การพัฒนาตัวแบบการทำนายระยะเวลาหมดอายุของยา

นางสาวนัจชลี ศรีมณีกาญจน์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2554

A Development of Drug Expiration Prediction Model

Miss Natchalee Srimaneekarn

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Statistics
Department of Statistics
Faculty of Commerce and Accountancy
Chulalongkorn University
Academic Year 2011

Thesis Title	A Development of Drug Expiration Prediction Model
By	Miss Natchalee Srimaneekarn
Field of Study	Statistics
Thesis Advisor	Assistant Professor Seksan Kiatsupaibul, Ph.D.

Accepted by the Faculty of Commerce and Accountancy, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree of Science Program in Statistics

…….…… Dean of Faculty of Commerce and Accountancy
(Associate Professor Pasu Decharin, Ph.D.)

THESIS COMMITTEE

…….…… Chairman
(Associate Professor Kanlaya Vanichbuncha, Ph.D.)

…….…… Thesis Advisor
(Assistant Professor Seksan Kiatsupaibul, Ph.D.)

…….…… Examiner
(Akarin Phaibulpanich, Ph.D.)

…….…… External Examiner
(Professor Anthony Hayter, Ph.D.)

นัจชลี ศรีมณีกาญจน์: การพัฒนาตัวแบบการทำนายระยะเวลาหมดอายุของยา.
(A DEVELOPMENT OF DRUG EXPIRATION PREDICTION MODEL)
อ. ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ.ดร.เสกสรร เกียรติสุไพบูลย์, 46 หน้า.

ผลิตภัณฑ์ทางการแพทย์ทุกชนิดจำเป็นต้องระบุวันหมดอายุบนฉลากผลิตภัณฑ์ซึ่งรวมทั้งยาด้วย วันหมดอายุนั้นมาจากระยะเวลาหมดอายุซึ่งได้มาจากการคำนวณจากตัวแบบทางสถิติ อย่างไรก็ตามวิธีการคำนวณหาระยะเวลาหมดอายุของยาที่แน่นอนยังไม่สามารถระบุได้ชัดเจนว่าควรใช้วิธีใด การศึกษานี้จึงเสนอตัวแบบใหม่สำหรับการทำนายระยะเวลาหมดอายุของยาโดยใช้ขอบเขตล่างของการทำนาย ซึ่งมีสัดส่วนค่าความแปรปรวนของยาแต่ละบรรจุภัณฑ์ที่เหมาะสม จากผลการศึกษาพบว่าเมื่อสัดส่วนค่าความแปรปรวนของยาแต่ละบรรจุภัณฑ์เพิ่มจะทำให้ตัวแบบการทำนายระยะเวลาหมดอายุของยาโดยใช้ขอบเขตล่างของความเชื่อมั่นมีความแม่นยำในการทำนายลดลง ดังนั้น ตัวแบบขอบเขตล่างของการทำนายที่มีสัดส่วนค่าความแปรปรวนของยาแต่ละบรรจุภัณฑ์ที่เหมาะสม จึงเป็นตัวแบบการทำนายที่เหมาะสมกว่าตัวแบบขอบเขตล่างของความเชื่อมั่นในการใช้ทำนายระยะเวลาหมดอายุของยา

ภาควิชา สถิติ ลายมือชื่อนิสิต .....

สาขาวิชา สถิติ ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก .....

ปีการศึกษา 2554

5381818026: MAJOR STATISTICS
KEYWORDS: EXPIRATION DATE / LINEAR MODEL / VARIANCE COMPONENTS / LOWER PREDICTION INTERVAL

NATCHALEE SRIMANEEKARN: A DEVELOPMENT OF DRUG EXPIRATION PREDICTION MODEL. ADVISOR: ASSIST. PROF. SEKSAN KIATSUPAIBUL, Ph.D., 46 pp.

Every pharmaceutical product is obliged to indicate its expiration date on its package. There are several alternative drug expiration prediction models that provide a sound statistical prediction, but to date there has been no single model that stands out. This study presents a drug expiration prediction model using the lower prediction interval with proper proportion of lot variability. The results illustrate that when the proportion of lot variability increased, the prediction interval model with proper proportion of lot variability provides a more appropriate drug expiration period than that provided by the confidence interval model. Hence, with a proper proportion of lot variability, the lower prediction interval model is recommended over the traditional lower confidence interval model.

Department: Statistics Student's Signature .....
Field of Study: Statistics Advisor's Signature .....
Academic Year: 2011

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Assistant Professor Seksan Kiatsupaibul, for his good advice and guidance since this thesis started until successful. Special thanks to Professor Anthony Hayter for his kindness that he initiated this thesis topic to me. And I also wish to thank the thesis committee, Associate Professor Kanlaya Vanichbuncha and Dr. Akarin Phaibulpanich, for all of comment and good suggestion.

I would also like to thank faculty staff in Department of Anatomy, Faculty of Dentistry, Mahidol University that they gave me a very good opportunity that allowed me to study Statistics at Department of Statistics, Faculty of Commerce and Accountancy, Chulalongkorn University.

Finally, I wish to express my heartiest thanks to my beloved family for their supporting and blessing, and to my friends for their help and encouragement.

CONTENTS

Page

LIST OF TABLES

Page

Page

Page

CHAPTER I

INTRODUCTION

1.1 Background and statement of problem

Every pharmaceutical product is obliged to indicate an expiration date on its packaging, which is estimated using a statistical model. According to the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use guideline (ICH, 2003) currently used in Europe, the United States and Japan, drug expiration study is divided into two parts: determining the shelf life of each batch and assaying whether the data from different batches can be combined for an overall estimation of a single shelf life. A one-sided 95% confidence limit was suggested in the guideline for expiration period determination. Although widely used, the one-sided 95% confidence limit inherently provides an overestimation of the drug expiration. There are several alternative models that provide a more sound statistical prediction, but to date there has been no single model that stands out.

In a previous study, Shao and Chen (1997) estimated drug expiration periods using different packaging in general experiments. The data was collected at 0, 3, 6, 9, 12, and 18 months after manufacture, then a lower predicted bound was used to predict the drug expiration period for comparison with the United States Food and Drug Administration's (U.S. FDA) approach. The results showed that the drug's labeled shelf life under their study were 27 and 26 months for a bottle container and blister package respectively, while the FDA's approach provided a labeled shelf life for a bottle container of 26 months.

Komka et al. (2010) used a tolerance interval for drug expiration period prediction for comparison with the lower 95 % confidence interval which is suggested by the ICH guideline. The results showed a much wider interval in their new approach

than that of the ICH guideline, resulting in a significantly shorter expiration period. They also compared the interval width against different sources of variation, which was narrower when tablet-to-tablet variability was lower compared with the analytical measurement error.

Moreover, Srimaneekarn et al. (2012) studied drug expiration period based on simulated drug strength data. They determined that the suitable proportion value of lot variability in lower prediction interval model was more effective than the lower confidence interval model or the lower prediction interval model. The results showed that different experimental designs and different total variabilities presented very similar simulation results for the expiration period estimation.

According to the ICH guideline (2003), a model for drug expiration prediction is a linear model with a time variable. The aim of this study is to develop a linear model for single batch drug expiration prediction using a prediction interval where the variation comes from two sources: the lot variation (σ_l^2) and the measurement error (σ_m^2). An approach for determining the proportion of lot variation (τ) is proposed.

1.2 Objective

1. To determine the proportion of lot variation (τ)
2. To develop a model for drug expiration prediction for single lot using a prediction interval with the proper proportion of lot variation
3. To investigate the number of drug samples for expiration period prediction

1.3 Area of Study

1. Simulated data is used.
2. The given drug strength (y_{ijk}) is a linear mixed-effects model, which is reduced as time increases:

$$y_{ijk} = \beta_0 + \beta_1 t_i + \alpha_{ij} + \varepsilon_{ijk}\,.$$

   y_{ijk} is the drug strength at time t_i.

   β_0 is the drug strength at manufacturing time, given $\beta_0 = 100\%$.

   β_1 is the fixed constant for fixed effect t_i, given $\beta_1 = -0.5$.

   t_i is the fixed effect time, given $t_i = 0, 3, 6, 9, 12, 18, 24, 36$.

   α_{ij} is the random effect between lots (pill to pill, or bottle to bottle)

   ε_{ijk} is the random error from measurement.
3. The given proportion of lot variation is

$$\tau = \frac{\sigma_l^2}{\sigma^2} = \frac{\sigma_l^2}{\sigma_l^2 + \sigma_m^2},$$

$$\sigma^2 = 1\,, \quad \alpha_{ij} \underset{\sim}{^{iid}} N(0, \sigma_l^2) \quad \text{and} \quad \varepsilon_{ij} \underset{\sim}{^{iid}} N(0, \sigma_m^2).$$

4. Probability that the drug strength is greater or equal to *L* at time *T* is calculated as the criterion.

1.4 Significance of the study

1. Estimating the proportion of lot variation (τ)
2. Developing a proper drug expiration prediction model
3. Investigating the number of drug sample for expiration period prediction

1.5 Methodology

1. Literature review
2. Study 1: Proportion of Lot Variability Determination and Model Development
    a. Simulation of drug strength with given parameters
    b. Estimation of parameters: $\hat{\beta}_0,\ \hat{\beta}_1,\ \hat{\sigma}$
    c. Estimation of the proportion of lot variation (τ)
    d. Calculation of the drug expiration period
        i. True drug expiration period
        ii. Predicted drug expiration period
3. Study 2: Investigation the number of samples
    a. Simulation of drug strength with different numbers of samples and replications
    b. Comparison of the results
4. Conclusion and discussion

CHAPTER II

LITERATURE REVIEW

2.1 Theoretical Basis

Linear Regression

The simple regression model with one variable is

$$Y = \beta_0 + \beta_1 X + \quad , \tag{2.1}$$

where

Y is a response vector size $n \times 1$,

X is a independent variable vector size $n \times 1$,

β_i is regression parameter; $I = 0$ and 1

is an error vector size $n \times 1$; $\sim N(0, \sigma^2)$.

Parameter β_0 and β_1 can be estimated from n observations by ordinary least square method, and the predictor,

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x\,, \tag{2.2}$$

is used for the study.

Confidence intervals for the regression surface

y_* is estimated from equation (2.2) by given $x = x_*$. Thus, a lower

$100\,(1-\alpha)\%$ confidence intervals for $\hat{y}_*$ is

$$\hat{y}_* - t_{\alpha,n-2}\sigma\sqrt{v_*},$$

(2.3)

where $t_{\alpha,n-2}$ is the upper α critical point of a t-distribution with $n-2$ degrees of freedom,

σ is a standard error,

$$v_* = x_*(X^TX)^{-1}x_*{}^T$$

$$= \frac{\{\sum x_i^2 - 2x_* n\bar{x} + nx_*^2\}}{n\sum(x_i-\bar{x})^2}$$

$$= \frac{\{\sum x_i^2 - n\bar{x}^2 + n(x_*-\bar{x})^2\}}{n\sum(x_i-\bar{x})^2}$$

$$= \frac{1}{n} + \frac{(x_*-\bar{x})^2}{\sum(x_i-\bar{x})^2} \quad ,$$

x_* is a x observation.

Prediction intervals for the response

Given value of $x = x_*$, from equation (2.1), then

$$y_* = \beta_0 + \beta_1 x_* + \ _*.$$

If we assume that

$$_* \sim N(0, \sigma^2)$$

and $_*$ is independent of $\varepsilon^T = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$,
then

$$E[\hat{y}_* - y_*] = x_*\beta - x_*\beta = 0$$

$$Var[\hat{y}_* - y_*] = Var[\hat{y}_*] + Var[y_*]$$

$$= \sigma^2 x_*(X^TX)^{-1}x_*{}^T + \sigma^2$$

$$= \sigma^2 v_* + \sigma^2$$

And

$$(\hat{y}_* - y_*) \sim N(0, (\sigma^2 v_* + \sigma^2).$$

Thus a lower $100(1-\alpha)\%$ confidence intervals for y_*(prediction interval) is

$$\hat{y}_* - t_{\frac{\alpha}{2},n-p}\sqrt{\sigma^2 v_* + \sigma^2}$$

$$= \hat{y}_* - t_{\frac{\alpha}{2},n-p}\sqrt{\sigma^2\left(\frac{1}{n} + \frac{(x_*-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right) + \sigma^2} \qquad (2.4)$$

Mixed-Effects Model

A mixed-effects model for one fixed-effect is written as

$$y_{ijk} = \mu + \beta_i + \alpha_{ij} + \varepsilon_{ijk}, \qquad (2.5)$$

where

μ is the general mean,

β_i is the fixed effect,

α_{ij} is a random variable representing the deviation from population of the sample,

ε_{ijk} is a random error representing the deviation from measurement of the sample j,

and

$$\alpha_{ij} \overset{iid}{\sim} N(0, \sigma_l^2),$$

$$\varepsilon_{ijk} \overset{iid}{\sim} N(0, \sigma_m^2).$$

The model with two sources of random variation, α_{ij} and ε_{ijk}, is sometimes called a hierarchical model or a multilevel model. The variance between observations is σ_l^2 corresponding to a correlation of

$$\tau = \frac{\sigma_l^2}{\sigma^2} = \frac{\sigma_l^2}{\sigma_l^2 + \sigma_m^2}. \tag{2.6}$$

Variance Components

According to α_{ij}s are independently and identically distributed,

$$\alpha_{ij} \overset{iid}{\sim} N(0, \sigma_l^2).$$

Consequently,

$$E(\alpha_{ij}) = 0, \quad \forall i, j$$

$$COV(\alpha_{ij}, \alpha_{i'j'}) = 0, \quad \forall i, i' \text{ and } j, j' \text{ except } i = i' \text{ and } j = j'.$$

From mixed-effects model (2.5),

$$y_{ijk} = \mu + \beta_i + \alpha_{ij} + \varepsilon_{ijk},$$

α_{ij}s are random effects, so $E(y_{ijk})$ should be conditional mean

$$E(y_{ijk}|\alpha_{ij}) = \mu + \beta_i + \alpha_{ij}.$$

Thus, the residual can be written as

$$\varepsilon_{ijk} = y_{ijk} - E(y_{ijk}|\alpha_{ij}).$$

Consequently,

$$E(\varepsilon_{ijk}) = 0, \quad \forall i, j, k$$

and

$$COV(\varepsilon_{ijk}, \varepsilon_{i'j'k'}) = 0, \quad \forall i, i', j, j' \text{and } k, k' \text{ except } i = i', j = j' \text{ and } k = k'.$$

and also

$$COV(\varepsilon_{ijk}, \alpha_{i'j'}) = 0, \quad \forall i, i', j, j' \text{and } k.$$

From mixed-effects model (2.5),

$$y_{ijk} = \mu + \beta_i + \alpha_{ij} + \varepsilon_{ijk},$$

$$\begin{aligned} Var(y_{ijk}) &= Var(\mu + \beta_i + \alpha_{ij} + \varepsilon_{ijk}) \\ &= Var(\alpha_{ij}) + Var(\varepsilon_{ijk}) \\ \sigma_y^2 &= \sigma_l^2 + \sigma_m^2 \end{aligned}$$

Since σ_l^2 and σ_m^2 are the components of the variance of y, they are called "variance components".

Variance Components Estimation

ANOVA estimator from balanced data was used for variance estimation. The methodology is to do ANOVA method as if the model is a fixed effects model and then derive the expected mean squares under mixed effects model.

Table 2.1: Analysis of variance estimators for variance estimation

Source of Variation	d.f.	Sum of Squares	Mean Square	Expected Mean Squares
Group (α_{ij})	*(a*-1*)*	*SSA*	*MSA = SSA/(a*-1*)*	$n\sigma_l^2 + \sigma_m^2$
Within Group	*a(n*-1*)*	*SSE*	*MSE = SSE/a(n*-1*)*	σ_m^2

From Table 2.1,

$$MSA = SSA/(a\text{-}1) = n\sigma_l^2 + \sigma_m^2$$

and $$MSE = SSE/a(n\text{-}1) = \sigma_m^2,$$

so σ_l^2 and σ_m^2 can be estimated as

$$\sigma_l^2 = (MSA - MSE)/n$$

and $$\sigma_m^2 = MSE.$$

2.2 Related Studies

From previous study, Shao and Chen (1997) estimated drug expiration period in difference package in general experiment. The data was collected at 0, 3, 6, 9, 12, 18 months after manufactured, then lower predicted bound was used to predict the drug expiration period comparing with the United States Food and Drug Administration (U.S. FDA) approach. The result shown that the drug labeled shelf life under their study were 27 and 26 months for bottle container and blister package respectively, while the FDA approach gave a labeled shelf life for bottle container of 26 months.

Komka et al. (2010) used tolerance interval comparing with lower 95 % confidence interval which is suggest by ICH guideline. The result found a much wider interval from their new approach than that from ICH guideline, resulting in a significant shorter expiration period. They also compared the interval width among difference sources of variation, which was narrower when tablet-to-tablet variability was lower comparing with analytical measurement error.

Moreover, Srimaneekarn et al. (2012) studied drug expiration period on simulated drug strength data. They indicated that the suitable value of proportion of lot variability in lower prediction interval model produced more effective than lower confidence interval model or lower prediction interval model. The results present that the different experimental designs and different total variability presented very similar simulation results for expiration period estimation.

CHAPTER III

METHODOLOGY

Simulate Drug Strength Data

According to ICH guideline, the frequency of testing should be every 3 months over the first year, every 6 months over the second year, and annually thereafter through the proposed period, so the time points (t_i) are 0, 3, 6, 9, 12, 18, 24 and 36 months. As the strength of drug reduces when the time increases, given the percentage of drug strength at manufacturing time is β_0 and $\beta_0 + \beta_1 t$ at time t, where $\beta_1 < 0$. Simulate drug strength (y_{ijk}) at each time point (t_i) from

$$y_{ijk} = \beta_0 + \beta_1 t_i + \alpha_{ij} + \varepsilon_{ijk} . \quad (3.1)$$

Random variable in equation (2.1), simple linear model, should be the lot variability (σ_l^2) which is the different among tablets or bottles in the same batch; however, during measurement process the measurement error (σ_m^2) also exists. Thus the equation is changed to equation (3.1), which there are two random variables; α_{ij} is a random variable representing the difference between drug samples, and ε_{ijk} is a random error representing the error from measuring the drug sample *j*, where $\alpha_{ij} \overset{iid}{\sim} N(0, \sigma_l^2)$ and $\varepsilon_{ijk} \overset{iid}{\sim} N(0, \sigma_m^2)$. Accordingly, the total error (σ^2) is the combination of the lot variability (σ_l^2) and the measurement error (σ_m^2),

$$\sigma^2 = \sigma_l^2 + \sigma_m^2. \quad (3.2)$$

Study1 : the proportion of lot variability determination and model development

Supposed that

$$\beta_0 = 100\%\,,\ \beta_1 = -0.5,\quad \sigma^2 = 1,\quad \alpha_{ij} \overset{iid}{\sim} N(0, \sigma_l^2),\ \varepsilon_{ijk} \overset{iid}{\sim} N(0, \sigma_m^2),$$

and given

Case 1	$\sigma_l^2 = 0$	$\tau = 0$
Case 2	$\sigma_l^2 = \frac{\sigma_m^2}{3} = \frac{\sigma^2}{4}$	$\tau = 0.25$
Case 3	$\sigma_l^2 = \sigma_m^2 = \frac{\sigma^2}{2}$	$\tau = 0.50$
Case 4	$\sigma_l^2 = 3\sigma_m^2 = \frac{3\sigma^2}{4}$	$\tau = 0.75$
Case 5	$\sigma_l^2 = \sigma^2$	$\tau = 1$

From equation (3.1), simulate drug strength (y_{ijk}) 5 samples (j = 1, 2, 3, 4, 5) at each time point (t_i), and repeat them 5 times per sample (k = 1, 2, 3, 4, 5 for each j) for the measurement error evaluation. Then $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}^2$ were estimated by ordinary least square method. The 10,000 simulations were performed for each case.

Estimate the proportion of lot variability(τ)

$$\tau = \frac{\sigma_l^2}{\sigma^2} = \frac{\sigma_l^2}{\sigma_l^2+\sigma_m^2} \tag{3.3}$$

As the mixed-effects model from equation (3.1),

$$y_{ijk} = \beta_0 + \beta_1 t_i + \alpha_{ij} + \varepsilon_{ijk} ,$$

the random effect variance (or lot variability, σ_l^2) and the random error (or measurement error, σ_m^2) can be estimated by ANOVA method for variance components estimation of mixed-effects model. Thus the proportion of lot variability (τ) can be estimated by

$$\hat{\tau} = \frac{\hat{\sigma}_l^2}{\hat{\sigma}^2} = \frac{\hat{\sigma}_l^2}{\hat{\sigma}_l^2+\hat{\sigma}_m^2} \tag{3.4}$$

Estimate Drug Expiration Period

Suppose the lower acceptance limit of drug strength is L, and drug expiration period is T, probability that the drug strength is greater or equal to L at time T is

$$P(N(\beta_0 + \beta_1 T, \sigma_l^2) \geq L) = 1 - \alpha . \tag{3.5}$$

1. <u>Calculate True Expiration Period</u>

According to ICH guideline, from equation (3.5), the true expiration period $(T_{true,\alpha})$ can be calculated as

$$T_{true,\alpha} = \frac{L-\beta_0+\sigma_l Z_{1-\alpha}}{\beta_1}. \tag{3.6}$$

And from equation (3.3), $\tau = \frac{\sigma_l^2}{\sigma^2} = \frac{\sigma_l^2}{\sigma_l^2+\sigma_m^2}$,

$$\sigma_l^2 = \sigma^2\tau,$$

hence $$T_{true,\alpha} = \frac{L-\beta_0+\sigma_l Z_{1-\alpha}}{\beta_1} = \frac{L-\beta_0+\sigma\sqrt{\tau} Z_{1-\alpha}}{\beta_1} . \quad (3.7)$$

2. Estimate Predicted Expiration Period

From equation (3.3) $\tau = \frac{\sigma_l^2}{\sigma^2} = \frac{\sigma_l^2}{\sigma_l^2+\sigma_m^2}$

$$\sigma_l^2 = \sigma^2\tau,$$

hence $$Var\left(\hat{\beta}_0 + \hat{\beta}_1 t\right) + \sigma_l^2 = \sigma^2\left(\frac{1}{n} + \frac{(T_2-\bar{t})^2}{\sum(t_i-\bar{t})^2}\right) + \sigma^2\tau.$$

2.1 Estimate by lower prediction interval method

From equation (2.4) lower prediction interval

$$\hat{y}_* - t_{\frac{\alpha}{2},n-p}\left(\sqrt{\sigma^2\left(\frac{1}{n} + \frac{(x_*-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right) + \sigma_l^2}\right),$$

$$= \hat{y}_* - t_{\frac{\alpha}{2},n-p}\left(\sqrt{\sigma^2\left(\frac{1}{n} + \frac{(x_*-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right) + \sigma^2\tau}\right),$$

The drug expiration period (T_{pd}) is estimated by $100(1-\alpha)\%$ lower prediction interval,

$$L = \hat{\beta}_0 + \hat{\beta}_1 T_{pd} - t_{\alpha,n-2}\hat{\sigma}\sqrt{\left(\frac{1}{n} + \frac{(T_{pd}-\bar{t})^2}{\sum(t_i-\bar{t})^2}\right) + \tau}\,, \quad (3.8)$$

for each τ, $\tau = 0, 0.25, 0.50, 0.75, 1$ and $\hat{\tau}$, which was previously estimated.

Please note that, when $\tau = 0, 1$ and $\hat{\tau}$, the predicted expiration periods are the lower confidence interval, the lower prediction interval and the newly proposed model (lower prediction interval with proper proportion of lot variability) respectively.

2.2 Estimate by direct method

From (3.5), $P(N(\beta_0 + \beta_1 T, \sigma_l^2) \geq L) = 1 - \alpha$, given $\alpha = 0.05,\ L = 90$. Hence, the predicted drug expiration period (T_{pd}) can be written as

$$T_{pd} = \frac{90 - \hat{\beta}_0 + \hat{\sigma}\sqrt{\tau}\, Z_{0.95}}{\hat{\beta}_1}, \tag{3.9}$$

for each $\tau,\ \tau = 0, 0.25, 0.50, 0.75, 1$ and $\hat{\tau}$, which was previously estimated.

Study2 : the number of drug sample investigation

Supposed that

$\beta_0 = 100\%$, $\beta_1 = -0.5$, $\sigma^2 = 1$, $\alpha_{ij} \overset{iid}{\sim} N(0, \sigma_l^2)$, $\varepsilon_{ijk} \overset{iid}{\sim} N(0, \sigma_m^2)$ and $\tau = 0.50$ or $\sigma_l^2 = \sigma_m^2 = \frac{\sigma^2}{2}$.

From equation (3.1), simulate drug strength (y_{ijk}) m samples (j = 1,…, m) at each time point (t_i), and repeat them p times per sample (k = 1,…, p for each j) for the measurement error evaluation. The number of samples (m) and replication (p) were in the table 3.1. Then $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}^2$ were estimated by ordinary least square method, and the variance components, $\hat{\sigma}_l^2$ and $\hat{\sigma}_m^2$, were estimated by ANOVA method for variance components estimation of mixed-effects model. The 10,000 simulations were performed.

Table 3.1: The number of samples (m) and replication (p)

	m	p	$n = mp$
Case1	5	2	10
Case2	3	3	9
Case3	2	5	10

The $T_{true,\alpha}$ was calculated from equation (3.7),

$$T_{true,\alpha} = \frac{L-\beta_0+\sigma_l Z_{1-\alpha}}{\beta_1} = \frac{L-\beta_0+\sqrt{\rho}\,\tau\, Z_{1-\alpha}}{\beta_1}.$$

The $T_{true,\alpha}$ is 17.6738, when L = 90 and $\alpha = 0.05$ for above data set. The Prediction expiration periods were calculated using the lower prediction interval method as in study 1, but given only $\tau = 0, 1$ and $\hat{\tau}$ these are the lower confidence interval, the lower prediction interval and the newly proposed model (lower prediction interval with proper proportion of lot variability) respectively.

CHAPTER IV

RESULTS

Study1: Proportion of Lot Variability Determination and Model Development

The results of the proportion of lot variability estimated by the ANOVA method of variance components estimation present a close estimation ($\hat{\tau}$) to the given the proportion of lot variability (τ) as shown in Table 4.1 below.

Table 4.1: Proportion of lot variability estimation

τ	τ	$\hat{\sigma}_l^2$	$\hat{\sigma}_m^2$
0	0 ± 0.0507	0 ± 0.0509	0.9990 ± 0.1120
0.25	0.2453 ± 0.0755	0.2489 ± 0.0926	0.7496 ± 0.0831
0.50	0.4912 ± 0.0770	0.4979 ± 0.1370	0.4997 ± 0.0554
0.75	0.7413 ± 0.0534	0.7468 ± 0.1822	0.2499 ± 0.0277
1	1	0.9958 ± 0.2278	0

The predicted expiration period simulation results with different cases using lower prediction interval method and using direct method are shown in Table 4.2 and table 4.3 respectively. The tables show the mean and standard deviation of the predicted expiration period (T_{pd}) with different cases.

The results show that the predicted period from the model with proportion of lot variability close to the true proportion gave the predicted expiration period, which is close to the true expiration period (T_{true}). That is the suitable predicted period for case 1, case 2, case 3, case 4, and case 5 are the lower prediction interval model with proportion of lot variability equal to 0, 0.25, 0.5, 0.75, and 1 respectively. Moreover, the predicted expiration period of the newly proposed model (the lower prediction interval with estimated proportion of lot variability) also produced the results close to the true expiration period as shown in table 4.2 and 4.3.

Table 4.2: Results of expiration period prediction (T_{pd}) with different cases using lower prediction interval

		T_{pd}						
		$\tau = 0$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 1$	$\tau = \hat{\tau}$	T_{true}
case1	mean	19.6857	18.0374	17.2415	16.6283	16.1106	19.4542	20.0000
	SD	0.1592	0.1806	0.2029	0.2237	0.2432	0.3517	0
case2	mean	19.6862	18.0465	17.2548	16.6449	16.1299	18.0813	18.3551
	SD	0.2263	0.2416	0.2608	0.2798	0.2984	0.4245	0
case3	mean	19.6876	18.0583	17.2716	16.6655	16.1538	17.2932	17.6738
	SD	0.2772	0.2988	0.3241	0.3490	0.3730	0.4561	0
case4	mean	19.6896	18.0720	17.2908	16.6891	16.1810	16.7015	17.151
	SD	0.3203	0.3540	0.3901	0.4246	0.4572	0.5019	0
case5	mean	19.7062	18.1010	17.3258	16.7286	16.2244	16.2244	16.7103
	SD	0.3540	0.4020	0.4516	0.4981	0.5414	0.5414	0

Table 4.3: Results of expiration period prediction (T_{pd}) with different cases using direct method

		T_{pd}						
		$\tau = 0$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 1$	$\tau = \hat{\tau}$	T_{true}
case1	mean	19.9988	19.1782	18.3575	17.5368	16.7162	19.9337	20.0000
	SD	0.1622	0.1627	0.1739	0.1941	0.2207	0.1916	0
case2	mean	19.9977	19.1813	18.3650	17.5486	16.7323	19.1906	18.3551
	SD	0.2310	0.2288	0.2365	0.2531	0.2771	0.3529	0
case3	mean	19.9971	19.1859	18.3748	17.5636	16.7524	18.3908	17.6738
	SD	0.2826	0.2809	0.2918	0.3140	0.3454	0.4342	0
case4	mean	19.9970	19.1916	18.3862	17.5808	16.7753	17.5959	17.1510
	SD	0.3261	0.3262	0.3435	0.3758	0.4196	0.4813	0
case5	mean	20.0115	19.2122	18.4129	17.6136	16.8143	16.8143	16.7103
	SD	0.3606	0.3619	0.3871	0.4320	0.4913	0.4913	0

The predicted expiration period simulation results using lower prediction interval method and using direct method for only case 3 are shown in table 4.4 and table 4.5 respectively. The tables show the mean and standard deviation of the predicted expiration period (T_{pd}) with different τ, mean of drug strength of all simulations which were calculated by each T_{pd}, the mean difference between T_{pd} and true expiration period(T_t), proportion of T_{pd} which was less than T_t ($T_{pd} \leq T_t$) and the mean of probability that drug strength was more than L at each T_{pd}. The other cases are presented in the Appendix A.

Using this simulation, $\sigma^2 = 1, \tau = 0.5$, and the equation (3.6), $T_{true,\alpha} = \frac{L-\beta_0+\sigma_l Z_{1-\alpha}}{\beta_1}$, the true expiration period can be calculated, which is 17.6738. The suitable model for drug expiration estimation should be a model with proper proportion of lot variability that can be estimated as with the experiment. As shown in Table 4.4, the results indicate that the predicted period of the lower prediction model with proportion of lot variability was close to the true proportion (0.5 for this study) and the given expiration period was close to the true expiration period (that is when $\tau = 0.5$ and $\hat{\tau}$). On the other hand, when $\tau = 0$ and 1, the lower confidence interval and the lower prediction interval, the prediction periods were longer and shorter comparing with T_t respectively.

Table 4.5 is similar table to table 4.4 for case 3, but the predicted period came from the direct method. The results presented longer prediction period than the lower prediction model, and also present less probability that drug strength was more than lower acceptance level at predicted period than the lower prediction model.

Table 4.4: Results of expiration period prediction (T_{pd}) with different τ using lower prediction interval method for case 3

	T_{pd}					
	$\tau = 0$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 1$	$\tau = \hat{\tau}$
Mean	19.6876	18.0583	17.2716	16.6655	16.1538	17.2932
SD	0.2772	0.2988	0.3241	0.3490	0.3730	0.4561
Drug Strength	90.1562	90.9709	91.3642	91.6672	91.9231	91.3534
Difference	2.0137	0.3845	-0.4023	-1.0083	-1.5201	-0.3806
$T_{pd} \leq T_{true}$	0	0.0967	0.8939	0.9985	1	0.7929
P(Dst $\geq L$)	0.5858	0.9104	0.9700	0.9890	0.9957	0.9657

Table 4.5: Results of expiration period prediction (T_{pd}) with different τ using direct method for case3

	T_{pd}					
	$\tau = 0$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 1$	$\tau = \hat{\tau}$
Mean	19.9971	19.1859	18.3748	17.5636	16.7524	18.3908
SD	0.2826	0.2809	0.2918	0.3140	0.3454	0.4342
Drug Strength	90.0014	90.4070	90.8126	91.2182	91.6238	90.8046
Difference	2.3233	1.5121	0.7009	-0.1103	-0.9215	0.7169
$T_{pd} \leq T_{true}$	0	0	0.0084	0.6376	0.9974	0.0508
P(Dst $\geq L$)	0.5008	0.7138	0.8698	0.9537	0.9872	0.8617

Mean: mean of T_{pd}

SD : standard deviation of T_{pd}

Drug Strength: drug strength calculated by each T_{pd}

Difference: the difference between T_{pd} and $T_{true,\alpha}(T_t)$

$T_{pd} \leq T_{true}$: proportion of T_{pd} which was less than or equal to T_{true}

P(Dst $\geq L$): probability that drug strength was more than or equal to L at T_{pd}

According to the ICH guideline, probability that the drug strength is greater or equal to L at a predicted time is 1- α. That probability also acts as a criterion for selection of the proper model. Table 4.6 is a table of that probability on different cases from the lower prediction interval method with different τ. The results show that the newly proposed model presents probability close to 0.95 in every case. The lower confidence interval (when τ = 0) is good only when there is no lot variability, and the lower prediction interval (when τ = 1) is good only when the measurement error are not presented.

Table 4.6: Probability that drug strength was more than L at T_{pd} with different cases using the lower prediction interval method

	T_{pd}					
	$\tau = 0$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 1$	$\tau = \hat{\tau}$
Case1	0.9765	1	1	1	1	0.9872
Case2	0.6202	0.9712	0.9961	0.9994	0.9999	0.9611
Case3	0.5858	0.9104	0.9700	0.9890	0.9957	0.9657
Case4	0.5700	0.8623	0.9365	0.9683	0.9835	0.9663
Case5	0.5575	0.824	0.9039	0.9438	0.9658	0.9658

The results in table 4.7 show that is not suitable for adding the proportion of lot variability in the direct model. However, it is better than direct method without the proportion of lot variability.

Table 4.7: Probability that drug strength was more than L at T_{pd} with different cases using the direct method

	T_{pd}					
	$\tau = 0$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 1$	$\tau = \hat{\tau}$
Case1	0.5036	1	1	1	1	0.6250
Case2	0.5009	0.7876	0.9442	0.9913	0.9992	0.7773
Case3	0.5008	0.7138	0.8698	0.9537	0.9872	0.8617
Case4	0.5007	0.6768	0.8196	0.9139	0.9648	0.9094
Case5	0.4978	0.6509	0.7821	0.8783	0.9391	0.9391

Moreover, the results also presented the difference between the lower prediction interval method and the direct method. The predicted expiration periods from the lower prediction interval method were shorter than the direct method, due to their formulas. For example, table 4.8 shows the expiration period predictions from the lower prediction interval method and the direct method with proportion of lot variability from different cases, $\sigma^2 = 1, \tau = 0.5$. All expiration periods from the direct method are greater than the lower prediction interval method and mostly are greater than a true expiration period.

Table 4.8: Results of expiration period prediction (T_{pd}) from the lower prediction interval method and the direct method with proportion of lot variability from different cases

Method		T_{pd}				
		$case1$	$case2$	$case3$	$case4$	$case5$
LPI	mean	19.4542	18.0813	17.2932	16.7015	16.2244
	SD	0.3517	0.4245	0.4561	0.5019	0.5414
DM	mean	19.9337	19.1906	18.3908	17.5959	16.8143
	SD	0.1916	0.3529	0.4342	0.4813	0.4913
T_t		20	18.3551	17.6738	17.1510	16.7103

LPI: Lower prediction interval method

DM: Direct method

T_t: True expiration period

Study2 : Number of Samples Investigation

According to the ICH guideline, the number of samples in a stability study is not more than ten (including replications), so there are three possible methods: 5 samples with 2 replications, 3 samples with 3 replications and 2 samples with 5 replications.

The predicted expiration period results with different numbers of samples and replications are shown in table 4.9, table 4.10 and table 4.11, which were calculated using the lower confidence interval model, the lower prediction interval model and the newly proposed model respectively. Each table includes the mean and standard deviation of the predicted expiration period (T_{pd}) with different numbers of samples, the mean square of the difference between T_{pd} and T_t, $\sum_{i=1}^{N}(T_{pd} - T_t)^2/N$, mean of drug strength of all simulations which were calculated by each T_{pd}, the mean difference between T_{pd} and T_t, proportion of T_{pd} which was less than T_t ($T_{pd} \leq T_t$) and the mean of probability that drug strength was more than L at each T_{pd}.

The results show that when the number of lots is increased, the standard deviation decreases. In addition, as the number of lots is increased, the probability that drug strength is greater than the lower acceptance level at the predicted period is also increased as in table 4.10 and table 4.11 respectively. The mean square of the difference between T_{pd} and T_t increased when using the lower confidence interval model, reduced when using the lower prediction interval model and the least when using the newly proposed model respectively.

Table4.9: Results of expiration period prediction (T_{pd}, the confidence interval model) with different numbers of samples and replications

	T_{pd}(confidence interval)		
	5:2	3:3	2:5
Mean	19.5148	19.4978	19.5285
SD	0.3118	0.3794	0.4438
MSDiff	3.4865	3.4709	3.6366
Drug St	90.2426	90.2511	90.2358
Difference	1.8410	1.8240	1.8546
$T_{pd} \leq T_t$	0	0	0
P(Dst $\geq L$)	0.6312	0.6343	0.625

MSDiff is mean square of the difference between T_{pd} and T_t, $\sum_{i=1}^{N}(T_{pd} - T_t)^2/N$.

Table4.10: Results of expiration period prediction (T_{pd}, the prediction interval model) with different numbers of samples and replications

	T_{pd}(prediction interval)		
	5:2	3:3	2:5
Mean	16.133	16.1733	16.2195
SD	0.4495	0.5292	0.5940
MSDiff	2.5761	2.5317	2.4679
Drug St	91.9335	91.9134	91.8903
Difference	-1.5408	-1.5006	-1.4543
$T_{pd} \leq T_t$	0.9997	0.9979	0.9933
P(Dst $\geq L$)	0.9954	0.9944	0.9932

MSDiff is mean square of the difference between T_{pd} and T_t, $\sum_{i=1}^{N}(T_{pd} - T_t)^2/N$.

Table4.11: Results of expiration period prediction (T_{pd}, the newly proposed model) with different numbers of samples and replications

	T_{pd}(prediction interval)		
	5:2	3:3	2:5
Mean	17.2860	17.3233	17.3601
SD	0.5880	0.6559	0.7278
MSDiff	0.4961	0.5531	0.6281
Drug St	91.3570	91.3384	91.3199
Difference	-0.3878	-0.3506	-0.3137
$T_{pd} \leq T_t$	0.7529	0.7073	0.6688
P(Dst $\geq L$)	0.9615	0.9568	0.9515

MSDiff is mean square of the difference between T_{pd} and T_t, $\sum_{i=1}^{N}(T_{pd} - T_t)^2/N$.

CHAPTER V

DISCUSSION AND CONCLUSION

As illustrated in Table 4.1, the results show that the predicted period of the model with proportion of lot variability was close to the true proportion and also the given expiration period was close to the true expiration period (T_t). In contrast, the lower confidence interval and the lower prediction interval offer the prediction period these are considerably longer and shorter than the true expiration period, respectively. In the former case, when put into use, the predicted expiration dates entrust a significant portion of expired drugs to patients The expired drugs may be ineffective or can cause adverse effect on the patients' health. In the latter case, usable drugs are wasted, causing unnecessary increase of the overall medical and social security cost. Therefore, we recommend our newly proposed model over both the conventional confidence interval model and the prediction interval model.

According to the proposed model, in equation (3.1), there are two random variables: α_{ij} is a random variable representing the drug sample's deviated from the population, and ε_{ijk} is a random variable representing the deviation from repetition of the drug sample j, where $\alpha_{ij} \overset{iid}{\sim} N(0, \sigma_l^2)$ and $\varepsilon_{ijk} \overset{iid}{\sim} N(0, \sigma_m^2)$. Accordingly, the total variance (σ^2) is the combination of the lot variability (σ_l^2) and the measurement error (σ_m^2). The expiration period prediction model should have an additional term, the proportion of lot variability, which is estimated by ANOVA method for variance components estimation. Table 4.6 shows that incorporating this parameter into the prediction model improves the quality of the predicted expiration date.

For the proposed model, keeping the total number of repetitions constant, the standard deviation decreases as the number of lots increases, as shown in table 4.11. In addition, as the number of lots is increased, the probability that drug strength is

greater than the lower acceptance level at the predicted period is also increased. Both effects of large number of lots are beneficial. Therefore, according to the data model under study, with the constraint of 10 repetitions per time point required by the regulator, we recommend the number of samples be 5 lots with 2 replications. As a caveat, it should be noted that other selection criteria can lead to different recommendations for the number of samples and replications.

REFERENCES

Faraway, J. J. (2006). *Extending the Linear Model with R.* Boca Raton: Chapman & Hall/CRC.

Food and Drug Administration. (2004). ICH harmonised tripartite guideline, evaluation for stability data Q1E. *The Federal Register*, 69, 110: 32010-32011.

Food and Drug Administration. (2004). ICH harmonised tripartite guideline, stability testing of new drug substances and products, Q1A (R2). *The Federal Register*, 68, 225: 65717-65718.

Komka, K., Kemeny, S., and Banfai, B. (2010). Novel tolerance interval model for the estimation of the shelf life of pharmaceutical products. *Journal of Chemometrics*, 24, 131-139.

McCulloch, C. E., and Searle, S.R. (2001). *Generalized, Linear, and Mixed Models*, New York: John Wiley and Sons.

Pinheiro, J.C., and Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS.* New York: Springer.

Searle, S.R., Casella, G., and McCulloch, C. E. (2006). *Variance Components*, New Jersey: John Wiley and Sons.

Searle, S.R. (1971). *Linear Models*, New York: John Wiley and Sons.

Seber, G.A.F. (1977). *Linear Regression Analysis.* New York: John Wiley and Sons.

Shao, J., and Chen, L. (1997). Prediction bounds for random shelf-life. *Statistics in Medicine*, 16, 1167-1173.

Srimaneekarn, N., Kiatsupaibul, S., Hayter, A. J., and Liu, W. Estimating drug shelf-life with unknown lot-to-lot variability. *Working paper*.

Appendices

Appendix A

A: Additional Results

Table A1.1: Results of expiration period prediction (T_{pd}) with different τ using lower prediction interval method case 1

	T_{pd}					
	$\tau = 0$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 1$	$\tau = \hat{\tau}$
Mean	19.6857	18.0374	17.2415	16.6283	16.1106	19.4542
SD	0.1592	0.1806	0.2029	0.2237	0.2432	0.3517
Drug Strength	90.1571	90.9813	91.3793	91.6858	91.9447	90.2729
Difference	-0.3143	-1.9626	-2.7585	-3.3717	-3.8894	-0.5458
$T_{pd} \leq T_t$	0.9765	1	1	1	1	0.9872
P(Dst $\geq L$)	0.9765	1	1	1	1	0.9872

Table A1.2: Results of expiration period prediction (T_{pd}) with different τ using direct method case 1

	T_{pd}					
	$\tau = 0$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 1$	$\tau = \hat{\tau}$
Mean	19.9988	19.1782	18.3575	17.5368	16.7162	19.9337
SD	0.1622	0.1627	0.1739	0.1941	0.2207	0.1916
Drug Strength	90.0006	90.4109	90.8212	91.2316	91.6419	90.0331
Difference	-0.0012	-0.8218	-1.6425	-2.4632	-3.2838	-0.0663
$T_{pd} \leq T_t$	0.5036	1	1	1	1	0.6250
P(Dst $\geq L$)	0.5036	1	1	1	1	0.6250

Mean: mean of T_{pd}

SD: standard deviation of T_{pd}

Drug Strength: drug strength calculated by each T_{pd}

Difference: the difference between T_{pd} and $T_{true,\alpha}(T_t)$

$T_{pd} \leq T_t$: proportion of T_{pd} which was less than or equal to T_t

P(Dst $\geq L$): probability that drug strength was more than or equal to L at T_{pd}

Table A2.1: Results of expiration period prediction (T_{pd}) with different τ using lower prediction interval method case 2

	T_{pd}					
	$\tau = 0$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 1$	$\tau = \hat{\tau}$
Mean	19.6862	18.0465	17.2548	16.6449	16.1299	18.0813
SD	0.2263	0.2416	0.2608	0.2798	0.2984	0.4245
Drug Strength	90.1569	90.9767	91.3726	91.6775	91.9350	90.9594
Difference	1.3310	-0.3086	-1.1003	-1.7102	-2.2252	-0.2739
$T_{pd} \leq T_t$	0	0.8972	1	1	1	0.7455
P(Dst $\geq L$)	0.6202	0.9712	0.9961	0.9994	0.9999	0.9611

Table A2.2: Results of expiration period prediction (T_{pd}) with different τ using direct method case 2

	T_{pd}					
	$\tau = 0$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 1$	$\tau = \hat{\tau}$
Mean	19.9977	19.1813	18.365	17.5486	16.7323	19.1906
SD	0.2310	0.2288	0.2365	0.2531	0.2771	0.3529
Drug Strength	90.0012	90.4093	90.8175	91.2257	91.6339	90.4047
Difference	1.6425	0.8262	0.0098	-0.8065	-1.6228	0.8354
$T_{pd} \leq T_t$	0	0.0002	0.4870	0.9996	1	0.0112
P(Dst $\geq L$)	0.5009	0.7876	0.9442	0.9913	0.9992	0.7773

Mean: mean of T_{pd}

SD: standard deviation of T_{pd}

Drug Strength: drug strength calculated by each T_{pd}

Difference: the difference between T_{pd} and $T_{true,\alpha}(T_t)$

$T_{pd} \leq T_t$: proportion of T_{pd} which was less than or equal to T_t

P(Dst $\geq L$): probability that drug strength was more than or equal to L at T_{pd}

Table A3.1: Results of expiration period prediction (T_{pd}) with different τ using lower prediction interval method case 3

	T_{pd}					
	$\tau = 0$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 1$	$\tau = \hat{\tau}$
Mean	19.6876	18.0583	17.2716	16.6655	16.1538	17.2932
SD	0.2772	0.2988	0.3241	0.3490	0.3730	0.4561
Drug Strength	90.1562	90.9709	91.3642	91.6672	91.9231	91.3534
Difference	2.0137	0.3845	-0.4023	-1.0083	-1.5201	-0.3806
$T_{pd} \leq T_t$	0	0.0967	0.8939	0.9985	1	0.7929
P(Dst $\geq L$)	0.5858	0.9104	0.97	0.989	0.9957	0.9657

Table A3.2: Results of expiration period prediction (T_{pd}) with different τ using direct method case 3

	T_{pd}					
	$\tau = 0$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 1$	$\tau = \hat{\tau}$
Mean	19.9971	19.1859	18.3748	17.5636	16.7524	18.3908
SD	0.2826	0.2809	0.2918	0.3140	0.3454	0.4342
Drug Strength	90.0014	90.4070	90.8126	91.2182	91.6238	90.8046
Difference	2.3233	1.5121	0.7009	-0.1103	-0.9215	0.7169
$T_{pd} \leq T_t$	0	0	0.0084	0.6376	0.9974	0.0508
P(Dst $\geq L$)	0.5008	0.7138	0.8698	0.9537	0.9872	0.8617

Mean: mean of T_{pd}

SD: standard deviation of T_{pd}

Drug Strength: drug strength calculated by each T_{pd}

Difference: the difference between T_{pd} and $T_{true,\alpha}(T_t)$

$T_{pd} \leq T_t$: proportion of T_{pd} which was less than or equal to T_t

P(Dst $\geq L$): probability that drug strength was more than or equal to L at T_{pd}

Table A4.1: Results of expiration period prediction (T_{pd}) with different τ using lower prediction interval method case 4

	T_{pd}					
	$\tau = 0$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 1$	$\tau = \hat{\tau}$
Mean	19.6896	18.072	17.2908	16.6891	16.1810	16.7015
SD	0.3203	0.3540	0.3901	0.4246	0.4572	0.5019
Drug Strength	90.1552	90.964	91.3546	91.6555	91.9095	91.6493
Difference	2.5386	0.9209	0.1398	-0.4619	-0.9701	-0.4496
$T_{pd} \leq T_t$	0	0.0043	0.3624	0.8582	0.9858	0.8133
P(Dst $\geq L$)	0.5700	0.8623	0.9365	0.9683	0.9835	0.9663

Table A4.2: Results of expiration period prediction (T_{pd}) with different τ using direct method case 4

	T_{pd}					
	$\tau = 0$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 1$	$\tau = \hat{\tau}$
Mean	19.9970	19.1916	18.3862	17.5808	16.7753	17.5959
SD	0.3261	0.3262	0.3435	0.3758	0.4196	0.4813
Drug Strength	90.0015	90.4042	90.8069	91.2096	91.6123	91.2020
Difference	2.8460	2.0406	1.2351	0.4297	-0.3757	0.4449
$T_{pd} \leq T_t$	0	0	0.0004	0.1219	0.8128	0.1761
P(Dst $\geq L$)	0.5007	0.6768	0.8196	0.9139	0.9648	0.9094

Mean: mean of T_{pd}

SD: standard deviation of T_{pd}

Drug Strength: drug strength calculated by each T_{pd}

Difference: the difference between T_{pd} and $T_{true,\alpha}(T_t)$

$T_{pd} \leq T_t$: proportion of T_{pd} which was less than or equal to T_t

P(Dst $\geq L$): probability that drug strength was more than or equal to L at T_{pd}

Table A5.1: Results of expiration period prediction (T_{pd}) with different τ using lower prediction interval method case 5

	T_{pd}					
	$\tau = 0$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 1$	$\tau = \hat{\tau}$
Mean	19.7062	18.101	17.3258	16.7286	16.2244	16.2244
SD	0.3540	0.4020	0.4516	0.4981	0.5414	0.5414
Drug Strength	90.1469	90.9495	91.3371	91.6357	91.8878	91.8878
Difference	2.9959	1.3907	0.6155	0.0184	-0.4859	-0.4859
$T_{pd} \leq T_t$	0	0.0003	0.0832	0.4881	0.8137	0.8137
P(Dst $\geq L$)	0.5575	0.8240	0.9039	0.9438	0.9658	0.9658

Table A5.2: Results of expiration period prediction (T_{pd}) with different τ using direct method case 5

	T_{pd}					
	$\tau = 0$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 1$	$\tau = \hat{\tau}$
Mean	20.0115	19.2122	18.4129	17.6136	16.8143	16.8143
SD	0.3606	0.3619	0.3871	0.4320	0.4913	0.4913
Drug Strength	89.9943	90.3939	90.7936	91.1932	91.5929	91.5929
Difference	3.3012	2.5019	1.7026	0.9033	0.1040	0.1040
$T_{pd} \leq T_t$	0	0	0	0.0180	0.4154	0.4154
P(Dst $\geq L$)	0.4978	0.6509	0.7821	0.8783	0.9391	0.9391

Mean: mean of T_{pd}

SD: standard deviation of T_{pd}

Drug Strength: drug strength calculated by each T_{pd}

Difference: the difference between T_{pd} and $T_{true,\alpha}(T_t)$

$T_{pd} \leq T_t$: proportion of T_{pd} which was less than or equal to T_t

P(Dst $\geq L$): probability that drug strength was more than or equal to L at T_{pd}

Appendix B

B : R Code

```
#***********************************************#
#***  Simulation for Drug Expiration Period Prediction  ***#
#***********************************************#

set.seed(101)
nround<-10000
L<-90
beta0<-100
beta1<- (-0.5)
sample<-2
repl<-5
tvec<-c(0,3,6,9,12,18,24,36)
time<-rep(rep(tvec,each=sample),each=repl)

alpha<-.05
talpha<-qt((1-alpha),length(tvec)-2)

sigma<-1                                    # Sigma = 1
rhovec<-c(0,1/4,1/2,3/4,1)                  # Rhovec = 0,1/4,2/4,3/4,1
rho<-1/2

gammavec<-c()
errorvec<-c()
strength<-c()
drug<-c()
Tt<-c()
Tp<-c()
Tdi<-c()
outputmat<-matrix(0,nrow=nround,ncol=13)
outputmat2<-matrix(0,nrow=nround,ncol=7)
for (i in 1:nround){
```

```
#*****************************#
#***   Simulate Drug Strength    ***#
#*****************************#

# Lot variation
        gammavec<-
rep(rnorm(length(tvec)*sample,0,sqrt(rho*(sigma^2))),each=repl)

# Variation of measurement
        errorvec<-rnorm(length(time),0,sqrt((1-rho)*(sigma^2)))

# Simulate drug strength
        strength<-beta0 + beta1*time + gammavec + errorvec
        drug<-rep(1:(length(tvec)*sample),each=repl)
        fdrug<-factor(drug)
        drugdata<-data.frame(strength=strength,time=time,fdrug=fdrug)

# Estimate beta0, beta1, sigma
        model<-lm(strength~time)
```

```
#*************************************#
#***   Calculate Predicted Expiration Period   ***#
#*************************************#

# Lower prediction interval method
        st<-sum((time-mean(time))^2)                    # St

        for (k in 1:length(rhovec)){
                f1<-function(T){
                        model$coef[1]+model$coef[2]*T-
                        talpha*summary(model)$sigma*sqrt(rhovec[k]+
                        1/length(time)+(T-mean(time))^2/st)-L
                }
                Tp[k]<-uniroot(f1,c(0,100))$root
        }

# Direct method
        for (k in 1:length(rhovec)){
                f2<-function(T){
                        (L-model$coef[1]+
                        rhovec[k]*summary(model)$sigma*qnorm(.95))/
                        model$coef[2]-T
                }
                Tdi[k]<-uniroot(f2,c(0,100))$root
        }
```

```
#**********************************#
#***   Calculate True Expiration Period    ***#
#**********************************#

        Tt<-(L-beta0-qnorm(0.05)*sqrt(rho)*sigma)/beta1


#**************************************#
#***  Fitted by Analysis of Variance Model    ***#
#**************************************#

        modav<-aov(strength~time+fdrug,drugdata)
        sigmalsqhat<-(anova(modav)$Mean[2]-anova(modav)$Mean[3])/repl
        sigmamsqhat<-anova(modav)$Mean[3]
        rhohat<-sigmalsqhat/(sigmalsqhat+sigmamsqhat)


#************************************************#
#***   Calculate Prediction Expiration Period by rhohat   ***#
#************************************************#

# Lower prediction interval method
        f3<-function(T){
                model$coef[1]+model$coef[2]*T-talpha*summary(model)$sigma*
                sqrt(max(rhohat,0)+1/length(time)+(T-mean(time))^2/st)-L
        }
        Tn<-uniroot(f3,c(0,100))$root


# Direct method
        f4<-function(T){
                (L-model$coef[1]+
                max(rhohat,0)*summary(model)$sigma*qnorm(.95))/model$coef[2]
                -T
        }
        Tndi<-uniroot(f4,c(0,100))$root
```

```
        #**************************#
        #**********  Output  ********#
        #**************************#

        # Lower prediction interval method
                outputmat[i,1:2]<-model$coef[1:2]

                outputmat[i,3]<-summary(model)$sigma

                outputmat[i,4:8]<-Tp[1:5]

                outputmat[i,9]<-Tn

                outputmat[i,10]<-Tt

                outputmat[i,11]<-rhohat

                outputmat[i,12:13]<-c(sigmalsqhat,sigmamsqhat)


        # Direct method

                outputmat2[i,1:5]<-Tdi[1:5]

                outputmat2[i,6]<-Tndi

                outputmat2[i,7]<-Tt

}
```

```
#*******************************#
#********** E valuation  ***********#
#*******************************#

        beta<-matrix(0,nrow=2,ncol=3)                    # bata0 beta1 sigma
        beta[1,1:3]<-apply(outputmat[,1:3],2,mean)
        beta[2,1:3]<-apply(outputmat[,1:3],2,sd)

        b1<-sum(outputmat[,2]>0)                         # n(beta1>0)



        #*****************************************#
        #********** Function for Evaluation ***********#
        #*****************************************#

        # function for Average drug strength at time x
                dst<-function(x)mean(beta0+beta1*x)
        # function for average diff  Tt & x
                dif<-function(x)mean(x-Tt)
        # function for Proportion < T(true,0.05)

                por<-function(x)mean(x<=Tt)
        # function for Average P( N(b0+b1*x,sigmal) > L )
                exc<-function(x)mean(1-pnorm(L,beta0+beta1*x,sqrt(rho*(sigma^2))))
```

```
#************************************#
#********** Evaluation Results  **********#
#************************************#

result<-matrix(NA,nrow=6,ncol=10)
result[1,1:10]<-round(apply(outputmat[,4:13],2,mean),4)
result[2,1:10]<-round(apply(outputmat[,4:13],2,sd),4)
result[3,1:7]<-round(apply(outputmat[,4:10],2,dst),4)
result[4,1:7]<-round(apply(outputmat[,4:10],2,dif),4)
result[5,1:7]<-round(apply(outputmat[,4:10],2,por),4)
result[6,1:7]<-round(apply(outputmat[,4:10],2,exc),4)
colnames(result)<- c("Tp,r=0","Tp,r=1/4","Tp,r=2/4","Tp,r=3/4","Tp,r=1",
                     "Tn","Tt","rho","sigmal^2","sigmam^2")
rownames(result)<-c("Mean","SD","Drug St","Tp-Tt","%(Tp<=Tt)","P(DSt>=L)")

result2<-matrix(NA,nrow=6,ncol=7)
result2[1,1:7]<-round(apply(outputmat2[,1:7],2,mean),4)
result2[2,1:6]<-round(apply(outputmat2[,1:6],2,sd),4)
result2[3,1:6]<-round(apply(outputmat2[,1:6],2,dst),4)
result2[4,1:6]<-round(apply(outputmat2[,1:6],2,dif),4)
result2[5,1:6]<-round(apply(outputmat2[,1:6],2,por),4)
result2[6,1:6]<-round(apply(outputmat2[,1:6],2,exc),4)
colnames(result2)<-
c("Tdi,r=0","Tdi,r=1/4","Tdi,r=2/4","Tdi,r=3/4","Tdi,r=1","Tndi","Tt")
rownames(result2)<-c("Mean","SD","Drug St","Tp-Tt","%(Tp<=Tt)","P(DSt>=L)")
```

```
a<-c("sigma","rho","sample","repl")
b<-c(sigma,rho,sample,repl)
rbind(a,b)

b1                                  # n(beta1>0)
result                              # Lower prediction interval method
result2                             # Direct method
write.csv(result,file="result.csv")
write.csv(result2,file="result2.csv")
```

BIOGRAPHY

Natchalee Srimaneekarn was born on January 21st, 1984 in Ratchaburi, Thailand. After she graduated with Doctor of Dental Surgery (D.D.S.) from Faculty of Dentistry, Mahidol University in 2008, she has been a faculty staff until now. In 2010, she enrolled in Master of Science in Statistics program at Department of Statistics, Faculty of Commerce and Accountancy, Chulalongkorn University.