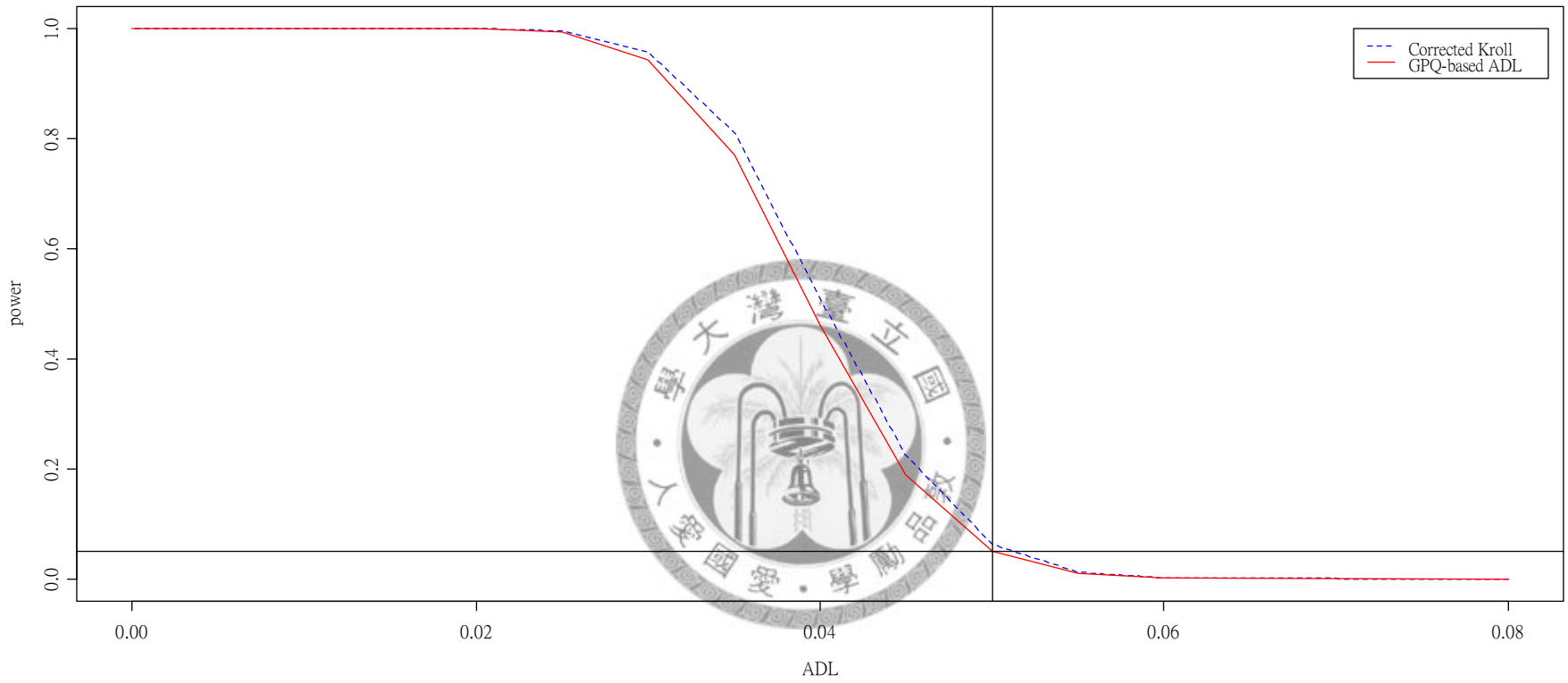


Figure 5.5.1 Empirical powers when standard deviation of normal random error is 0.1, number of solutions is 5, and number of replicates is 3 (Corrected Kroll's method vs. GPQ-based ADL method)

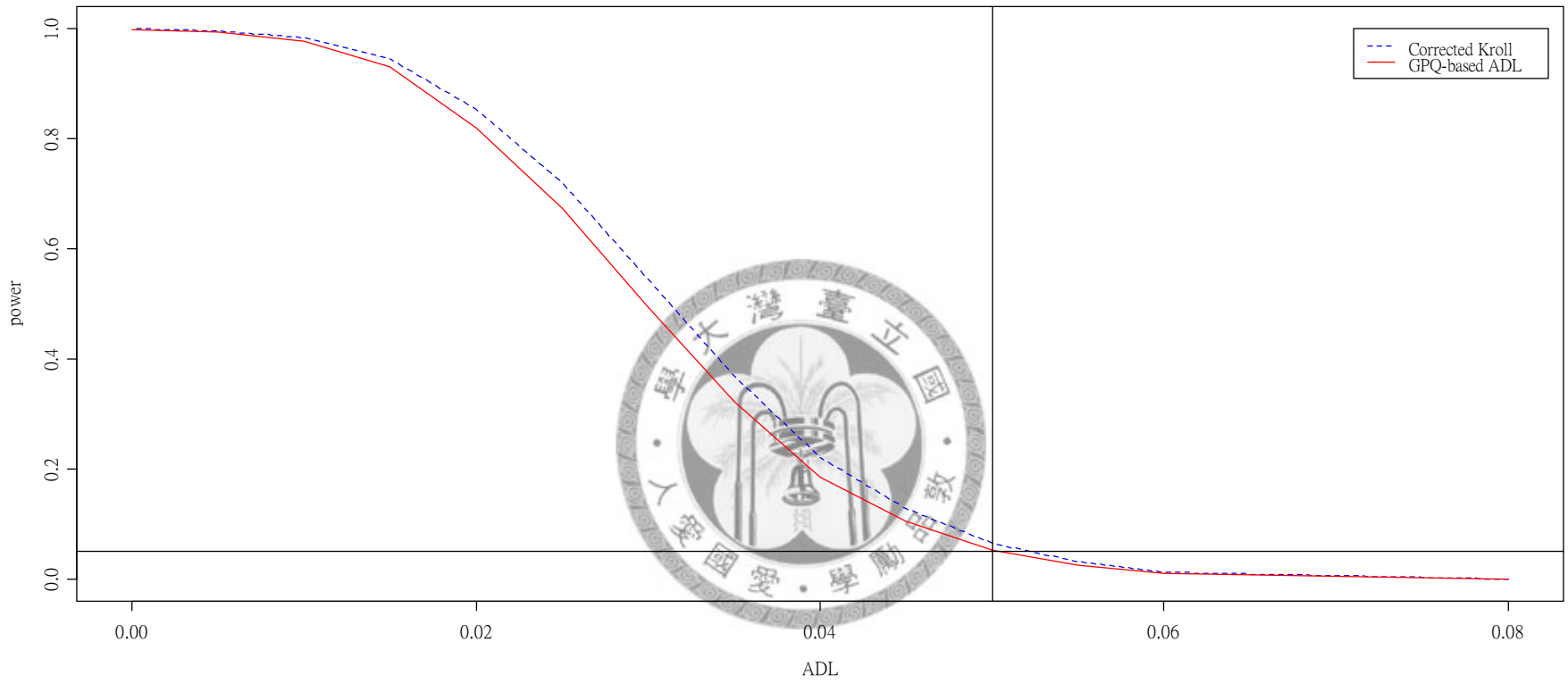


Figure 5.5.2 Empirical powers when standard deviation of normal random error is 0.2, number of solutions is 5, and number of replicates is 3 (Corrected Kroll's method vs. GPQ-based ADL method)

manufacturer. Furthermore, R^2 is also above 0.99. As a result, the quadratic model is the best-fitted model among the three models recommended by the approved CLSI guideline EP6-A (Tholen et al., 2003).

The observed predicted means from the quadratic and linear regression models at each of the five dilutions as well as their corresponding differences are given in Table 5.6.3. The results of the corrected Kroll's and the GPQ-based ADL methods are provided in Table 5.6.4. From the differences in the observed predicted means between the quadratic and linear regression models and the observed mean concentrations, the observed ADL yields a value of 0.0146. With respect to a margin of percent bound of 5%, the critical value is 0.0437 which is greater than the observed ADL of 0.0146, According to the decision rule of the corrected Kroll method, the analytical method can be concluded linear at the 5% significance level. The 95% upper confidence limit for the ADL computed by the GPQ-based ADL method is 0.0218 which is smaller than the allowable upper limit of 0.05. Hence, the linearity of the analytical procedure can be concluded at the 5% significance level by the GPQ-based ADL procedure.

5.7 Summary

The ADL proposed by Kroll et al. (Kroll, 2000) is an aggregate criterion constructed from the deviations from linearity scaled by the mean concentrations. However, the sampling distribution of the observed ADL involves unknown nuisance parameters μ and σ. On the other hand, the observed values of GPQs are free of the nuisance parameters. As a result, we apply the GPQ method to the inference of evaluation of linearity based ADL. The simulation results presented above show that the corrected Kroll's method inflates the type I error rate and the GPQ-based ADL method can control

Table 5.6.1 Measurement of calcium (mg/dL)

Dilution	Replicate 1	Replicate 2
1	4.7	4.6
2	7.8	7.6
3	10.4	10.2
4	13.0	13.1
5	15.5	15.3

Source : The approved CLSI guideline EP6-A (2003)

Table 5.6.2 Summary of results of regression analyses for the example of calcium

Order	Coefficient	LS Estimates	SE	t-test	SE $S_{y.x}$	Degrees freedom
Linear	α'	2.16	0.15	14.3		
	β_1'	2.68	0.05	59.0	0.204	8
Quadratic	α''	1.54	0.19	8.2		
	β_1''	3.22	0.14	22.4		
	β_2''	-0.09	0.02	**-3.8**	0.124	7
Cubic	α'''	1.47	0.47	3.15		
	β_1'''	3.32	0.61	5.45		
	β_2'''	-0.13	0.23	-0.56		
	β_3'''	0.004	0.02	0.17	0.134	6

Source : The approved CLSI guideline EP6-A (2003)

the size at the nominal level. On the other hand, the GPQ-based ADL procedure not only adequately control the type I error rate but also has the similar performance of the power as the corrected Kroll's method. Therefore, we conclude the GPQ-based ADL procedure is better than the correct Kroll's method for evaluating the linearity in assay validation.

Table 5.6.3 Mean differences between the best-fitted curve and simple linear regression equation for the example of calcium

Result Mean	Predicted (Linear)	Predicted (Quadratic)	Difference	% Difference
4.65	4.85	4.67	-0.18	-3.9
7.70	7.54	7.62	0.08	1.0
10.30	10.22	10.40	0.18	1.8
13.05	12.90	12.99	0.09	0.7
15.40	15.59	15.41	-0.18	-1.2

Source : The approved CLSI guideline EP6-A (2003)

Table 5.6.4 Results of the linearity evaluation for the example of calcium by corrected Kroll's and GPQ-based ADL methods

Method	Sample Statistic / Critical Value or Allowable Bound		Conclusion
Corrected Kroll	Sample ADL	0.0146	
	Critical Value	0.0437	**Linear**
GPQ-based ADL	Upper 95% C.L.	0.0218	
	Allowable Upper Bound	0.05	**Linear**

95% C.L. : Upper 95% Confidence limit

Chapter 6

Alternative Aggregate Criterion - Sum of Square of the Deviation from Linearity (SSDL)

In this chapter, we propose a new measure of the assessment of linearity named Sum of Square of the Deviation from Linearity (SSDL). As mentioned in Section 3.2.1 of Chapter 3, SSDL is formulated directly by the nature of disaggregate criterion proposed by CLSI guideline as the form of model-by-dilution interaction. However, its corresponding statistical hypothesis and testing procedure is not to detect existence of the model-by-dilution interaction but rather to verify whether the model-by-dilution interaction is within some pre-specified allowable upper limit.

6.1 SSDL and Statistical Hypothesis

Recall our introduction for SSDL in Chapter 3, a natural aggregate metric for assessment of assay linearity is the sum of squares of deviations from linearity (SSDL) denoted by τ defined as

$$\tau = \sum_{i=1}^{L} (\mu_{Pi} - \mu_{Li})^2. \qquad (6.1.1)$$

The hypotheses for proving the assay linearity in then formulated as follows:

$$H_0: \sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2 \geq L\delta_0^2 \text{ vs. } H_0: \sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2 < L\delta_0^2 \qquad (6.1.2)$$

or equivalently

$$H_0: \sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2/L \geq \delta_0^2 \text{ vs. } H_0: \sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2/L < \delta_0^2$$

where δ_0 is the allowable limit of μ_{Pi}-μ_{Li} for evaluation procedure suggested by EP6-A guideline (Tholen et al., 2003). The generalized pivotal quantity approach to hypothesis (6.1.2) of evaluation for linearity of assay validation based on the SSDL is provided in the following subsequent subsections.

6.2. Generalized Pivotal Quantity of SSDL

Most of work for deriving a GPQ of SSDL actually has been carried out in Chapter 5 for deriving a GPQ of ADL. According to the definition of SSDL in Eq. (6.1.1), a GPQ of SSDL can be obtained as:

$$R_\tau = \frac{1}{J}\left(\boldsymbol{R}_{\mu_P-\mu_L}\right)'\left(\boldsymbol{R}_{\mu_P-\mu_L}\right) \qquad (6.2.1)$$

where $\boldsymbol{R}_{\mu_P-\mu_L}$ was derived as Eq. (5.2.4) in Section 5.2 of Chapter 5.

6.3 Generalized Confidence Interval of SSDL

An upper 100(1-α)th percentile GCI for SSDL can be obtained from the following Monte-Carlo algorithm:

Step 1: Choose a large simulation sample size, say K=10,000. For k equal to 1 through K, carry out the following two steps.

Step 2: Generate LRx1 standard normal random vector **Z** and central chi-square random variable U with degree of freedom LJ-d-1.

Step 3: For the realized values of **Y** and S^2, compute $R_{\tau,k}$ defined in Eq. (6.2.1).

The required upper 100(1-α)th percentiles of the distribution of GPQ for SSDL, which is also the upper 100(1-α)th generalized confidence limit for SSDL, is then estimated by the 100(1-α)th sample percentiles of the collection of K=10,000 realizations $R_{\tau,1}$, $R_{\tau,2}$ ……., $R_{\tau,10000}$.

6.4 Statistical Testing Procedure

The upper 100(1-α)% generalized confidence limit for SSDL based on GPQ can be used to test the statistical hypothesis in (6.1.2) for linearity. The null hypothesis in (6.1.2) is rejected and the linearity of a analytical method is concluded at the α significance level if the upper 100(1-α)% generalized confidence limit for SSDL is less than $L\delta_0^2$.

6.5 Simulation Study

We conducted a simulation study to compare the empirical sizes and powers of the corrected Kroll's and GPQ-based SSDL methods. Following the specification of the experiment designs for evaluation of linearity, the number of solutions (or dilutions) of different concentrations is set to be 5 or 7 and the number of replications at each concentration is 2, 3, or 4. Throughout the simulation, mean concentration μ is assumed to be 4 and the allowable margin of linearity based on ADL, θ_0, is specified 0.05 as recommended by Kroll et al. (Kroll, 2000). From the relationship that SSDL = $L(\mu\theta)^2$,

where μ and θ are the mean of the concentrations, and ADL, respectively. It follows that the margin for SSDL for 5 and 7 concentrations are 0.2 and 0.28, respectively. In addition, standard deviation of normal random error is specified as 0.1 and 0.2. For each of 12 combinations, ten thousand (10,000) random samples are generated. For the 5% nominal significance level, a simulation study with 10,000 random samples implies that 95 percent of the empirical sizes evaluated at the allowable margins will be within 0.0457 and 0.0543 if the proposed methods can adequately control the size at the nominal level of 0.05.

Table 6.5.1 presents the results of the empirical sizes. All of empirical sizes for the corrected Kroll's is larger than 0.0543. On the other hand, all of empirical sizes of the GPQ method are within the range and showed that it has a better ability for controlling the size at the nominal level than the corrected Kroll's method. It was introduced in the previous chapters that the poor performance for the corrected Kroll's method in controlling the size results from the variability of estimators of non-centrality parameters for non-central χ^2 distribution of the observed ADL being estimated by the square root of residual mean square obtained from best-fitted polynomial model. On the contrary, since one requirement for GPQ is that $R_{\mu_P-\mu_L}$ is free of nuisance parameter σ, the GPQ approach can control the size at the nominal level.

Table 6.5.2 presents the results of the empirical powers. For the simulation, the true ADL is specified as 0.005 for both number of solutions of 5 and 7. The results given in Table 6.5.2 also show that the empirical power increases as the numbers of replicates or concentrations increases. Both the methods provide comparable powers except for the one of the GPQ-based SSDL method is 0.6962 when number of solution is 5, number of replicates is 2, and standard deviation of normal random error is 0.2. However, all

Table 6.5.1 Empirical sizes (corrected Kroll's method vs. GPQ-based SSDL method)

No. of Solution	No. of Replicate	Standard Deviation	Corrected Kroll	GPQ-based SSDL
5	2	0.1	0.0769	0.0535
		0.2	0.0734	0.0503
	3	0.1	0.0679	0.0523
		0.2	0.0643	0.0501
	4	0.1	0.0569	0.0476
		0.2	0.0596	0.0504
7	2	0.1	0.0670	0.0532
		0.2	0.0671	0.0532
	3	0.1	0.0573	0.0502
		0.2	0.0557	0.0476
	4	0.1	0.0563	0.0506
		0.2	0.0595	0.0529

empirical powers of the GPQ-based SSDL method for other combinations of parameters are still greater than 90%. In addition, from Table 6.5.1, the corrected Kroll's method fails to control the size at the nominal level. Therefore, the advantage of power by the corrected Kroll's method comes at the expense of inflated type I error rates.

Figure 6.5.1 and 6.5.2 present the empirical powers when the standard deviations of normal random error are 0.1 and 0.2, respectively with number of solutions is 5, number of replicates is 3, and the true ADLs are ranged from 0 to 0.08. Figure 6.5.1 shows that when standard deviation is 0.1, the empirical size at ADL=0.05 for the corrected Kroll's method is 0.0679, while the empirical size of the GPQ-based SSDL method is 0.0521. It shows that the GPQ method can control the size better than the other methods at the nominal level. In addition, the powers reach 0 and 1 at ADL=0.08 and 0.005, respectively for both methods. On the other hand, the power of the GPQ-based SSDL method is quite competitive to the corrected Kroll's method although it is little lower. The similar results are observed in Figure 6.5.2 when standard deviation of normal random error is 0.2. The empirical sizes for the corrected Kroll's and the GPQ-based SSDL methods at ADL=0.05 are 0.0827 and 0.0494, respectively. In addition, the powers for both methods when the standard deviation is 0.1 are lower than those when the standard deviation is 0.1.

6.6 Numerical Example

The same example of calcium used in Chapter 5 from Example 2 of CLSI guideline EP6-A (Tholen et al., 2003) is used to illustrate the proposed testing procedures. In this example, the allowable margin of percent bound for ADL is set as 0.05. As indicated in EP6-A (ICH Expert Working Group, 1995), the criteria of $|\mu_{Pi}-\mu_{Li}|$ for claiming

Table 6.5.2 Empirical powers with the true ADL=0.005 (corrected Kroll's method vs. GPQ-based SSDL method)

No. of Solution	No. of Replicate	Standard Deviation	Corrected Kroll	GPQ
5	2	0.1	1.0000	0.9994
		0.2	0.9261	0.6962
	3	0.1	1.0000	1.0000
		0.2	0.9454	0.9256
	4	0.1	1.0000	1.0000
		0.2	0.9828	0.9781
7	2	0.1	1.0000	1.0000
		0.2	0.9327	0.9078
	3	0.1	1.0000	1.0000
		0.2	0.9901	0.9873
	4	0.1	1.0000	1.0000
		0.2	0.9980	0.9972

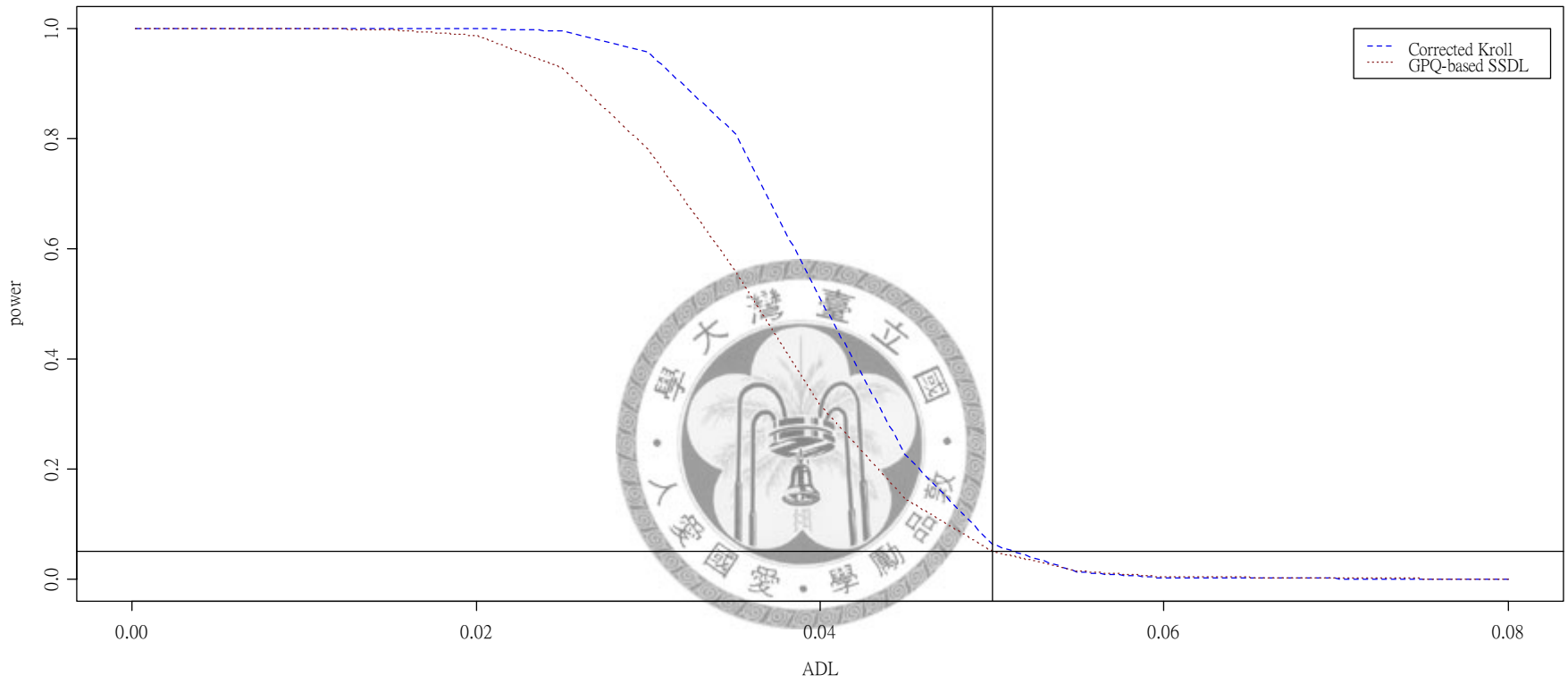


Figure 6.5.1 The empirical powers when standard deviation of normal random error is 0.1, number of solutions is 5, and number of replicates is 3 (corrected Kroll's method vs. GPQ-based SSDL method)

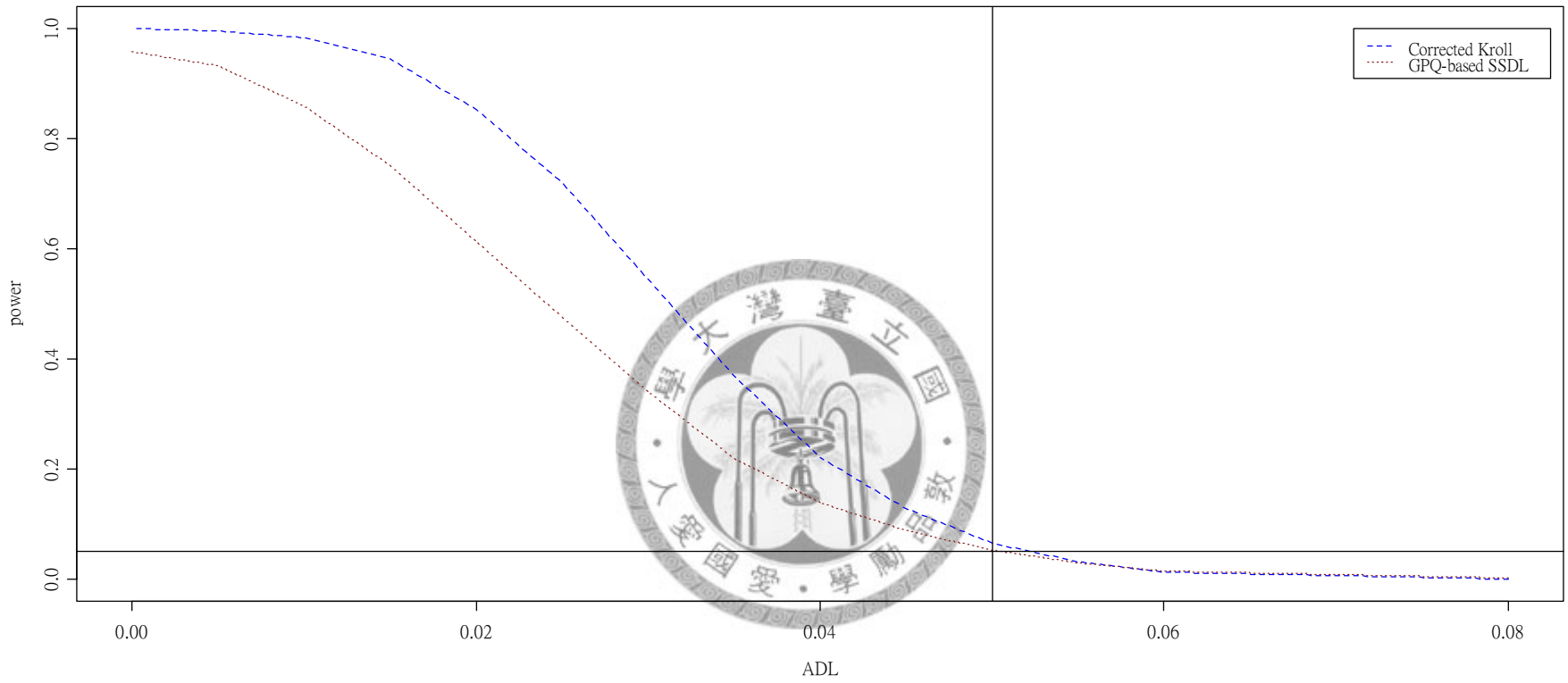


Figure 6.5.2 The empirical powers when standard deviation of normal random error is 0.2, number of solutions is 5, and number of replicates is 3 (corrected Kroll's method vs. GPQ-based SSDL method)

linearity is set as 0.2 mg/dL, the allowable limit of SSDL is set as 0.2 which is calculated by square of 0.2 mg/dL multiplying 5 concentrations. Table 6.6.1 presents the results of the two testing procedures. According to the decision rule of the corrected Kroll method, the analytical method can be concluded linear at the 5% significance level. On the other hand, the 95% upper limit confidence limit for SSDL of the GPQ methods is 0.2664, respectively. As a result, the GPQ-based SSDL method can not conclude that the analytical procedure is linear at the 5% significance level. The results presented above show the consistent results with the simulation results in Section 6.5 which the GPQ-based SSDL method is more conservative than the corrected Kroll's method. However, as demonstrated by the simulation, the GPQ-based SSDL method is the procedure that can adequately control the size at the nominal level.

6.7 Summary

The ADL is an aggregate criterion proposed by Kroll et al. (Kroll, 2000) for evaluating the linearity in assay validation. In this chapter, we propose an alternative criterion of SSDL based on the GPQ approach to assess the linearity. Simulation results show that the GPQ-based SSDL method not only can adequately control the type I error rate at the nominal level better than the corrected Kroll's method but also keep a competitive performance of the power. The reason for the poor performance corrected Kroll's method in controlling the size at the nominal level is the variability of estimators of non-central parameters for non-central χ^2 distribution of the observed ADL being estimated by the square root of residual mean square obtained from best-fitted polynomial model. Therefore, we can conclude the proposed statistical hypothesis based on the aggregate criteria SSDL in conjunction with the testing procedure derived from

Table 6.6.1 Results of the linearity evaluation for the example of calcium by the corrected Kroll's and GPQ-based SSDL methods

Method	Sample Statistic / Critical Value or Allowable Bound		Conclusion
Corrected Kroll	Sample ADL	0.0146	
	Critical Value	0.0437	**Linear**
GPQ-based SSDL	Upper 95% C.L.	0.2664	
	Allowable Upper Bound	0.2	Nonlinear

95% C.L. : Upper 95% Confidence limit. of SSDL

the GPQ method for evaluating the linearity in assay validation is better than the corrected Kroll's method.

Chapter 7

Alternative Criterion - Sum of Squares of the Deviation from Linearity Related to the Variation (CVDL)

The SSDL we introduced in Chapter 7 is based on the un-scaled deviations from linearity while ADL is based on the deviations from linearity scaled by the population average of concentrations of all solutions of the assay. Both ADL and SSDL do not take the experimental variability into consideration. As the repeatability is also the important characteristic which stands for reliability of a assay method, Wu (Wu, 2008) propose the coefficient of variation of the deviations from linearity (CVDL) as an alternative measure which can be used to evaluate the linearity and repeatability simultaneously.

7.1 CVDL and Statistical Hypothesis

As introduced in Section 3.2.3 of Chapter 3, CVDL is defined as:

$$\eta = \text{CVDL} = \frac{\sqrt{\sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2/L}}{\sigma} = \frac{\sqrt{\sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2/L}}{\sqrt{\sigma^2}}. \quad (7.1.1)$$

CVDL contains both information of the deviation from the linearity and the repeatability in term of $\sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2$ and σ, respectively, where $\sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2$ is the sum of squares for the difference in predicted values of the best-fitted model and σ is the residual mean square obtained from the best-fitted model as defined in the previous chapters.

The corresponding hypothesis for assessing linearity is then given as:

$H_0: \eta \geq \eta_0$ vs. $H_a: \eta < \eta_0$. (7.1.2)

where η_0 is the allowable limit of CVDL.

An estimator of CVDL can be also expressed in terms of SSDL as

$$\hat{\eta} = \frac{\sqrt{\sum_{i=1}^{L}\left(\hat{Y}_{Pi} - \hat{Y}_{Li}\right)^2/L}}{s}, \quad (7.1.3)$$

where s is the square root of the residual mean square obtained from the best-fitted model with degrees of freedom of LJ-d-1, and d is the degrees of freedom for regression of the best-fitted model.

7.2 Generalized Pivotal Quantity of CVDL

As it can be found in Eq. (7.1.1) for the definition of CVDL, the term of numerator is exactly the SSDL denoted by R_τ we introduced in Chapter 5. In addition, a GPQ of σ^2 can be obtained as

$$R_{\sigma^2} = \frac{(\text{LJ-d-1})s^2}{U} \qquad (7.2.1)$$

$$= \frac{s^2\sigma^2}{S^2} \qquad (7.2.2)$$

where U is the same chi-square random variable with degrees of LJ-d-1 we defined in Eq. (5.2.2).

From (7.2.1), R_{σ^2} has distribution that is free of parameters. In addition, when S^2 is substituted by its observed value s^2 in (7.2.2), then the observed value of R_{σ^2} denoted by r_{σ^2} is equal to σ^2 and free of any nuisance parameter. Hence, it fulfills the two requirements of (a) and (b) as described in Section 5.1 for being a GPQ for σ^2.

Therefore, a GPQ of CVDL can be obtained by:

$$R_\eta = \sqrt{\frac{R_\tau}{R_{\sigma^2}}} \qquad (7.2.3)$$

7.3 Generalized Confidence Interval of CVDL

An upper 100(1-α)th percentile GCI for CVDL can be obtained from the following Monte-Carlo algorithm:

Step 1: Choose a large simulation sample size, say K=10,000. For k equal to 1 through K, carry out the following two steps.

Step 2: Independently generate LJx1 standard normal random vector $\mathbf{Z}$ and U is the same chi-square random variable with degrees of LJ-d-1.

Step 3: For the realized values of $\mathbf{Y}$ and S^2, compute R_η as defined in (7.2.3).

The required upper 100(1-α)th percentiles of the distribution of GPQ for CVDL is then

estimated by the 100(1-α)th sample percentiles of the collection of K=10,000 realizations $R_{\eta,1}$, $R_{\eta,2}$,……..,$R_{\eta,10000}$.

7.4 Statistical Testing Procedure

The upper 100(1-α)% generalized confidence limit for CVDL based on GPQ can be used to test their respective statistical hypotheses in (7.1.2) for linearity. The null hypothesis in (7.1.2) is rejected and the linearity of a analytical method is concluded at the α significance level if the upper 100(1-α)% generalized confidence limit for CVDL is less than η_0.

7.5 Simulation Study

A simulation study is performed to compare the empirical sizes and powers of the corrected Kroll's and GPQ methods based on CVDL. The specifications of the simulation study are given as follows: The number of solutions (or dilutions) of different concentrations is set to be 5 or 7 and the number of replications at each concentration is 2, 3, or 4. Throughout the simulation, mean concentration μ is assumed to be 4. If following that the allowable margin of linearity based on ADL, θ_0, is specified at 0.05 as recommended by Kroll et al. (Kroll, 2000). Using the relationship that CVDL $=\sqrt{\theta \times L \times \mu}/\sigma$, where μ is the population mean concentrations for all solutions of the assay and θ is ADL. The allowable limit η_0 is 2 and 1 for σ being 0.1 and 0.2, respectively. For each of 12 combinations, ten thousand (10,000) random samples are generated. For the 5% nominal significance level, a simulation study with 10,000 random samples implies that 95 percent of the empirical sizes evaluated at the

allowable margins will be within 0.0457 and 0.0543 if the proposed methods can adequately control the size at the nominal level of 0.05.

The results of the empirical sizes are provided in Table 7.5.1. All the empirical sizes of the GPQ method based on CVDL are within the range between 0.0457 and 0.0543, while all empirical sizes of the corrected Kroll's method are larger than 0.0543. The simulation results reveal that the GPQ-based CVDL method can adequately control the size at the nominal level. The reason for a better performance of the GPQ-based CVDL method may be that the distribution of GPQ is free of their respective nuisance parameters. On the other hand, the corrected Kroll's method fails to take into account the variability in estimator of the non-centrality parameter of the non-central chi-square distribution.

The results of the empirical powers are presented in Table 7.5.2. In Table 7.5.2, the true value of ADL is assumed to be 0.005 for both number of solutions of 5 and 7. The results in Table 7.5.2 also show that the empirical power of both methods is an increasing function of the number of replicates and number of solutions. In addition, the empirical power of the GPQ-based CVDL method is competitive to the corrected Kroll's method. Although the empirical power of the corrected Kroll's method is larger than that of the GPQ-based CVDL method, its better performance on the empirical power results from inflation of the size above the nominal level.

Figure 7.5.1 and 7.5.2 present the empirical powers of the four methods when σ are 0.1 and 0.2, respectively with number of solutions is 5, number of replicates is 3. The true values of ADL are ranged from 0 to 0.08. A comparison of Figure 7.5.1 and Figure 7.5.2 reveals that the power of both methods is a deceasing function of σ. The power curve of the GPQ-based CVDL method is uniformly lower than that of the corrected

Kroll's method. However, the empirical power of GPQ-based CVDL method at ADL=0.05 is 0.0511 while which for corrected Kroll's method is 0.0623. Therefore, it show that show that the GPQ-based CVDL method can control the size at the nominal level while corrected Kroll's method cannot.

7.6 Numerical Example

The same numerical data of calcium in the previous chapters is used to illustrate the proposed testing procedures in evaluation of linearity of an analytical procedure. Following EP6-A (Tholen et al., 2003), the criterion of $|\mu_{Pi}-\mu_{Li}|$ for linearity is set as 0.2 mg/dL for all 5 concentrations. In this example, the allowable margin of percent bound for ADL is set as 0.05. On the other hand, the allowable limit of the GPQ-based SSDL is set as 0.2 which is calculated by square of 0.2 mg/dL multiplying 5 concentrations. We also assume that the allowable repeatability set by the manufacturer is 0.2. Therefore, the allowable margin of the GPQ-based CVDL is 1 which is equal to the allowable margin of 0.2 for SSDL divided by the product of 5 (concentrations) and square of the repeatability of 0.2, i.e., $\eta=\sqrt{\tau/(L\sigma^2)}$. The results of the corrected Kroll's and the GPQ-based CVDL methods are provided in Table 7.6.1. The linearity is concluded by corrected Kroll's method since the observed ADL yields a value of 0.0146 is less than the critical value of 0.0437 with respect to a margin of percent bound of 5%. On the other hand, the 95% upper confidence limits for CVDL methods is 1.9125. Its 95% upper confidence limits is larger than their respective allowable upper limits of 1. As a result, the GPQ-based CVDL method can not conclude the linearity of the analytical procedure at the 5% significance level. As shown in simulation results and conservative than the corrected Kroll's method.

Table 7.5.1 Empirical sizes (corrected Kroll's method vs. GPQ-based CVDL method)

No. of Solutions	No. of Replicates	Standard Deviation	Corrected Kroll	GPQ-based CVDL
5	2	0.1	0.0702	0.0540
		0.2	0.0763	0.0467
	3	0.1	0.0623	0.0513
		0.2	0.0655	0.0511
	4	0.1	0.0594	0.0489
		0.2	0.0595	0.0509
7	2	0.1	0.0655	0.0490
		0.2	0.0635	0.0504
	3	0.1	0.0592	0.0473
		0.2	0.0583	0.0504
	4	0.1	0.0562	0.0529
		0.2	0.0571	0.0452

Table 7.5.2 Empirical powers with the true ADL=0.005 (corrected Kroll's method vs. GPQ-based CVDL method)

No. of Solutions	No. of Replicates	Standard Deviation	Corrected Kroll	GPQ-based CVDL
5	2	0.1	1.0000	0.9876
		0.2	0.9670	0.7754
	3	0.1	1.0000	0.9989
		0.2	0.9965	0.9212
	4	0.1	1.0000	0.9998
		0.2	0.9996	0.9678
7	2	0.1	1.0000	0.9979
		0.2	0.9923	0.8994
	3	0.1	1.0000	0.9999
		0.2	0.9996	0.9742
	4	0.1	1.0000	1.0000
		0.2	1.0000	0.9932

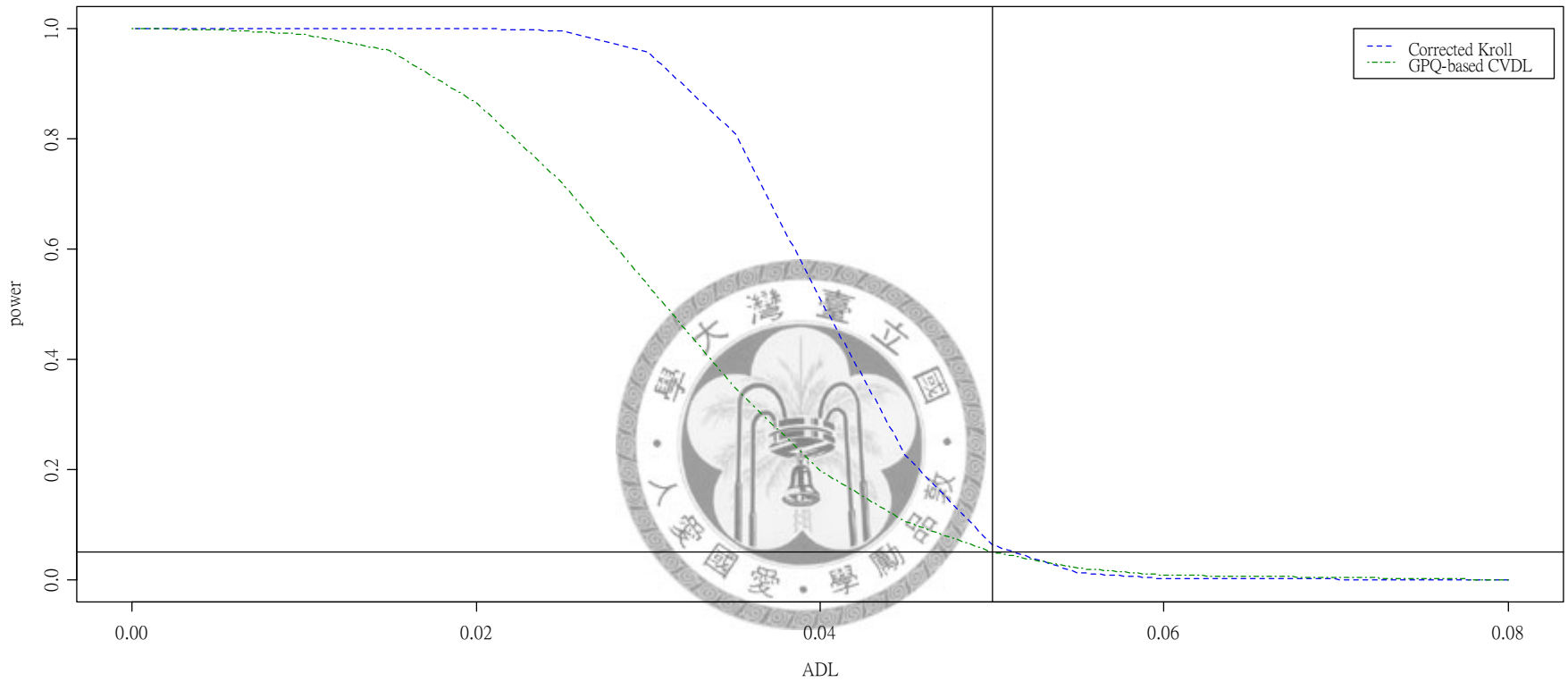


Figure 7.5.1 The empirical powers when standard deviation of normal random error is 0.1, number of solutions is 5, and number of replicates is 3 (corrected Kroll's method vs. GPQ-based CVDL method)

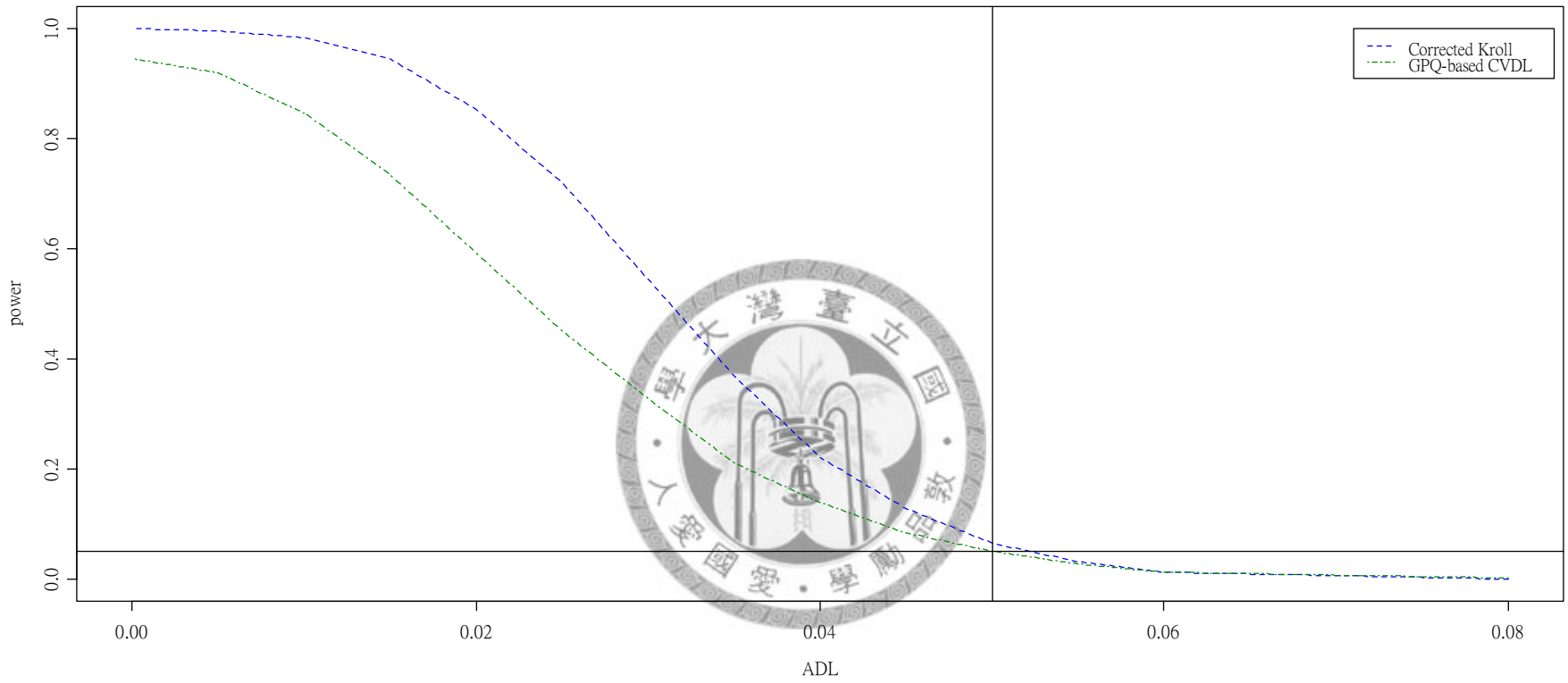


Figure 7.5.2 The empirical powers when standard deviation of normal random error is 0.2, number of solutions is 5, and number of replicates is 3 (corrected Kroll's method vs. GPQ-based CVDL method)

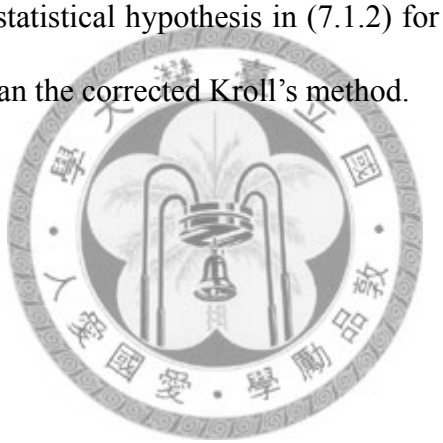
Table 7.6.1 Results of the linearity evaluation for the example of calcium by corrected Kroll's and GPQ-based CVDL methods

Method	Sample Statistic / Critical Value or Allowable Bound		Conclusion
Corrected Kroll	Sample ADL	0.0146	
	Critical Value	0.0437	**Linear**
GPQ-based CVDL	Upper 95% C.L.	1.9125	
	Allowable Upper Bound	1	Nonlinear

95% C.L. : Upper 95% Confidence limit

7.7 Summary

Both ADL and CVDL are the aggregate criterion for assessment of linearity in assay validation. The main difference between these two criteria is the proposed CVDL is an criterion not only contain the information of the deviations from linearity but also the repeatability of the analytical procedure. The simulation results presented above show that the corrected Kroll's method inflates the type I error rate and the GPQ-based CVDL methods can control the size at the nominal level. In addition, the GPQ-based CVDL method also keep the good power performance. Therefore, we conclude the GPQ-based CVDL with respect to the statistical hypothesis in (7.1.2) for evaluating the linearity in assay validation is better than the corrected Kroll's method.



Chapter 8

Discussion and Summary

Various aggregate criteria including ADL, SSDL and CVDL for evaluating the linearity in assay validation were introduced in Chapter 2 to 7. Although these criteria are formulated by different components which provide the different characteristics, however, the common feature of these criteria is that all of them contain the sum of square for the deviations from linearity as the major component. In this chapter, we discuss the relationship among these criteria. In addition, the results of the simulation study and numerical example are used to compare the performances and characteristics of each aggregate criterion for the assessment of linearity in assay validation.

8.1 Relationship of the Aggregate Criteria

Recall the definition for ADL, SSDL and CVDL are defined in the previous chapters as follows:

$$\theta = \mathrm{ADL} = \frac{\sqrt{\sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2/L}}{\mu}$$

$$\tau = \mathrm{SSDL} = \sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2.$$

$$\eta = \mathrm{CVDL} = \frac{\sqrt{\sum_{i=1}^{L}(\mu_{Pi} - \mu_{Li})^2/L}}{\sqrt{\sigma^2}}$$

where μ_{Pi} and μ_{Li} are the predicted mean of the best-fit polynomial model and linear regression model at ith concentration with i = 1,.., L, respectively; μ is the population mean concentration for all solutions of the assay and σ^2 is the variance of residual under the best-fitted model. It can be found that the SSDL is the common component for each criterion. Their relationship can easily be constructed as the following:

$$\tau = L(\mu\theta)^2 = L(\eta\sigma)^2 \qquad (8.1.1)$$

Unlike that SSDL is the unscaled deviation defined as the pure sum of square of the deviations from the linearity, both CVDL and ADL are the scaled deviations. ADL is the square root of the average sum of squares of the scaled deviations by μ, while CVDL is scaled by the variability or repeatability of the best-fitted model.

8.2 Comparison by Simulation Study

A simulation studies was employed to compare the empirical sizes and powers among three GPQ-based ADL, SSDL and CVDL methods. Parts of the results from the same simulation study were presented in Chapter 5, 7 for comparing the performance of the corrected Kroll's method with GPQ-based ADL and GPQ-based CVDL methods, respectively. As described in Chapter 5 and 7, the specifications of the simulation study are given as follows: The number of solutions (or dilutions) of different concentrations is set to be 5 or 7 and the number of replications at each concentration is 2, 3, or 4. Throughout the simulation, mean concentration μ is assumed to be 4. The allowable margin of linearity based on ADL, θ_0, is specified at 0.05. From the relationship of $\tau = L(\mu\theta)^2 = L(\eta\sigma)^2$ in Eq. (8.1.1), it follows that the margin for SSDL are 0.2 and 0.28 for 5 and 7 concentrations, respectively. In addition, under the specification of standard

deviation of normal random error is specified as 0.1 and 0.2, the allowable limit η_0 is 2 and 1 for σ being 0.1 and 0.2, respectively. For each of 12 combinations, ten thousand (10,000) random samples are generated. For the 5% nominal significance level, a simulation study with 10,000 random samples implies that 95 percent of the empirical sizes evaluated at the allowable margins will be within 0.0457 and 0.0543 if the proposed methods can adequately control the size at the nominal level of 0.05.

The results of the empirical sizes are provided in Table 8.2.1. All of empirical sizes of the GPQ methods based on ADL, SSDL and CVDL are within the range between 0.0457 and 0.0543. The simulation results reveal that the GPQ-based methods for SSDL, ADL, and CVDL can adequately control the size at the nominal level. On the other hand, according to the empirical powers of three GPQ-based methods presented in Table 8.2.2, the empirical power of the GPQ-based ADL method is larger than that of the GPQ-based SSDL which is in turn larger than that of the GPQ-based CVDL method.

Figure 8.2.1 and 8.2.2 present the empirical powers of the three GPQ-based methods when σare 0.1 and 0.2, respectively with number of solutions is 5, number of replicates is 3. For Figure 8.2.1, when the ADL = 0.05, the empirical size for the three GPQ-based methods are 0.0499, 0.0513, and 0.0502 for SSDL, CVDL, and ADL respectively. Similar findings are observed in Figure 8.2.2. Again these results show that the three GPQ-based procedures can adequately control the size at the nominal level of 5%. Moreover, Both figures demonstrate that the GPQ-based ADL procedure is uniformly more powerful than the GPQ-based SSDL method which is in turn uniformly more powerful than the GPQ-based CVDL method. For example, in Figure 8.2.1 when ADL is 0.03, the empirical powers are 0.9426, 0.7803, and 0.5337, respectively for the GPQ-based ADL, SSDL, and CVDL methods. In other words, the GPQ-based ADL

Table 8.2.1 Empirical sizes (GPQ-based SSDL vs. GPQ-based CVDL vs. GPQ-based ADL methods)

No. of Solutions	No. of Replicates	Standard Deviation	GPQ-based SSDL	GPQ-based CVDL	GPQ-based ADL
5	2	0.1	0.0462	0.0540	0.0467
		0.2	0.0523	0.0467	0.0517
	3	0.1	0.0499	0.0513	0.0502
		0.2	0.0522	0.0511	0.0517
	4	0.1	0.0498	0.0489	0.0505
		0.2	0.0504	0.0509	0.0508
7	2	0.1	0.0504	0.0490	0.0501
		0.2	0.0495	0.0504	0.0494
	3	0.1	0.0505	0.0473	0.0509
		0.2	0.0495	0.0504	0.0498
	4	0.1	0.0498	0.0529	0.0498
		0.2	0.0498	0.0452	0.0510

Table 8.2.2 Empirical powers with the true ADL=0.005 (GPQ-based SSDL vs. GPQ-based CVDL vs. GPQ-based ADL methods)

No. of Solutions	No. of Replicates	Standard Deviation	GPQ-based SSDL	GPQ-based CVDL	GPQ-based ADL
5	2	0.1	0.9995	0.9876	1.0000
		0.2	0.6976	0.7754	0.9331
	3	0.1	1.0000	0.9989	1.0000
		0.2	0.9326	0.9212	0.9942
	4	0.1	1.0000	0.9998	1.0000
		0.2	0.9814	0.9678	0.9995
7	2	0.1	1.0000	0.9979	1.0000
		0.2	0.9123	0.8994	0.9888
	3	0.1	1.0000	0.9999	1.0000
		0.2	0.9850	0.9742	0.9994
	4	0.1	1.0000	1.0000	1.0000
		0.2	0.9981	0.9932	1.0000

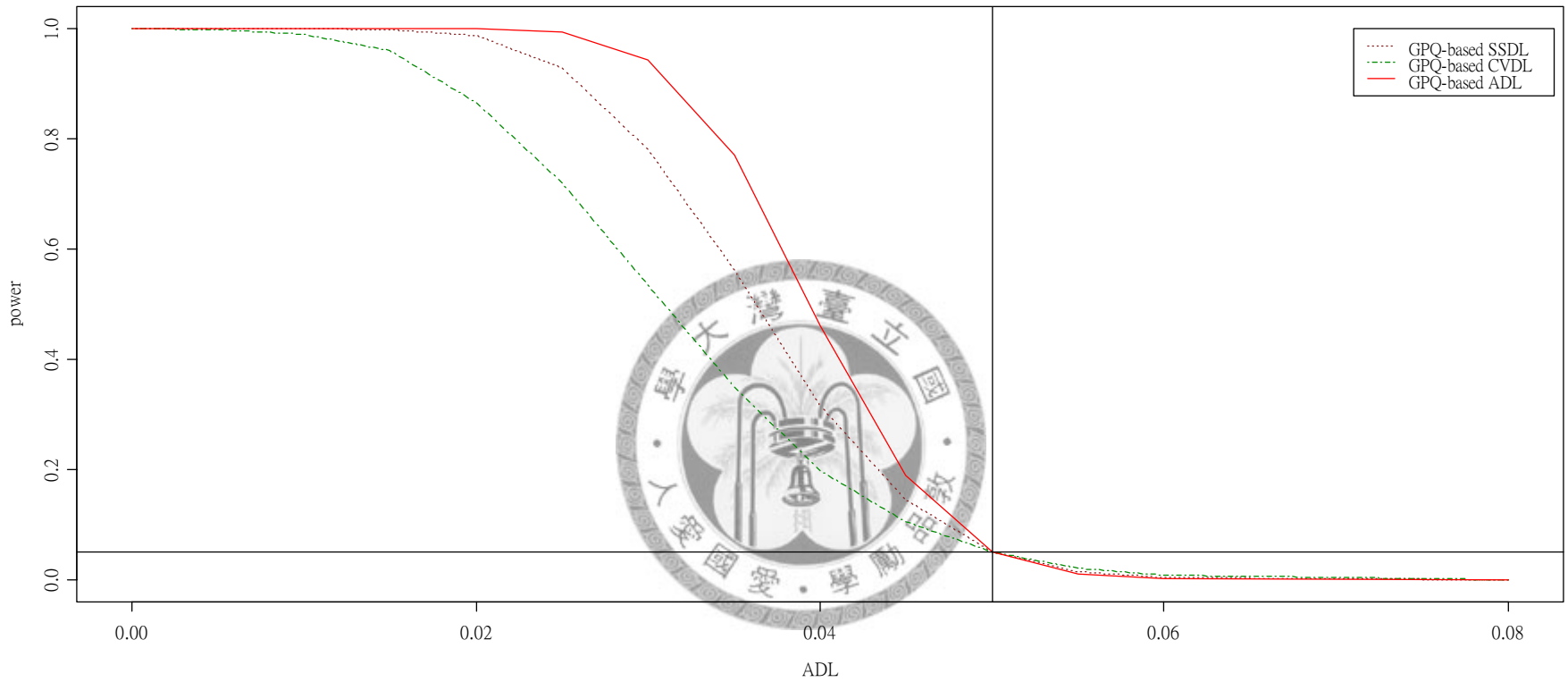


Figure 8.2.1 The empirical powers when standard deviation of normal random error is 0.1, number of solutions is 5, and number of replicates is 3 (GPQ-based SSDL vs. GPQ-based CVDL vs. GPQ-based ADL methods)

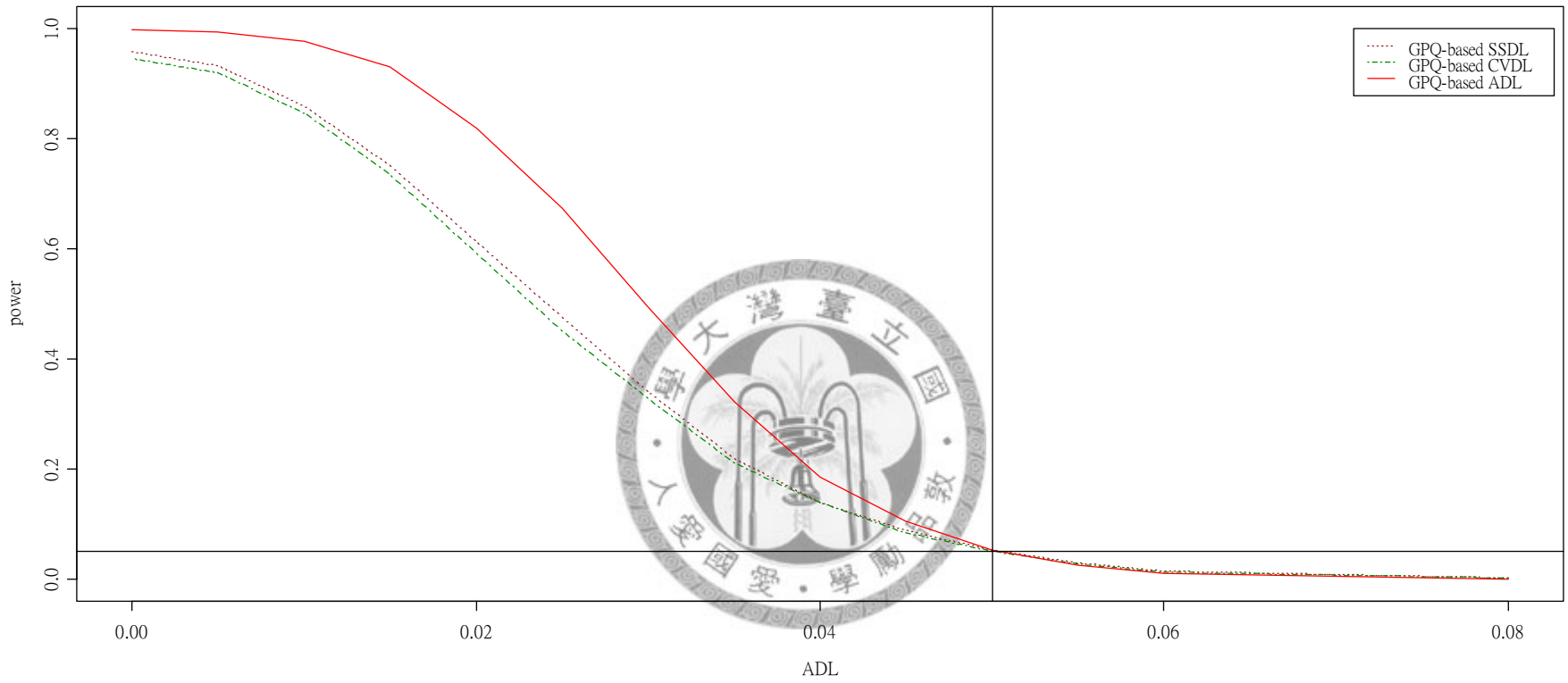


Figure 8.2.2 The empirical powers when standard deviation of normal random error is 0.2, number of solutions is 5, and number of replicates is 3 (GPQ-based SSDL vs. GPQ-based CVDL vs. GPQ-based ADL methods)

procedure is 40% more powerful than the GPQ-based CVDL method and is 16% more powerful than the GPQ-based SSDL method at the ADL of 0.03. Therefore, the improvement of the power provided by the GPQ-based ADL method is impressively substantial.

8.3 Numerical Example

The previous example of calcium is used to illustrate the proposed testing procedures in evaluation of linearity of an analytical procedure. Under the criteria of $|\mu_{Pi}-\mu_{Li}|$ for linearity and repeatability are 0.2mg/dL and 0.2mg/dL, respectively, and the allowable margin of percent bound for ADL is set as 0.05, the corresponding criteria and results of three GPQ-based ADL, SSDL and CVDL methods are presented in Table 8.3.1. The results show that the 95% upper confidence limit for the ADL computed by the GPQ method is 0.0218 which is smaller than the allowable upper limit of 0.05. Hence, the linearity of the analytical procedure can be concluded at the 5% significance level by the GPQ-based ADL procedure. On the other hand, the 95% upper confidence limits for SSDL of the GPQ-based SSDL and CVDL methods are 0.2471 and 1.9125, respectively. Both 95% upper confidence limits are larger than their respective allowable upper limits of 0.2 and 1. As a result, both methods can not conclude the linearity of the analytical procedure at the 5% significance level. The results presented above show the different conclusions between the GPQ-based methods. As shown in simulation results, all three GPQ-based methods can control the size at the nominal size of 0.05, the GPQ-based ADL method is uniformly more powerful than the other two GPQ-based methods. This might be one of the reasons why the linearity can be claimed by the GPQ-based

Table 8.3.1 Results of the linearity evaluation by three different methods

Method	Sample Statistic / Critical Value or Allowable Bound		Conclusion
GPQ-based SSDL	Upper 95% C.L.	0.2471	Nonlinear
	Allowable Upper Bound	0.2	
GPQ-based CVDL	Upper 95% C.L.	1.9125	Nolinear
	Allowable Upper Bound	1	
GPQ-based ADL	Upper 95% C.L.	0.0218	**Linear**
	Allowable Upper Bound	0.05	

95% C.L. : Upper 95% Confidence limit

ADL method.

8.4 Summary

In this chapter, we discuss the relationship among three different aggregate criteria of ADL, SSDL and CVDL. As we mentioned in Section 8.1, the SSDL, i.e., the sum of square of the deviation from the linearity is the basis of three criteria. On the other hand, ADL and CVDL are the scale measures scaled by the average concentration and repeatability, respectively. As the demonstrated by the simulation results, all three GPQ-based ADL, SSDL, CVDL methods can control the size at the nominal level. Moreover, simulation results reveal that the GPQ-based ADL procedure is uniformly more powerful than the GPQ-based SSDL and CVDL methods. In addition, CVDL method is the most conservative procedure among all three GPQ-methods. This may be due to the reason that it is scaled by the repeatability and it requires both the predicted means and repeatability of the best-fitted model to meet the allowable limits. On the other hand, the GPQ-based ADL procedure not only adequately control the type I error rate but also is uniformly more powerful than the other GPQ-based method. Therefore, the GPQ-based ADL procedure will be recommended to be the better procedure for evaluating the linearity in assay validation among the three GPQ-based methods. However, as the GPQ-based CVDL procedure considers linearity and repeatability simultaneously in one measure, one may consider using CVDL as the criterion for assay validation if he/she would like to evaluate accuracy and reliability simultaneously.

Chapter 9

Concluding Remarks

9.1 Conclusion

One of the most important characteristics for evaluation of accuracy and precision in assay validation is linearity. Even though the best-fitted model is not linear, linearity of the analytical procedure can still be claimed if the difference in the predicted means between the best-fitted and linear models is smaller than some pre-specified allowable limit at all concentrations employed in the validation experiment. As a result, the deviation from linearity is the fundamental unit for assessment of bias for evaluation of linearity.

With respect to the disaggregate criterion, the approved CLSI EP6-A guideline proposes the estimation method by comparing the estimates of the differences in the predicted means with the pre-specified allowable limit directly without the formal statistical inference procedure. The method completely ignores the variation of the estimate and inflates the type I error of the results of the evaluation. On the hand, the ADL proposed by Kroll et al. (Kroll, 2000) is an aggregate criterion constructed by the sum of square of from the deviations from linearity scaled by the mean concentrations. However, the statistical testing procedure proposed by Kroll et al. (Kroll, 2000) not only incorrectly formulates the hypothesis for proving linearity but also contained the unknown nuisance parameters in the distribution of ADL which causes the problem for

controlling the size at the nominal level. Therefore, we propose the TOST procedure and corrected Kroll's method to improve the shortcomings of the above two methods by providing the formal statistical testing procedure instead of the estimation method and reformulating the correct hypothesis for the uncorrected Korll's method, respectively. The simulation results show the proposed methods can control the size better than the two current methods. On the other hand, to overcome the issue raised by the unknown nuisance parameters of the distribution of ADL, we propose the GPQ-based ADL method for eliminating the unknown parameter in the distribution by applying the concept of generalized confidence interval proposed by Weerahadi (Weerahandi, 1993). The proposed GPQ method not only can control the size at the nominal level better than the corrected Kroll's method but also keep the good performance of the power for assessment of linearity in assay validation.

In addition to ADL proposed by Kroll et al. (Kroll, 2000), we also introduce two new alternative criteria SSDL and CVDL. SSSL is an un-scaled measure which is formulate by the sums of the square of the deviation from linearity, while CVDL is a scaled measure which is scaled by the variability of the best-fitted model for assessment of linearity. The major difference of CVDL with other two aggregate criteria is that CVDL considers both accuracy and reliability with respect to an analytical method into one measure simultaneously. With respect to SSDL, one may consider the following test statistic for evaluating linearity using F-test:

$$\psi = \frac{\sum_{i=1}^{L}\sum_{j=1}^{J}\left(\hat{\mu}_{Pi} - \hat{\mu}_{Li}\right)^2 / (d\text{-}1)}{\sum_{i=1}^{L}\sum_{j=1}^{J}\left(Y_{ij} - \hat{\mu}_{Pi}\right)^2 / (LJ\text{-}d\text{-}1)}$$

Under the null hypothesis of hypothesis (3.2.2.2), i.e., $\sum_{i=1}^{L}(\mu_{Pi}-\mu_{Li})^2=L\delta_0^2$, ψ is distributed as an non-central $F_{d\text{-}1,LJ\text{-}d\text{-}1}$ distribution with non-centrality parameter of $\frac{LJ\delta_0^2}{\sigma^2}$. However, there is still unknown parameter σ^2 in the non-centrality parameter of the distribution. If the statistical testing is performed based on ψ with non-central $F_{d\text{-}1,LJ\text{-}d\text{-}1}$ distribution by substituting σ^2 using its estimates, the type-I error may still be inflated due to the variability of estimates of σ^2. Therefore, the GPQ approach is proposed to solve the issue of the unknown parameters in the distribution of the estimators of each aggregate criterion. Our simulation results show all three GPQ-based ADL, SSDL and CVDL method can not only control the size better than corrected Kroll's method but also maintain the good performance of the power. On the other hand, it also show the GPQ-based ADL procedure is uniformly more powerful than the GPQ-based SSDL and CVDL methods.

In addition to the proposed GPQ approach, a bootstrap procedure may be a reasonable approach to evaluation of linearity for the proposed aggregate criteria. However, bootstrap procedures may suffer a disadvantage that the sampling distributions of the observed ADL, SSDL and CVDL involve unknown nuisance parameters which need to be substituted by their estimates when generating the bootstrap samples. Bootstrap procedures will still inflate type I error rate due to variability of estimates of unknown parameters. On the other hand, derivation of generalized pivotal quantities is based on the sampling distribution of the sample mean, the mean square of the best-fitted model. As result, our proposed GPQ procedures do incorporate the sampling variability of the estimated parameters. In addition, the observed GPQ is free of the nuisance parameters. This is another novelty of our

proposed procedure which applies the technique of GPQ to resolve the issue of nuisance parameters for the inference of the proposed aggregate criteria on evaluation of linearity.

The other issue needs to be noted is about the design of experiment for evaluation of linearity. As it has already known that the variability of the predicted values of the fitted regression models will become larger at the concentration levels which are close to the start and end points of the range of selected concentration levels. Therefore, the optimal design with the selection of appropriate concentration levels including the number of concentration levels, the value of concentration levels and the number of samples at each concentration levels by considering the change of the variability for the predicted values at different concentration levels needs to be considered. As one of the purposes for the evaluation of linearity is to decide the range of concentration levels with linearity, after selecting out the concentration levels without nonlinearity according to the criteria of EP6-A guideline (Tholen et al., 2003), an equal space design, i.e., equal difference between each two neighbor concentration levels, which is the design with most efficiency is recommend.

In our research, we introduce the TOST procedure for the disaggregate criterion as well as the GPQ-based procedure for the different aggregate criteria. All of the proposed procedures show the good performance in controlling the size and power for assessment of linearity in assay validation. In addition, the evaluation procedure based on the disaggregate criterion is more conservative than which based on the aggregate criterion because it requires that the differences in predicted means between the best-fitted model and linear models for all solutions be within the pre-specified limit, while the aggregate criterion only requires the magnitude of sum of deviations from linearity be controlled

within a aggregate limit. The choice of the disaggregate-based procedure and aggregate based procedure may depend on how accuracy the assay method is required. In addition, although the GPQ-based ADL procedure is recommended to be used for assessment of linearity in assay validation since it is the uniformly more powerful than the other two GPQ-based methods. However, one may consider using CVDL as the criterion for assay validation if he/she would like to evaluate accuracy and reliability simultaneously.

9.2 Other Application and Future Research

As we introduced in Chapter 5 that SSDL is an aggregate criterion formulated by the nature of disaggregate criterion proposed by the CLSI EP6-A guideline (Tholen et al., 2003) as the form of model-by-dilution interaction. The similar concept for aggregate criterion of model-by-dilution interaction for assessing linearity can also be implemented to the area of investigating the consistency of treatment differences of a pharmaceutical product among different populations. For instance, the pharmaceutical companies conduct the bridging study as a supplementary study in the new region to provide pharmacodynamic or clinical data on efficacy, safety, dosage and dose regimen to allow extrapolation of the foreign clinical data to the population of the new region for getting the drug approval in the new region. The interest of the center effect is also for a multi-center clinical trial to evaluate consistence of the drug effect among study centers. In addition, scientists or medical expert may have interest in evaluating the similar effect of the pharmaceutical product can be obtained for adults and children. The treatment-by-population interaction can be considered as a measure for evaluating if the effects of a pharmaceutical product are consistent among populations based on the aggregate criteria, where population can be the region, center or age group according to

the application. However, our aim is not to detect existence of the treatment-by-population interaction but rather to verify whether the treatment-by-population interaction is within some pre-specified allowable upper limit. Therefore one of possible hypothesis can be formulated as follows:

$$H_0 : \frac{\sum_{k=1}^{L}\left(\mu_{\Delta k} - \overline{\mu_\Delta}\right)^2}{L} \geq \theta_0^2 \quad \text{vs.} \quad H_a : \frac{\sum_{k=1}^{L}\left(\mu_{\Delta k} - \overline{\mu_\Delta}\right)^2}{L} < \theta_0^2$$

where $\mu_{\Delta k}$ and $\overline{\mu_\Delta}$ are the treatment difference in population k and overall mean treatment difference among L populations, respectively, between investigational and control products. θ_0 is the allowable upper bound of treatment difference for each of L populations. The research for this application has already been investigated as the separate topic.

With respect to the further research, there are some topics related to our current research which can be considered as the following:

(1) As both accuracy and reliability are important for an assay method, the proposed CVDL is proposed as a criterion for not only assessing accuracy but also evaluating reliability simultaneously. The alternative approach for evaluating accuracy and reliability simultaneously is to perform a multiple comparison by conducting two statistical testing procedures for assessment of linearity and repeatability, respectively.

(2) CVDL is actually an aggregate criterion for aggregating the measures of accuracy and reliability by their ratio. In some situation, the different importance of accuracy and reliability for some specific assay method may be considered. The weighted sum of the measures for accuracy and reliability probably can be

considered as an alternative aggregate criterion for this type of evaluation.

(3) CLSI EP6-A guideline (Tholen et al., 2003) suggests an experiment with 5 to 7 concentration levels and at least two replicates of samples should be employed for evaluation of linearity. However, the determination of sample size in experiment should not be fixed but based on the consideration of the magnitude of the allowable limit of evaluation criterion. Therefore, the selection the appropriate sample size with desired power and significant level under the pre-specified allowable limit of the evaluation criterion is one of the topics for future research.

Bibliography

Hanning J, Iyer HK, Patterson P. (2006). Fiducial generalized confidence intervals. J. Am. Stat. Assoc. **101**: 254-269.

ICH Expert Working Group, International Conference on Harmonization Tripartite Guideline Q2A: Test on Validation of Analytical Procedures. (1995).

Tholen DW, Kroll M, Astles JR, Caffo AL, Happe TM, Krouwer J, Lasky F. EP6A: Evaluation of the Linearity of Quantitative Measurement Procedures: A Statistical Approach; Approved Guideline, Clinical Laboratory Standard Institute, Wayne, PA, U.S.A. 2003.

Tsui KW, Weerahandi S. (1989). Generalized P-values in Significance Testing of Hypotheses in the Presence of Nuisance Parameters. J. Am. Stat. Assoc.; 84: 602-607.

Weerahandi S. (1993). Generalized Confidence Intervals. J. Am. Stat. Assoc. 88: 899-905.

Wu HJ. (2008). A study on statistical methods for evaluation of linearity in assay validation. Unpublished Master Thesis, Division of Biometry, Department of Agronomy, National Taiwan University.

Appendix

Instruction for Using the Fortran Program for Linearity Evaluation of Assay Validation

Introduction:

The program is developed based on the statistical methods presented in the published paper entitled "Statistical Methods for Evaluating the Linearity in Assay Validation" (Eric Hsieh, and Chin-fu Hsiao, Jen-pei Liu, 2008).

System Requirement

Fortran Version: Compaq Visual Fortran 6.6 with IMSL Library

Instruction

The program includes the following functions:

1. Input Data:
   There is the user friendly interface which allows users can input the following information based on their experiment design:
   (1) Desired statistical significant level
   (2) Desired allowable limit for different of predicted values between linear and polynomial models
   (3) Number of Solutions
   (4) Number of Replicates
   (5) Measures obtained from their experiments

2. Output
   There are two types of output provided in the applications

(1) Output on the screen

The output on the screen provides the basic information of the design and results of the linearity evaluation by each statistical procedure in the paper.

(2) Output to the text file

The application allows user to specify the desired path and file name to save the detail results of the evaluation including:

- Inputted measures obtained from the experiment
- Summary of Regression Analysis
- Mean Differences between the Best-fitted Curve and Simple Linear Regression Equation
- Results of the linearity evaluation by each statistical procedure

3. Demo of the Operation

Please refer to the Fig 1 and 2 for demo of the operation.

4. Sample Content in the Output File for Detail Results

Please refer to the output in Example.txt for sample detail Results.

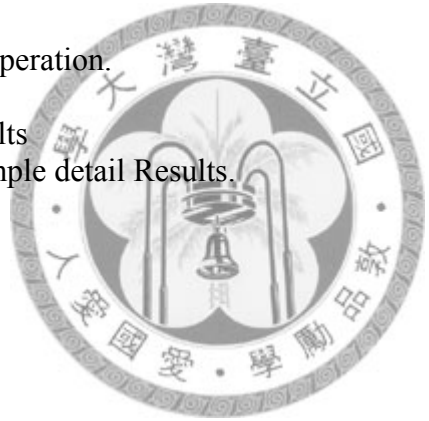


Figure- 1

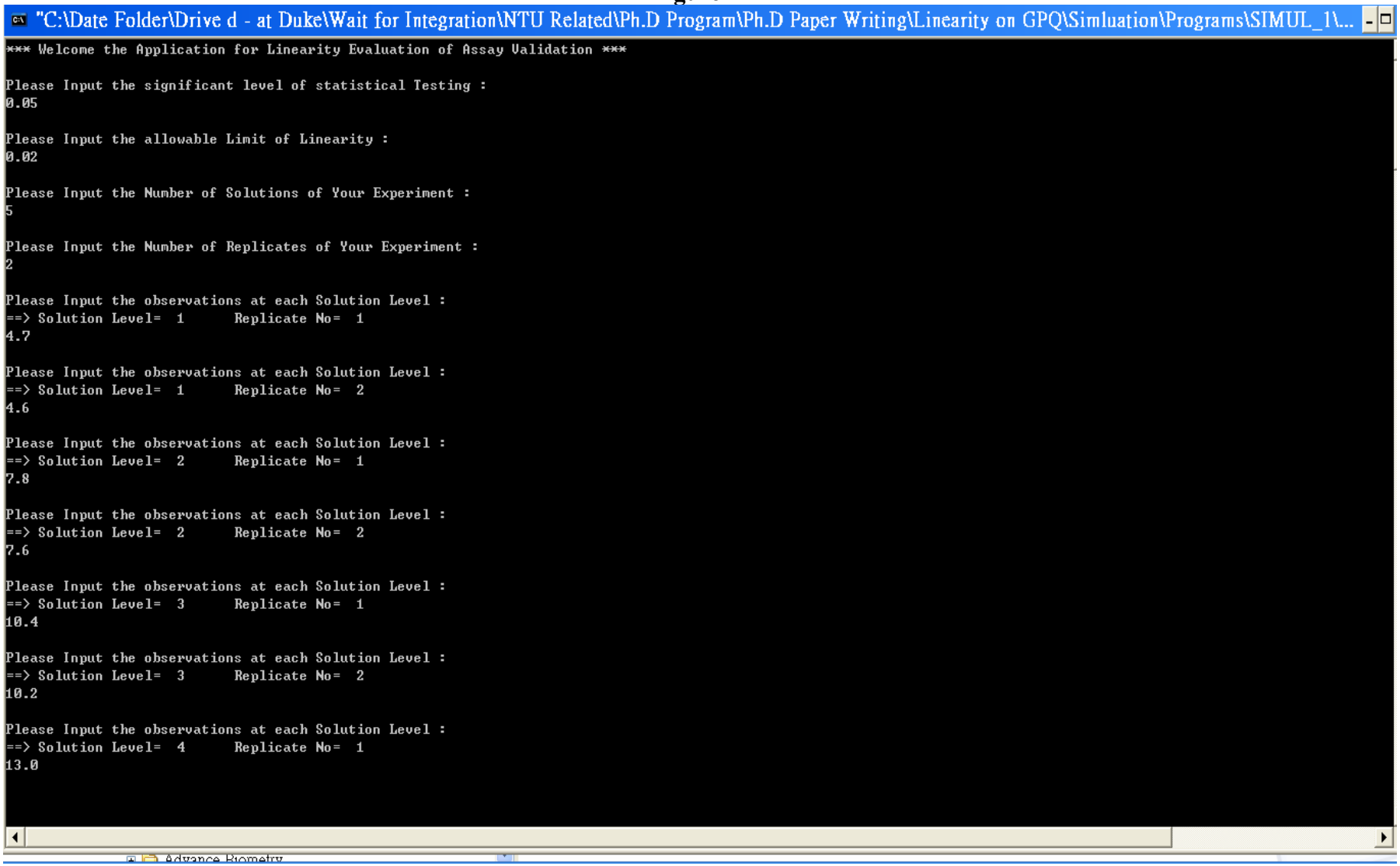
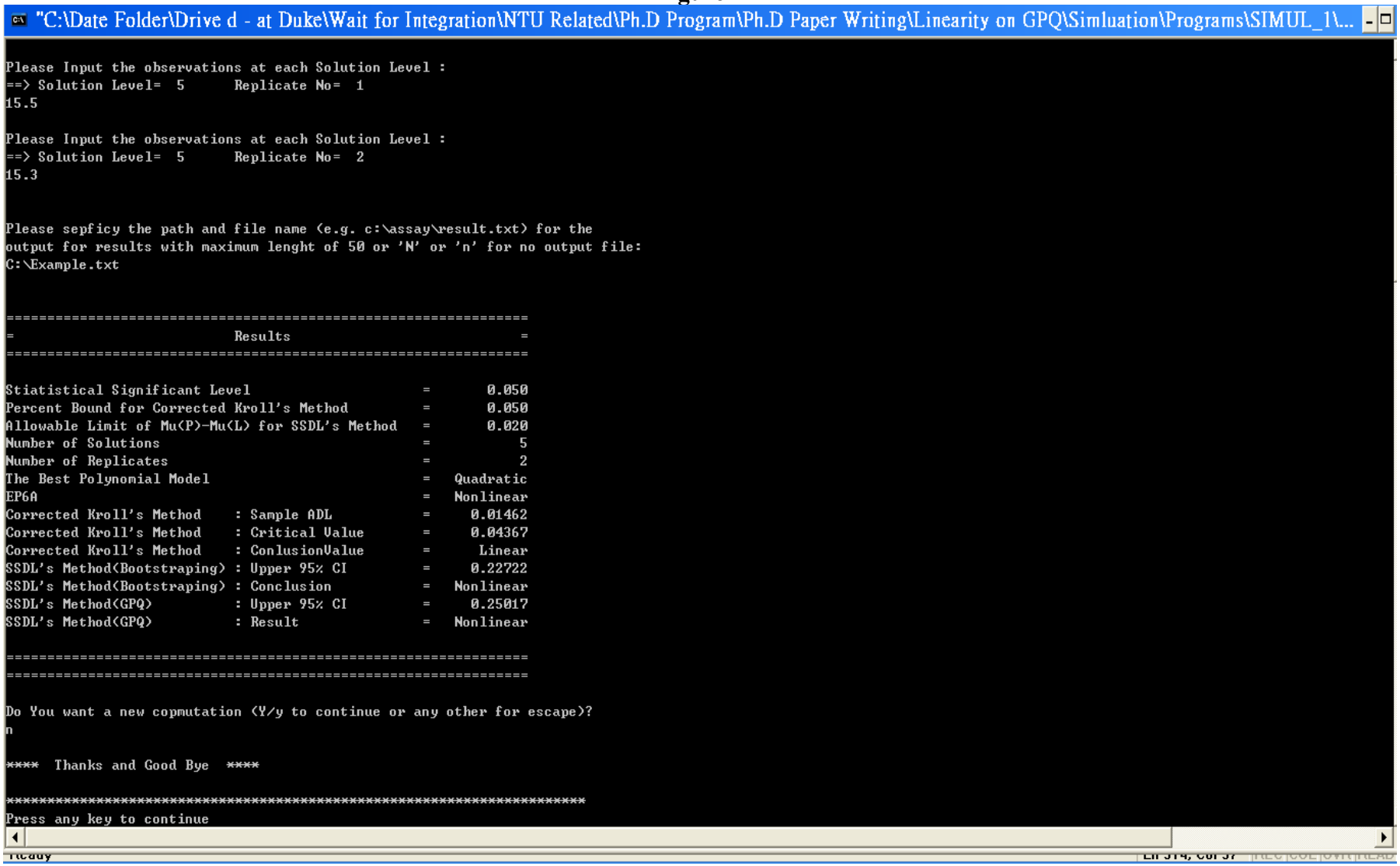

"C:\Date Folder\Drive d - at Duke\Wait for Integration\NTU Related\Ph.D Program\Ph.D Paper Writing\Linearity on GPQ\Simluation\Programs\SIMUL_1\...
*** Welcome the Application for Linearity Evaluation of Assay Validation ***
Please Input the significant level of statistical Testing :
0.05
Please Input the allowable Limit of Linearity :
0.02
Please Input the Number of Solutions of Your Experiment :
5
Please Input the Number of Replicates of Your Experiment :
2
Please Input the observations at each Solution Level :
==> Solution Level= 1 Replicate No= 1
4.7
Please Input the observations at each Solution Level :
==> Solution Level= 1 Replicate No= 2
4.6
Please Input the observations at each Solution Level :
==> Solution Level= 2 Replicate No= 1
7.8
Please Input the observations at each Solution Level :
==> Solution Level= 2 Replicate No= 2
7.6
Please Input the observations at each Solution Level :
==> Solution Level= 3 Replicate No= 1
10.4
Please Input the observations at each Solution Level :
==> Solution Level= 3 Replicate No= 2
10.2
Please Input the observations at each Solution Level :
==> Solution Level= 4 Replicate No= 1
13.0

Figure 2


"C:\Date Folder\Drive d - at Duke\Wait for Integration\NTU Related\Ph.D Program\Ph.D Paper Writing\Linearity on GPQ\Simluation\Programs\SIMUL_1\...
Please Input the observations at each Solution Level :
==> Solution Level= 5 Replicate No= 1
15.5
Please Input the observations at each Solution Level :
==> Solution Level= 5 Replicate No= 2
15.3
Please sepficy the path and file name (e.g. c:\assay\result.txt) for the
output for results with maximum lenght of 50 or 'N' or 'n' for no output file:
C:\Example.txt
= Results =
Stiatistical Significant Level = 0.050
Percent Bound for Corrected Kroll's Method = 0.050
Allowable Limit of Mu(P)-Mu(L) for SSDL's Method = 0.020
Number of Solutions = 5
Number of Replicates = 2
The Best Polynomial Model = Quadratic
EP6A = Nonlinear
Corrected Kroll's Method : Sample ADL = 0.01462
Corrected Kroll's Method : Critical Value = 0.04367
Corrected Kroll's Method : ConlusionValue = Linear
SSDL's Method(Bootstraping) : Upper 95% CI = 0.22722
SSDL's Method(Bootstraping) : Conclusion = Nonlinear
SSDL's Method(GPQ) : Upper 95% CI = 0.25017
SSDL's Method(GPQ) : Result = Nonlinear
Do You want a new copmutation (Y/y to continue or any other for escape)?
n
**** Thanks and Good Bye ****
Press any key to continue

Sample Output in Example.txt

```
================================================================================
=     Statistical Methods for Evaluating the Linearity in Assay Validation     =
=                                                                              =
=                  Eric Hsieh1 and Chin-fu Hsiao, Jen-pei Liu (2008)           =
=                              Jounral of Chemometrics                         =
================================================================================
```

=== Measurement ===

Solution	Replicate	Result
1	1	4.70000
1	2	4.60000
2	1	7.80000
2	2	7.60000
3	1	10.40000
3	2	10.20000
4	1	13.00000
4	2	13.10000
5	1	15.50000
5	2	15.30000

=== Summary of Regression Analysis ===

Order	Coefficient	Value	SE	t-test	Signi-ficant	Std err Sy.x	D.F.
Linear	b0	2.16500	0.15097	14.341	*		
	b1	2.68500	0.04552	58.988	*	0.20356	8
Quadratic	b0	1.54000	0.18863	8.164	*		
	b1	3.22071	0.14375	22.406	*		
	b2	-0.08929	0.02350	-3.799	*	0.12438	7
Cubic	b0	1.47000	0.46623	3.153	*		
	b1	3.31905	0.60943	5.446	*		
	b2	-0.12679	0.22620	-0.561			
	b3	0.00417	0.02498	0.167		0.13403	6

=== Mean Differences between the Best-fitted Curve and Simple Linear Regression Equation ===

Result Mean	Predicted (Linear)	Predicted (Quadratic)	Difference	%Difference
4.65000	4.85000	4.67143	-0.17857	-3.82263
7.70000	7.53500	7.62429	0.08929	1.17107
10.30000	10.22000	10.39857	0.17857	1.71727
13.05000	12.90500	12.99429	0.08929	0.68712
15.40000	15.59000	15.41143	-0.17857	-1.15869

Results		
Stiatistical Significant Level	=	0.050
Percent Bound for Corrected Kroll's Method	=	0.050
Allowable Limit of Mu(P)-Mu(L) for SSDL's Method	=	0.020
Number of Solutions	=	5
Number of Replicates	=	2
The Best Polynomial Model	=	Quadratic
EP6A	=	Nonlinear
Corrected Kroll's Method : Sample ADL	=	0.01462
Corrected Kroll's Method : Critical Value	=	0.04367
Corrected Kroll's Method : ConlusionValue	=	Linear
SSDL's Method(Bootstraping) : Upper 95% CI	=	0.22722
SSDL's Method(Bootstraping) : Conclusion	=	Nonlinear
SSDL's Method(GPQ) : Upper 95% CI	=	0.25017
SSDL's Method(GPQ) : Result	=	Nonlinear

```
!=============================================================================
!=     Statistical Methods for Evaluating the Linearity in Assay Validation   =
!=                                                                            =
!=                  Eric Hsieh1 and Chin-fu Hsiao, Jen-pei Liu (2008)         =
!=                          Jounral of Chemometrics                           =
!=============================================================================


program LinearEval
      use IMSL
      USE DFPORT
      implicit none

      character*8 char_time

!Parameter Setting
      integer, parameter :: nomonte=10000, noboot=3000 ! Simulatio time for Monte-Carlo and Bootstrap

!Statistics
      real(kind=8), allocatable :: mu_p_hat(:), mu_l_hat(:), diffpl(:) !Working vectors of estimates of Mu(Pi),Mu(Li) and their difference
      real(kind=8) :: diffsrm !Sum of square of Mu(Pi)-Mu(Li)
      real(kind=8) :: ADL,crikroll,parmp2,cgpq !Sample Statistic
      real(kind=8) :: Mu_y !Mean of y
      real(kind=8) :: MSEP !MSE of the models i.e. estimate of Sigma^2
      real(kind=8), allocatable :: presid(:), lresid(:) !residuals    for polynominal model by original LSE and bootstrap sampling,respectively

!Data
      real(kind=8), allocatable :: obs_y(:) !Working vector of y

!Result
```

```
      character(len=12) :: EP6A,Cr_Kroll,Pbootstp2,GPQ

!Working Variable
      integer :: lMonte, lrep, lsol, lboot, lgpq      !loop indexes
      integer :: noobs_y !Working number of observation of y
      real(kind=8), allocatable :: betaw(:) !Working vector for model coefficient
      integer :: error=0 !Status check for memory allocation
      real(kind=8), allocatable :: pbdiffsrm(:), psbdiffsrm(:)!Arrays conating sum of square of Mu(Pi)-Mu(Li) generated by bootstrap
sampling before and after sorting
      integer :: i,j,k,m !Temporary working variables
      real(kind=8), allocatable :: GZ(:) !Standard normal random sample for GPQ
      real(kind=8) :: GU(1) !Chi-square random sample with df LR-d-1 for GPQ
      real(kind=8), allocatable :: Gdiffpl2(:), SGdiffpl2(:) !Arrays conating GPQ of (Mu(Pi)-Mu(Li))^2 generated by GPQ method before
and after sorting
      real(kind=8), allocatable :: GSIGMA12(:,:) !Sigma^1/2 generated by Spectral Decomposition Method for GPQ sampling
      real(kind=8), allocatable :: RSIG(:,:) !Upper triangular matrix decomposed by Cholesky factorization
      integer    :: IRANK !Returned parameter for
      real(kind=8), allocatable :: Mnormpl(:,:) !Random sample generated by Multinormal distribution of Mu(Pi)-Mu(Li)
      real(kind=8), parameter :: TOL=2.220446049250313E-14 !Tolerance limit for DCHFAC function of IMSL
      character(len=1) :: Start
      real(kind=8) :: delta, alpha
      integer :: nosol, norep
      real(kind=8), allocatable :: dm_d1(:,:),dm_d2(:,:),dm_d3(:,:),dm_dp(:,:) !working design matrix for model
      real(kind=8), allocatable :: WP(:,:),WL(:,:),WP2(:,:), WP3(:,:) !Working information matrix for polynomial model
      real(kind=8), allocatable :: mu_P2_hat(:),mu_P3_hat(:),p2resid(:),p3resid(:) !Working predict vector
      real(kind=8), allocatable :: W(:,:),IW(:,:) !Difference of working information matrixs between 1st and polynomial models
      real(kind=8) :: IT_WL(2,2),IT_WP2(3,3),IT_WP3(4,4)
      real(kind=8) :: betaL_hat(2),betaP2_hat(3),betaP3_hat(4),betaL_SE(2),betaP2_SE(3),betaP3_SE(4)
      real(kind=8) :: betaL_tv(2),betaP2_tv(3),betaP3_tv(4)
      real(kind=8) :: LSEYX,P2SEYX,P3SEYX,Ldf,P2df,P3df, MSEL,MSEP2, MSEP3,porder
      character(len=1) :: Lsig1,Lsig2,P2sig1,P2sig2,P2sig3,P3sig1,P3sig2,P3sig3,P3sig4
```

```
    character(len=50) :: flname


!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!! Start of Code !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!! Initial Setting !!!
allocate(pbdiffsrm(noboot),stat=error)
allocate(psbdiffsrm(noboot),stat=error)
allocate(Gdiffpl2(nomonte),stat=error)
allocate(SGdiffpl2(nomonte),stat=error)
Start="Y"
do while (start=="Y" .OR. start=="y")
  write(*,"(A76)") "*** Welcome the Application for Linearity Evaluation of Assay Validation ***"
  write(*,*) " "
  write(*,"(A76)") "Please Input the significant level of statistical Testing :                  "
  read(*,*) alpha
  write(*,*) " "
  write(*,"(A76)") "Please Input the allowable Limit of Linearity :                              "
  read(*,*) delta
  write(*,*) " "
  write(*,"(A76)") "Please Input the Number of Solutions of Your Experiment :                    "
  read(*,*) nosol
  write(*,*) " "
  write(*,"(A76)") "Please Input the Number of Replicates of Your Experiment :                   "
  read(*,*) norep
  write(*,*) " "
  noobs_y=nosol*norep !number of observation of y
  allocate(dm_d1(noobs_y,2),stat=error)
  allocate(dm_dp(noobs_y,4),stat=error)
  allocate(dm_d2(noobs_y,3),stat=error)
  allocate(dm_d3(noobs_y,4),stat=error)
```

```
allocate(obs_y(noobs_y),stat=error)
allocate(presid(noobs_y),stat=error)
allocate(p2resid(noobs_y),stat=error)
allocate(p3resid(noobs_y),stat=error)
allocate(lresid(noobs_y),stat=error)
allocate(mu_p_hat(noobs_y),stat=error)
allocate(mu_p2_hat(noobs_y),stat=error)
allocate(mu_p3_hat(noobs_y),stat=error)
allocate(mu_l_hat(noobs_y),stat=error)
allocate(WP(noobs_y,noobs_y),stat=error)
allocate(WP2(noobs_y,noobs_y),stat=error)
allocate(WP3(noobs_y,noobs_y),stat=error)
allocate(WL(noobs_y,noobs_y),stat=error)
allocate(IW(noobs_y,noobs_y),stat=error)
allocate(diffpl(noobs_y),stat=error)
allocate(RSIG(noobs_y,noobs_y),stat=error)
allocate(Mnormpl(noboot,noobs_y),stat=error)
allocate(GZ(noobs_y),stat=error)
allocate(GSIGMA12(noobs_y,noobs_y),stat=error)

do lsol=1,nosol
   do lrep=1,norep
      write(*,"(A76)") "Please Input the observations at each Solution Level :                                    "
      write(*,"(A19,I3,A19,I3)") "==> Solution Level=", lsol, "          Replicate No=", lrep
      read(*,*) obs_y((lsol-1)*norep+lrep)
      write(*,*) ""
   end do !lrep
end do !lsol

do lsol=1,nosol
   do lrep=1,norep
```

```
      dm_d1((lsol-1)*norep+lrep,1)=1
      dm_d1((lsol-1)*norep+lrep,2)=lsol
      dm_dp((lsol-1)*norep+lrep,1)=1
      dm_dp((lsol-1)*norep+lrep,2)=lsol
      dm_dp((lsol-1)*norep+lrep,3)=lsol*lsol
      dm_dp((lsol-1)*norep+lrep,4)=lsol*lsol*lsol
   end do !lrep
end do !lsol

dm_d2=dm_dp(:,1:3)
dm_d3=dm_dp
IT_WL= .i. (dm_d1 .tx. dm_d1)
IT_WP2=.i. (dm_d2 .tx. dm_d2)
IT_WP3=.i. (dm_d3 .tx. dm_d3)
WL=dm_d1 .x. IT_WL .xt. dm_d1
WP2=dm_d2 .x. IT_WP2 .xt. dm_d2
WP3=dm_d3 .x. IT_WP3 .xt. dm_d3
mu_L_hat=WL .x. obs_y !Compute the estimate of Mu(Li)
mu_P2_hat=WP2 .x. obs_y !Compute the estimate of Mu(Pi)
mu_P3_hat=WP3 .x. obs_y !Compute the estimate of Mu(Pi)
lresid=obs_y-mu_L_hat !Residuals for linear model
p2resid=obs_y-mu_p2_hat !Residuals for polynomial model
p3resid=obs_y-mu_p3_hat !Residuals for polynomial model
MSEL=norm(Lresid,2)**2/(noobs_y-1.0-1.0) !MSE of polynominal model i.e. estimate of Sigma^2
MSEP2=norm(p2resid,2)**2/(noobs_y-2.0-1.0) !MSE of polynominal model i.e. estimate of Sigma^2
MSEP3=norm(p3resid,2)**2/(noobs_y-3.0-1.0) !MSE of polynominal model i.e. estimate of Sigma^2
IT_WL= .i. (dm_d1 .tx. dm_d1)
IT_WP2=.i. (dm_d2 .tx. dm_d2)
IT_WP3=.i. (dm_d3 .tx. dm_d3)
betaL_hat= (IT_WL .xt. dm_d1) .x. obs_y
betaP2_hat= (IT_WP2 .xt. dm_d2) .x. obs_y
```

```
betaP3_hat= (IT_WP3 .xt. dm_d3) .x. obs_y
do i=1,2
   betaL_SE(i)= dsqrt(IT_WL(i,i)*MSEL)
end do
do i=1,3
   betaP2_SE(i)= dsqrt(IT_WP2(i,i)*MSEP2)
end do
do i=1,4
   betaP3_SE(i)= dsqrt(IT_WP3(i,i)*MSEP3)
end do
betaL_tv=betaL_hat/betaL_SE
betaP2_tv=betaP2_hat/betaP2_SE
betaP3_tv=betaP3_hat/betaP3_SE
LSEYX=dsqrt(MSEL)
P2SEYX=dsqrt(MSEP2)
P3SEYX=dsqrt(MSEP3)
Ldf=noobs_y-1.0-1.0
P2df=noobs_y-2.0-1.0
P3df=noobs_y-3.0-1.0
if (dabs(betaL_tv(1)) >dtin(1.0-alpha/2.0,Ldf)) then
   Lsig1="*"
else
   Lsig1=" "
end if
if (dabs(betaL_tv(2)) > dtin(1.0-alpha/2.0,Ldf)) then
   Lsig2="*"
else
   Lsig2=" "
end if
if (dabs(betaP2_tv(1)) > dtin(1.0-alpha/2.0,P2df)) then
   P2sig1="*"
```

```
else
   P2sig1=" "
end if
if (dabs(betaP2_tv(2)) > dtin(1.0-alpha/2.0,P2df)) then
   P2sig2="*"
else
   P2sig2=" "
end if
if (dabs(betaP2_tv(3)) > dtin(1.0-alpha/2.0,P2df)) then
   P2sig3="*"
else
   P2sig3=" "
end if
if (dabs(betaP3_tv(1)) > dtin(1.0-alpha/2.0,P3df)) then
   P3sig1="*"
else
   P3sig1=" "
end if
if (dabs(betaP3_tv(2)) > dtin(1.0-alpha/2.0,P3df)) then
   P3sig2="*"
else
   P3sig2=" "
end if
if (dabs(betaP3_tv(3)) > dtin(1.0-alpha/2.0,P3df)) then
   P3sig3="*"
else
   P3sig3=" "
end if
if (dabs(betaP3_tv(4)) > dtin(1.0-alpha/2.0,P3df)) then
   P3sig4="*"
else
```

```
    P3sig4=" "
  end if
  porder=0.0

  if ((P3Sig4=="*") .AND. (P3SEYX <= P2SEYX)) then
    porder=3.0
    WP=WP3
    mu_p_hat=mu_p3_hat
    MSEP=MSEP3
  else if ((P2Sig3=="*") .AND. (P2SEYX <= LSEYX)) then
    porder=2.0
    WP=WP2
    mu_p_hat=mu_p2_hat
    MSEP=MSEP2
  else if (LSig2=="*") then
    porder=1.0
  else
    porder=0.0
  end if


  IW=(WP-WL) .xt. (WP-WL)
  diffpl=mu_p_hat-mu_l_hat !Compute the Vector of Mu(Pi)-Mu(Li)
  diffsrm=norm(diffpl,2)**2 !Compute sum of square for Mu(Pi)-Mu(Li)
  mu_y=sum(obs_y)/noobs_y !Mean of y

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

  if (porder == 0.0) then
    write(*,"(A1)")
    write(*,"(A1)")
```

```
   write(*,"(A64)")          "================================================================="
   write(*,"(A64)")          "=                        Results                                ="
   write(*,"(A64)")          "================================================================="
   write(*,*) " "
   write(*,"(A64)")          "Neither one of linear, quadratic or cubic model fitted          "
   write(*,*) " "
   write(*,"(A64)")          "================================================================="
   write(*,"(A64)")          "================================================================="
else if (porder == 1.0) then
   write(*,"(A1)")
   write(*,"(A1)")
   write(*,"(A64)")          "================================================================="
   write(*,"(A64)")          "=                        Results                                ="
   write(*,"(A64)")          "================================================================="
   write(*,"(A1)")
   write(*,"(A64)")          "The linear model has already been the best model.               "
   write(*,"(A1)")
   write(*,"(A64)")          "================================================================="
   write(*,"(A64)")          "================================================================="
else
   !!!!!!!!!!!!    EP6A method !!!!!!!!!!!!!!!
   if (maxval(abs(diffpl)) < delta) then
      EP6A="          Linear"
   else
      EP6A="      Nonlinear"
   end if

   !!!!!!!!!!!! Corrected- Kroll method !!!!!!!!!!!!!!!
   ADL=dsqrt(diffsrm/noobs_y)/(mu_y)
   CriKroll=(dsqrt(MSEP)/mu_y)*dsqrt(DCSNIN(alpha, porder-1.0, noobs_y*(0.05**2)/(MSEP/(mu_y**2)))/noobs_y)
   if ( ADL < crikroll ) then
```

```
   Cr_Kroll="           Linear"
else
   Cr_Kroll="      Nonlinear"
end if

!!!!!!!!!!!!      Parametric Bootstratp      !!!!!!!!!!!!!!
CALL DCHFAC(noobs_y, MSEP*IW, noobs_y, TOL, IRANK, RSIG, noobs_y)
CALL DRNMVN(noboot, noobs_y, RSIG, noobs_y, Mnormpl, noboot)
do i=1,noboot
   pbdiffsrm(i)= norm((diffpl + Mnormpl(i,:)),2)**2
end do
call DSVRGN(noboot, pbdiffsrm , psbdiffsrm)
parmp2=psbdiffsrm(noboot+1- floor(alpha*noboot))/norep
if ( psbdiffsrm(noboot+1- floor(alpha*noboot)) < (noobs_y*(delta**2)) ) then
   Pbootstp2="           Linear"
else
   Pbootstp2="      Nonlinear"
end if

!!!!!!!!!!!!      GPQ method      !!!!!!!!!!!!!
do lgpq=1,nomonte
   CALL DRNNOR (noobs_y, GZ)
   call DRNCHI(1,noobs_y-porder-1.0, GU)
   call CPSIGMA(((noobs_y-porder-1.0)*MSEP/GU(1))*IW, noobs_y, GSIGMA12)
   Gdiffpl2(lgpq)=norm(diffpl-(GSIGMA12 .x. GZ),2)**2 !/norep
end do
call DSVRGN(nomonte,Gdiffpl2,sGdiffpl2)
CGPQ=sGdiffpl2(nomonte - floor(alpha*nomonte))/norep
if ( sGdiffpl2(nomonte - floor(alpha*nomonte)) < (noobs_y*(delta**2)) ) then
   GPQ="           Linear"
else
```

```
   GPQ="      Nonlinear"
end if
write(*,"(A1)") " "
write(*,"(A78)") "Please sepficy the path and file name (e.g. c:\assay\result.txt) for the           "
write(*,"(A78)") "output for results with maximum lenght of 50 or 'N' or 'n' for no output file:"
read(*,*) flname
if ((trim(flname) .NE. "n") .OR. (trim(flname) .NE. "N")) then
   OPEN(unit = 111, file = flname, status = "replace", action ="write")
end if

write(111,"(A1)") " "
write(111,"(A76)") "=========================================================================================="
write(111,"(A76)") "=     Statistical Methods for Evaluating the Linearity in Assay Validation     ="
write(111,"(A76)") "=                                                                              ="
write(111,"(A76)") "=                  Eric Hsieh1 and Chin-fu Hsiao, Jen-pei Liu (2008)           ="
write(111,"(A76)") "=                        Jounral of Chemometrics                               ="
write(111,"(A76)") "=========================================================================================="
write(111,"(A1)") " "
write(111,"(A1)") " "
write(111,"(A1)") " "
write(111,"(A32)") "=== Measurement ===                "
write(111,"(A1)") " "
write(111,"(A46)") "      Solution          Replcaite          Result       "
write(111,"(A46)") "----------------------------------------------"
write(111,"(A1)") " "
do lsol=1,nosol
   do lrep=1,norep
      write(111,"(A5,I5,A10,I5,A6,F12.5)") "        ",lsol,"              ",lrep,"          ",obs_y((lsol-1)*norep+lrep)
   end do
end do
write(111,"(A1)") " "
```

```
write(111,"(A46)")    "----------------------------------------------"
write(111,"(A1)") " "
write(111,"(A1)") " "
write(111,"(A1)") " "
write(111,"(A57)") "=== Summary of Regression Analysis ===                              "
write(111,"(A1)") " "
write(111,"(A106)") "      Order      Coefficient            Value                  SE                 t-test           Signi-              Std err
D.F."
write(111,"(A106)") "                                                                                                                        ficant
Sy.x                 "
write(111,"(A106)") "----------------------------------------------------------------------------------------------------------"
write(111,"(A1)") " "
write(111,"(A26,F12.5,F15.5,F14.3,A9,A1)") "       Linear              b0             ",betaL_hat(1),betaL_SE(1),betaL_tv(1),"
",Lsig1
write(111,"(A26,F12.5,F15.5,F14.3,A9,A1,F18.5,A4,I5)") "                           b1             ",betaL_hat(2),betaL_SE(2),betaL_tv(2),"
",Lsig2,LSEYX,"       ",ceiling(Ldf)
write(111,"(A1)") " "
write(111,"(A26,F12.5,F15.5,F14.3,A9,A1)") "    Quadratic          b0             ",betaP2_hat(1),betaP2_SE(1),betaP2_tv(1),"
",P2sig1
write(111,"(A26,F12.5,F15.5,F14.3,A9,A1)") "                           b1             ",betaP2_hat(2),betaP2_SE(2),betaP2_tv(2),"
",P2sig2
write(111,"(A26,F12.5,F15.5,F14.3,A9,A1,F18.5,A4,I5)") "                           b2
",betaP2_hat(3),betaP2_SE(3),betaP2_tv(3),"               ",P2sig3,P2SEYX,"       ",ceiling(P2df)
write(111,"(A1)") " "
write(111,"(A26,F12.5,F15.5,F14.3,A9,A1)") "       Cubic               b0             ",betaP3_hat(1),betaP3_SE(1),betaP3_tv(1),"
",P3sig1
write(111,"(A26,F12.5,F15.5,F14.3,A9,A1)") "                           b1             ",betaP3_hat(2),betaP3_SE(2),betaP3_tv(2),"
",P3sig2
write(111,"(A26,F12.5,F15.5,F14.3,A9,A1)") "                           b2             ",betaP3_hat(3),betaP3_SE(3),betaP3_tv(3),"
",P3sig3
write(111,"(A26,F12.5,F15.5,F14.3,A9,A1,F18.5,A4,I5)") "                           b3
```

```
",betaP3_hat(4),betaP3_SE(4),betaP3_tv(4)," ",P3sig4,P3SEYX," ",ceiling(P3df)
      write(111,"(A1)") " "
      write(111,"(A105")    "----------------------------------------------------------------------------------------------------"
      write(111,"(A1)") " "
      write(111,"(A1)") " "
      write(111,"(A1)") " "
      write(111,"(A92)") "=== Mean Differences between the Best-fitted Curve and Simple Linear Regression Equation ==="
      write(111,"(A1)") " "
      write(111,"(A102)")     "       Result                Predicted                Predicted                Difference
%Difference       "
      if (porder==2.0) then
         write(111,"(A102)") "         Mean                  (Linear)                 (Quadratic)
"
      else if (porder==3.0) then
         write(111,"(A102)") "         Mean                  (Linear)                 (Cubic)
"
      end if
      write(111,"(A102)")     "----------------------------------------------------------------------------------------------------"
      write(111,"(A1)") " "
      do lsol=1,nosol
         if (porder==2.0) then
            write(111,"(F11.5,F20.5,F21.5,F23.5,F23.5)"),sum(obs_y((lsol-1)*norep+1:(lsol-1)*norep+norep))/norep/1.0,
mu_l_hat((lsol-1)*norep+1),&
                                                  mu_p2_hat((lsol-1)*norep+1),mu_p2_hat((lsol-1)*norep+1)-mu_l_hat((lsol-1)*norep+1),&

      (mu_p2_hat((lsol-1)*norep+1)-mu_l_hat((lsol-1)*norep+1))/mu_p2_hat((lsol-1)*norep+1)*100
         else if (porder==3.0) then
            write(111,"(F11.5,F20.5,F21.5,F23.5,F23.5)"),sum(obs_y((lsol-1)*norep+1:(lsol-1)*norep+norep))/norep/1.0,
mu_l_hat((lsol-1)*norep+1),&
                                                  mu_p3_hat((lsol-1)*norep+1),mu_p3_hat((lsol-1)*norep+1)-mu_l_hat((lsol-1)*norep+1),&
```

```
(mu_p3_hat((lsol-1)*norep+1)-mu_l_hat((lsol-1)*norep+1))/mu_p3_hat((lsol-1)*norep+1)*100
   end if
end do
write(111,"(A1)") " "
write(111,"(A102)")     "----------------------------------------------------------------------------------------------------"
write(111,"(A1)") " "
write(111,"(A1)") " "
write(111,"(A1)") " "
write(111,"(A64)")              "================================================================"
write(111,"(A64)")              "=                          Results                             ="
write(111,"(A64)")              "================================================================"
write(111,"(A1)")
write(111,"(A52,F12.3)")      "Stiatistical Significant Level                    =", alpha
write(111,"(A52,F12.3)")      "Percent Bound for Corrected Kroll's Method         =", 0.05
write(111,"(A52,F12.3)")      "Allowable Limit of Mu(P)-Mu(L) for SSDL's Method  =", delta
write(111,"(A52,I12)")        "Number of Solutions                               =", nosol
write(111,"(A52,I12)")        "Number of Replicates                              =", norep
if (porder==2.0) then
   write(111,"(A64)")          "The Best Polynomial Model                          =     Quadratic"
else if (porder==3.0) then
   write(111,"(A64)")          "The Best Polynomial Model                          =         Cubic"
end if
write(111,"(A52,A12)")        "EP6A                                                 =", EP6A
write(111,"(A52,F12.5)")      "Corrected Kroll's Method       : Sample ADL       =", ADL
write(111,"(A52,F12.5)")      "Corrected Kroll's Method       : Critical Value   =", Crikroll
write(111,"(A52,A12)")        "Corrected Kroll's Method       : ConlusionValue   =", Cr_Kroll
write(111,"(A52,F12.5)")      "SSDL's Method(Bootstraping) : Upper 95% CI        =", parmp2
write(111,"(A52,A12)")        "SSDL's Method(Bootstraping) : Conclusion          =", Pbootstp2
write(111,"(A52,F12.5)")      "SSDL's Method(GPQ)             : Upper 95% CI     =", cgpq
write(111,"(A52,A12)")        "SSDL's Method(GPQ)             : Result           =", GPQ
write(111,"(A1)")
```

```
write(111,"(A64)")          "================================================================"
write(111,"(A64)")          "================================================================"


!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
write(*,"(A1)")
write(*,"(A1)")
write(*,"(A64)")          "================================================================"
write(*,"(A64)")          "=                          Results                             ="
write(*,"(A64)")          "================================================================"
write(*,"(A1)")
write(*,"(A52,F12.3)")    "Stiatistical Significant Level                      =", alpha
write(*,"(A52,F12.3)")    "Percent Bound for Corrected Kroll's Method          =", 0.05
write(*,"(A52,F12.3)")    "Allowable Limit of Mu(P)-Mu(L) for SSDL's Method    =", delta
write(*,"(A52,I12)")      "Number of Solutions                                 =", nosol
write(*,"(A52,I12)")      "Number of Replicates                                =", norep
if (porder==2.0) then
   write(*,"(A64)")       "The Best Polynomial Model                           =     Quadratic"
else if (porder==3.0) then
   write(*,"(A64)")       "The Best Polynomial Model                           =         Cubic"
end if
write(*,"(A52,A12)")      "EP6A                                                =", EP6A
write(*,"(A52,F12.5)")    "Corrected Kroll's Method      : Sample ADL          =", ADL
write(*,"(A52,F12.5)")    "Corrected Kroll's Method      : Critical Value      =", Crikroll
write(*,"(A52,A12)")      "Corrected Kroll's Method      : ConlusionValue      =", Cr_Kroll
write(*,"(A52,F12.5)")    "SSDL's Method(Bootstraping) : Upper 95% CI          =", parmp2
write(*,"(A52,A12)")      "SSDL's Method(Bootstraping) : Conclusion            =", Pbootstp2
write(*,"(A52,F12.5)")    "SSDL's Method(GPQ)            : Upper 95% CI        =", cgpq
write(*,"(A52,A12)")      "SSDL's Method(GPQ)            : Result              =", GPQ
write(*,"(A1)")
write(*,"(A64)")          "================================================================"
```

```
      write(*,"(A64)")                "================================================================="
end if

   deallocate(dm_d1)
   deallocate(dm_dp)
   deallocate(dm_d2)
   deallocate(dm_d3)
   deallocate(obs_y)
   deallocate(presid)
   deallocate(p2resid)
   deallocate(p3resid)
   deallocate(lresid)
   deallocate(mu_p_hat)
   deallocate(mu_p2_hat)
   deallocate(mu_p3_hat)
   deallocate(mu_l_hat)
   deallocate(WP)
   deallocate(WP2)
   deallocate(WP3)
   deallocate(WL)
   deallocate(IW)
   deallocate(diffpl)
   deallocate(RSIG)
   deallocate(Mnormpl)
   deallocate(GZ)
   deallocate(GSIGMA12)
      write(*,"(A1)")
   Write(*,"(A72)") "Do You want a new copmutation (Y/y to continue or any other for escape)?"
   read(*,*) Start
   if (Start .NE. "Y" .AND. Start .NE. "y") then
      write(*,"(A1)")
```

```
      write(*,"(A31)") "****    Thanks and Good Bye    ****"
      write(*,"(A1)") " "
      write(*,"(A71)") "***********************************************************************"
   else
      write(*,"(A1)")
      write(*,*) ""
   end if
end do

deallocate(pbdiffsrm)
deallocate(psbdiffsrm)
deallocate(Gdiffpl2)
deallocate(SGdiffpl2)
write(111,*) " "
call TIME(char_time)
stop
end

subroutine CPSIGMA(SW,SDIMEN,CGSIGMA12)
   use IMSL
   implicit none
   integer I,SDIMEN
   REAL(kind=8)    SW(SDIMEN,SDIMEN),B(SDIMEN,SDIMEN), eval(SDIMEN),
evec(SDIMEN,SDIMEN),K(SDIMEN+1,SDIMEN),X(SDIMEN+1,SDIMEN),E(SDIMEN,SDIMEN),LAMDA12(SDIMEN,SDIMEN),SI
GMA12(SDIMEN,SDIMEN),CGSIGMA12(SDIMEN,SDIMEN)
   INTEGER IRANK
   INTEGER, PARAMETER ::NKEY=1
   INTEGER NCX,NRX,LDX
   INTEGER       ICOMP, INDKEY(NKEY), IORDR, IRET, NGROUP
   DATA          INDKEY/1/
   integer, allocatable :: IPERM(:), NI(:)
```

```
    integer :: error=0

    NCX=SDIMEN
    NRX=SDIMEN+1
    LDX=NRX
    allocate(IPERM(NCX),stat=error)
    allocate(NI(NCX),stat=error)
    B=SW
    CALL DEVCSF(SDIMEN, B, SDIMEN, EVAL, EVEC, SDIMEN)
    do i=1,SDIMEN
       if (eval(i) .LE. 0.0) then
          eval(i)=0
       end if
    end do
    k(1,:)=eval
    k(2:SDIMEN+1,:)=evec
    X=K
    ICOMP = 0
    IORDR = 1
    IRET    = 0
    CALL DSCOLR (NRX, NCX, X, LDX, ICOMP, IORDR, IRET, NKEY, INDKEY,IPERM, NGROUP, NI)
    E=X(2:SDIMEN,:)
    LAMDA12=0
    do I=1,SDIMEN
       LAMDA12(i,i)=dsqrt(X(1,i))
    end do
    SIGMA12=E .x. LAMDA12
    CGSIGMA12=SIGMA12 .xt. E
    deallocate(IPERM)
    deallocate(NI)
end
```