Drug Development and Industrial Pharmacy, 22(2), 143–151 (1996)



# Calibration and Stability Analysis with a Simple Mixed Linear Model When the Experiment Is of a Split-Plot Type

Kwan R. Lee and Robert C. Gagnon

*Statistical Sciences Department, SmithKline Beecham Pharmaceuticals, 709 Swedeland Road, King of Prussia, Pennsylvania 19406*

**ABSTRACT**

*A typical calibration problem in bioassay involves several known concentrations of an analyte, with replicate determinations of the analyte at each level of concentration. A straight-line regression between the replicate determinations of the analyte and the known concentrations is sought. Common practice in this situation is to obtain several determinations of one analyte at each consecutive concentration level, instead of running the experiment in a completely randomized fashion. A proper model for this type of calibration experiment is a simple mixed linear model with two sources of variation: between analyte concentrations, and within analyte concentrations (model A). However, a practitioner may use a model based on the average determination at each concentration of analyte (model B), or he/she may adopt a model that assumes a completely independent error structure (model C) to fit a straight line. The purpose of this article is to compare the analysis based on models B and C to the supposedly correct analysis based on model A. It is shown that the statistical analysis based on model B is equivalent to the one based on model A except for the estimation of individual variance components. Simulated data sets generated from model A under various parameter value combinations are analyzed to provide the summary of performance for both models B and C and to point out the shortcomings of the usual practices based on model C. Application to pharmaceutical stability analysis is also given.*

## INTRODUCTION

Linear regression analysis is an important and often used tool for data analysis in the pharmaceutical industry. The simplicity of the method, the relative ease at which the method can be understood, and the wide availability of computer programs facilitate the use of this important statistical tool. Although linear regression is conceptually easy for practitioners to understand, many fail to recognize the importance of the statistical assumptions that underlie correct application of the method. In particular, assumptions about the underlying error structure of the linear regression model are not well understood by many practitioners. Further, most readily available linear regression computer programs provide little assistance to practitioners when it comes to evaluating these assumptions. As a result, linear regression models are at times misspecified. What is the impact of such misspecification? Is the linear regression model robust to misspecification about the underlying error structure of the model?

For example, in a typical HPLC (high performance liquid chromotography) calibration assay, $n$ known concentrations of the analyte are formulated. Usually each concentration is formulated by spiking reference material (e.g., plasma) with known concentrations of the analyte. From each concentration, $k$ replicate extraction procedures are performed to isolate the analyte. This type of calibration experiment produces a nested error structure, having within concentration components of variation nested within between-concentration components. The major source of between analyte concentration variability is due to preparation-to-preparation variation of the spiked reference material. The $nk$ isolated analytes are injected into the HPLC, and the calibration line is constructed based on the known (spiked) analyte levels and the $nk$ responses. A proper statistical model for the calibration line for this type of experiment is a simple mixed linear model with two sources of variation at each concentration: between analyte concentrations and within analyte concentrations (model A). However, a practitioner may adopt a statistical model based on the average of the $k$ determinations at each concentration level of the analyte (model B), or he/she may use a statistical model that assumes a completely independent error structure (model C). Statistical tests of the estimated slope parameter are often the goal of this type of calibration assay. Proper specification of the underlying statistical model is essential for obtaining the correct results. In this paper, we investigate the impact of model misspecification on the coverage probability of the slope parameter under different levels of intraclass (within concentration) correlation.

In a second example, a formulations group will generate a "stability batch" of a drug during its preclinical stage of development. This batch is used to check the stability of the drug prior to releasing it into clinical studies. Once the stability of the drug is confirmed, a "clinical batch" is generated, which is utilized for clinical trials of the drug. The clinical batch is checked for stability for the duration of the drug's use in clinical studies (5 years is not uncommon). In each case, stability is determined using well-known statistical methodology (1). The methodology is based on linear regression of the assayed quantity of the drug versus time. Typically, practitioners adopt statistical models based on averages (model B) or complete independence (model C). In reality, since several assays at the same time point tend to have positive correlation, the mixed linear model (model A with between and within time as the sources of variation) is appropriate. The FDA, in the *Guideline for Submitting Documentation for the Stability of Human Drugs and Biologics* (2), states that "procedures will remain valid even when . . . assumptions are mildly violated. If there is evidence of severe violation of the assumptions of the data, an alternate approach may be necessary." Thus, it is important to utilize the appropriate statistical model. Also, the FDA's use of the word *mildly* calls into question how far the model can stray from the true model before stability estimates become unacceptable. In this paper, we investigate the impact of model misspecification on estimates of shelf life.

Norwood (3) considered a mixed linear model with nested error structure for the stability analysis of several batches of a drug. He fitted the model using the approach outlined by Fuller and Battese (4), but his study did not include analysis of the effects of model misspecification.

In this paper, we present a detailed analysis of models A, B, and C. We compare these three models to each other for balanced designs, and we argue that model B is equivalent to model A for inference about the slope parameter and for estimation of shelf life. Therefore, unless we are interested in estimation of individual variance components and/or intraclass correlation, model B is sufficient. A numerical example is included for illustration, and a simulation study is summarized that gives the extent of model C's deviation from the correct model under various combinations of inter- and intra- class variation. In general, model C should not be used unless intraclass correlation is extremely small.

## METHODS

### A Simple Mixed Linear Model

An appropriate model is a simple mixed linear model given as

$$\text{Model A: } y_{ij} = \beta_0 + \beta_1(x_i - \bar{x}) + b_i + w_{ij}, \quad i = 1\ldots n, \; j = 1\ldots k, \tag{1}$$

where $y_{ij}$ is the $j$th determination of an analyte at concentration $x_i$. The intercept $\beta_0$ and the slope $\beta_1$ need to be determined. The symbol $\bar{x}$ is the average of $x_i$'s. The term $b_i$ is an independent random variable with mean 0 and variance $\sigma_b^2$; $w_{ij}$ is an independent random variable with mean 0 and variance $\sigma_w^2$, and is independent of $b_i$. It follows that the variance-covariance matrix $\Sigma$ of $y_{ij}$ is block diagonal. Elements of each diagonal block are given by $\sigma_b^2 + \sigma_w^2$ for the diagonal and $\sigma_b^2$ for the off-diagonal elements (compound symmetry) as follows:

$$\text{Let } \boldsymbol{B} = \begin{bmatrix} \sigma_b^2 + \sigma_w^2 & \cdots & \sigma_b^2 \\ \vdots & \ddots & \vdots \\ \sigma_b^2 & \cdots & \sigma_b^2 + \sigma_w^2 \end{bmatrix},$$

$$\text{then } \Sigma = \begin{bmatrix} \boldsymbol{B} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B} & \cdots & \mathbf{0} \\ \vdots & \cdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{B} \end{bmatrix}$$

where $\mathbf{0}$ = matrix of 0's.

In other words,

$$\begin{aligned} \operatorname{cov}(y_{ij}, y_{i'j'}) &= \sigma_b^2 + \sigma_w^2, \quad \text{if } i = i', \; j = j' \\ &= \sigma_b^2, \quad \text{if } i = i', \; j \neq j' \\ &= 0, \quad \text{if } i \neq i' \end{aligned} \tag{2}$$

Hence the intraclass correlation $\rho$ is proportional to $\sigma_b^2$ and is given by $\sigma_b^2/(\sigma_b^2 + \sigma_w^2)$. If we define $\beta = (\beta_0, \beta_1)$, then it is well known that the ordinary least squares estimator $\hat{\beta}$ of $\beta$ is BLUE (best linear unbiased estimator), and is equal to the generalized least squares estimator for this particular variance structure of compound symmetry (5,6). If we denote the least squares estimators of $\beta_0$ and $\beta_1$ as $\hat{\beta}_0$ and $\hat{\beta}_1$ respectively, it is not difficult to show the decompositions of the degrees of freedom, sums of squares, and the expected mean squares in an analysis of variance table (Table 1).

If the test for the slope under normality is desired, it should be done with the F-ratio of the mean squares from the first row and the second row, as can be seen from the expected mean squares (EMS) column of the ANOVA table (Table 1). The mean squares for the within variation should never be used for this test since it will unduly inflate the F-ratio. The variance components $\sigma_b^2$ and $\sigma_w^2$ can be estimated by the usual analysis of variance estimators from this table by equating mean squares to the EMS column. Restricted maximum likelihood (REML) estimators of the variance components are popular for the mixed model and they are equivalent to the usual ANOVA estimators for the balanced design model (8). In fact the above ANOVA estimators can be shown to be the REML estimators for the model A. The variance of the slope estimator, $\hat{\beta}_1$, can be shown (see appendix) to be

$$\operatorname{var}(\hat{\beta}_1) = \frac{\left(\sigma_b^2 + \dfrac{\sigma_w^2}{k}\right)}{S_{xx}} \tag{3}$$

The standard error of $\hat{\beta}_1$ can be estimated from equation 3 by inserting REML estimators of the variance components, or equivalently by replacing the numerator of equation 3 by the between mean squares (MSB) from Table 1 divided by $k$, i.e.,

*Table 1*

*Analysis of Variance for the Linear Mixed Model A*

| Source | *d.f.* | SS | MS | EMS |
|---|---|---|---|---|
| Slope | 1 | $k\hat{\beta}_1^2 S_{xx}$ | $k\hat{\beta}_1^2 S_{xx}$ | $k\hat{\beta}_1^2 S_{xx} + k\sigma_b^2 + \sigma_w^2$ |
| Between | $n - 2$ | $k\Sigma_i(\bar{y}_i - \hat{y}_i)^2$ | $k\Sigma_i(\bar{y}_i - \hat{y}_i)^2/(n-2)$ | $k\sigma_b^2 + \sigma_w^2$ |
| Within | $n(k-1)$ | $\Sigma_{ij}(y_{ij} - \bar{y}_i)^2$ | $\Sigma_{ij}(y_{ij} - \bar{y}_i)^2/n(k-1)$ | $\sigma_w^2$ |

*Note*: $S_{xx} = \Sigma_i(x_i - \bar{x})^2$, $\bar{y}_i = \Sigma_j y_{ij}/k$, $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1(x_i - \bar{x})$. Between sum of squares was adjusted for the slope and intercept, which are given by $\hat{\beta}_0 = \bar{\bar{y}} = \Sigma_{ij} y_{ij}/nk$ and $\hat{\beta}_1 = \Sigma_i(x_i - \bar{x})(\bar{y}_i - \bar{\bar{y}})/S_{xx}$, respectively.

$$se(\hat{\beta}_1) = \sqrt{\frac{\Sigma_i(\bar{y}_i - \hat{y}_i)^2/(n-2)}{S_{xx}}} = \sqrt{\frac{\text{MSB}}{kS_{xx}}} \quad (4)$$

If a confidence interval for the slope is desired it is given by the usual $t$ interval,

$$\hat{\beta}_1 \pm t_{1-\alpha/2,(n-2)} se(\hat{\beta}_1) \quad (5)$$

Note that the correct *d.f.* for $t$ is $n - 2$ not $n(k - 1)$. A new SAS® (version 6.07, SAS 1992) procedure PROC MIXED seems to give all computations correct except for the denominator *d.f.* for the $t$ test of $\hat{\beta}_1$. PROC MIXED erroneously used $n(k - 1)$ rather than $n - 2$ for the $t$ test.

## A Model Based on Mean Values

Sometimes practitioners use a model based on mean values $\bar{y}_i$, with the usual independent error assumption of the mean values,

$$\text{Model B: } \bar{y}_i = \beta_0 + \beta_1(x_i - \bar{x}) + \varepsilon_i, \ i = 1 \ldots n, \quad (6)$$

where $\varepsilon_i$ has mean zero and variance $\sigma^2_\varepsilon$. This is a familiar simple linear regression model applied to means. The independence assumption is satisfied from model A since observations in separate classes are in fact independent. It can be shown (see appendix) that $\sigma^2_\varepsilon$ is equal to $\sigma^2_b + \sigma^2_w/k$ from model A.

It is easy to see that the least squares estimators of the intercept $\beta_0$ ($\bar{\bar{y}}$ in both cases) and the slope $\beta_1$ in this model are exactly the same as that of model A above. The analysis of variance table for model B is given in Table 2.

Note that $\Sigma_i(\bar{y}_i - \hat{y}_i)^2/(n-2)$ is the mean squared error from model B and the usual unbiased estimator of $\sigma^2_\varepsilon$. We have seen that it is also the estimator used for the numerator of $se(\hat{\beta}_1)$ in model A (equation 4). Therefore, the standard error of the slope for model B is exactly the same as that for model A. The $n - 2$ *d.f.* $t$-test for the slope in model B is equivalent to the $F$ test in model A. Hence, everything seems to be in agreement between models A and B. Different from what some practitioners might think, model B does not lose any degrees of freedom; it does the correct analysis. Then what does model A have in addition to what model B has already? The individual variance components and the intraclass correlation can be estimated from model A but not from model B. In other words, if the practitioner's main objective is to conduct inference on the fixed effects, model B will be sufficient. Note that if we have an unequal number of determinations at each concentration (unbalanced design), the least squares estimator $\hat{\beta}_1$ of the slope from model A may not be the same as that from model B. In fact, it will be weighted by the number of determinations, $k_i$, for each $i$th class, i.e.,

$$\hat{\beta}_1 = \Sigma_i k_i(x_i - \bar{x})(\bar{y}_i - \bar{\bar{y}})/\Sigma_i k_i(x_i - \bar{x})^2 \quad (7)$$

Model B does not use any intraclass information directly. For an unbalanced design, the analysis will be different from that of model A, especially when the degree of the unbalancedness is severe.

## A Model Based on Complete Independence of the Observations

As we have seen in model A, the $y_{ij}$'s are not completely independent. However, some practitioners often erroneously adopt a model based on complete independence of the errors $d_{ij}$, which have mean 0 and common variance $\sigma^2_d$, as follows:

$$\text{Model C: } y_{ij} = \beta_0 + \beta_1(x_i - \bar{x}) + d_{ij}, \ i = 1 \ldots k \quad (8)$$

Then, statistical inference is conducted on the slope based on an $nk - 2$ degrees of freedom $t$-test . It is easy to show that for balanced designs, the estimates of $\beta_0$ and $\beta_1$ will be exactly the same as in the other two mod-

***Table 2***

*Analysis of Variance for Model B*

| Source | *d.f.* | SS | MS | EMS |
|---|---|---|---|---|
| Slope | 1 | $k\hat{\beta}_1^2 S_{xx}$ | $k\hat{\beta}_1^2 S_{xx}$ | $k\hat{\beta}_1^2 S_{xx} + k\sigma^2_b + \sigma^2_w$ |
| Between | $n - 2$ | $\Sigma_i(\bar{y}_i - \hat{y}_i)^2$ | $\Sigma_i(\bar{y}_i - \hat{y}_i)^2/(n-2)$ | $\sigma^2_b + \sigma^2_w/k$ |

*Note*: $S_{xx} = \Sigma_i(x_i - \bar{x})^2$, $\bar{y}_i = \Sigma_j y_{ij}/k$, $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1(x_i - \bar{x})$. $\hat{\beta}_0 = \bar{\bar{y}} = \Sigma_{ij} y_{ij}/nk$ and $\hat{\beta}_1 = \Sigma_i(x_i - \bar{x})(\bar{y}_i - \bar{\bar{y}})/S_{xx}$, respectively.

*Table 3*

*Analysis of Variance for Model C*

| Source | *d.f.* | SS | MS | EMS Under Model A |
|---|---|---|---|---|
| Slope | 1 | $k\hat{\beta}_1^2 S_{xx}$ | $k\hat{\beta}_1^2 S_{xx}$ | $k\hat{\beta}_1^2 S_{xx} + k\sigma_b^2 + \sigma_w^2$ |
| Within | $nk - 2$ | $\Sigma_{ij}(\bar{y}_{ij} - \hat{y}_i)^2$ | $\Sigma_{ij}(\bar{y}_{ij} - \hat{y}_i)^2/(nk - 2)$ | $m\sigma_b^2 + \sigma_w^2/k$ |

*Note*: $S_{xx} = \Sigma_i(x_i - \bar{x})^2$, $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1(x_i - \bar{x})$. $\hat{\beta}_0 = \bar{\bar{y}} = \Sigma_{ij} y_{ij}/nk$ and $\hat{\beta}_1 = \Sigma_i(x_i - \bar{x})(\bar{y}_i - \bar{\bar{y}})/S_{xx}$. $m$ is given in equation 10 and in the appendix.



els. The difference between model C and models A and B lies in the estimate of the standard error of $\hat{\beta}$ and the corresponding degrees of freedom. As is well known, the confidence interval for $\beta_1$ is given by

$$\hat{\beta}_1 \pm t_{1-\alpha/2,(nk-2)}\sqrt{\frac{\hat{\sigma}_d^2}{kS_{xx}}} \tag{9}$$

where $\hat{\sigma}_d^2$ is given by the within mean squares (MSW) from the analysis of variance in Table 3.

The last column of Table 3 shows the expected value of $\hat{\sigma}_d^2$ under the assumption that model A is the correct model. The expected value of $\hat{\sigma}_d^2$ shown in Table 3 is derived in the appendix. Since

$$m = \frac{(n-2)k}{nk-2} \tag{10}$$

and $m \leq 1$, we can see that $\hat{\sigma}_d^2$ underestimates $k\sigma_b^2 + \sigma_w^2$ unless $k = 1$, in which case it is the same as MSB and models A and C are equivalent. In general, however $k > 1$ and the expected value of $\hat{\sigma}_d^2$ is much smaller than the expected value of MSB in model A, resulting in undue inflation of the $F$ ratio under model C.

The length of a confidence interval is a function of the $t$ values and *d.f.*, as well as the estimates of the variance components. Hence, it is of interest to compare the expected lengths of the two confidence intervals obtained under models A and C, assuming that model A is correct. We compared the expected squared length's of the two intervals because of the ease of the algebra required for computing expectations. It can be shown that the ratio, $r$, of the expected squared lengths of the two confidence intervals from models A and C is given as (see appendix for derivation)

$$r = \frac{F_{1-\alpha,(1,n-2)}(k\sigma_b^2 + \sigma_w^2)}{F_{1-\alpha,(1,nk-2)}(m\sigma_b^2 + \sigma_w^2)} \tag{11}$$

The ratio $r$ increases as $k$ and/or the intraclass correlation $\rho$ increase. If $n$ increases, $r$ decreases. In other words, the length of the confidence interval from model C becomes much shorter than that for the true model A as $k$ or $\rho$ increase, but less so for larger $n$. In Table 4, the ratio $r$ was constructed for $\alpha = 0.05$, $\rho = 0.8$, $n = 7$, and $\sigma_b^2 + \sigma_w^2 = 1$. Although the metric is in squared units, it is clear that the ratio gets uncomfortably large as $k$ increases.

*Table 4*

*The Expected Squared Lengths of the Two Confidence Intervals from Models A and C as a Function of k ($\alpha = 0.05$, $\rho = 0.8$, $n = 7$ and $\sigma_b^2 + \sigma_w^2 = 1$)*

| $k$ | $m(n,k)$ | $k\sigma_b^2 + \sigma_w^2$ | $m\sigma_b^2 + \sigma_w^2$ | Numerator | Denominator | $r$ |
|---|---|---|---|---|---|---|
| 1 | 1.00 | 1.0 | 1.00 | 6.61 | 6.61 | 1.00 |
| 2 | 0.83 | 1.8 | 0.87 | 11.89 | 4.11 | 2.89 |
| 3 | 0.79 | 2.6 | 0.83 | 17.18 | 3.64 | 4.72 |
| 4 | 0.77 | 3.4 | 0.82 | 22.47 | 3.45 | 6.52 |
| 5 | 0.76 | 4.2 | 0.81 | 27.75 | 3.34 | 8.32 |
| 6 | 0.75 | 5.0 | 0.80 | 33.04 | 3.27 | 10.11 |
| 7 | 0.74 | 5.8 | 0.80 | 38.33 | 3.22 | 11.90 |
| 8 | 0.74 | 6.6 | 0.79 | 43.61 | 3.19 | 13.69 |
| 9 | 0.74 | 7.4 | 0.79 | 48.90 | 3.16 | 15.48 |
| 10 | 0.74 | 8.2 | 0.79 | 54.18 | 3.14 | 17.26 |

*Table 5*

*The Expected Squared Lengths of the Two Confidence Intervals from Models A and C as a Function of ρ ($\alpha = 0.05$, $n = 7$, $k = 5$, and $\sigma_b^2 + \sigma_w^2 = 1$)*

| Rho(ρ) | Numerator | Denominator | Ratio |
|---|---|---|---|
| 0.0 | 6.61 | 4.14 | 1.60 |
| 0.1 | 9.25 | 4.04 | 2.29 |
| 0.2 | 11.89 | 3.94 | 3.02 |
| 0.3 | 14.54 | 3.84 | 3.79 |
| 0.4 | 17.18 | 3.74 | 4.60 |
| 0.5 | 19.82 | 3.64 | 5.45 |
| 0.6 | 22.47 | 3.54 | 6.35 |
| 0.7 | 25.11 | 3.44 | 7.31 |
| 0.8 | 27.75 | 3.34 | 8.32 |
| 0.9 | 30.40 | 3.24 | 9.39 |
| 1.0 | 33.04 | 3.14 | 10.54 |

For fixed $n = 7$ and $k = 5$, the intraclass correlation ρ is changed from 0 to 1 by 0.1 increments and results are given in Table 5. As the intraclass correlation increases, $\sigma_b^2$ increases both in the numerator and the denominator but large $k$ (=5) compared to $m$ (=0.76) makes the ratio larger. For a study with a small $n$ with large $k$, together with a large intraclass correlation, the confidence interval from model C is very inaccurate.

## A Numerical Example

A set of data is generated from a population of $\beta_0 = 10$, $\beta_1 = 1$, $\sigma_b^2 = 0.8$, $\sigma_w^2 = 0.2$, $n = 7$, and $k = 5$, and is summarized in Table 6.

We used the SAS procedure PROC GLM and its type I sums of squares to generate the ANOVA results in Table 7.

Output from SAS PROC MIXED showed its REML estimators of the variance components to be exactly the same as the above ANOVA estimators based on type I sums of squares. The test of the slope can be done either from Table 7 or by the $t$ test from model B in Table 8.

PROC MIXED has an identical result except that it has a *d.f.* of 28 in the denominator, which is incorrect for computation of confidence intervals. The same *d.f.* of 28 was given for the $F$ test of the slope. The correct *d.f.* should be 5 for the denominator. Running model C using SAS PROC REG gave identical intercept and slope estimates, but the standard errors were underestimated (0.323 and 0.040 respectively) and the *d.f.* (=33) was inflated as expected. It showed a within mean square of 0.900 based on the 33 *d.f.*, which is an estimate of $(25/33)\sigma_b^2 + \sigma_w^2$ under model A, as can be seen from the EMS column of Table 3. The correct 95% confidence interval for $\beta_1$ should be $1.05 \pm$

*Table 6*

*A Simulated Data for Numerical Example*

| Sample Number | Concentration | Determinations | | | | |
|---|---|---|---|---|---|---|
| 1 | 1 | 8.78 | 9.24 | 9.25 | 8.91 | 9.00 |
| 2 | 3 | 13.57 | 13.62 | 13.30 | 13.12 | 14.11 |
| 3 | 5 | 14.48 | 14.81 | 14.45 | 14.45 | 14.24 |
| 4 | 7 | 16.97 | 17.63 | 17.72 | 17.30 | 17.74 |
| 5 | 9 | 19.97 | 19.74 | 19.87 | 18.85 | 18.99 |
| 6 | 11 | 20.06 | 19.65 | 19.05 | 20.06 | 20.58 |
| 7 | 13 | 22.57 | 23.12 | 22.38 | 22.98 | 23.23 |

*Table 7*

*Analysis of Variance for the Simulated Data*

| Source | *df* | SS | MS | EMS | Variance Components |
|---|---|---|---|---|---|
| Slope | 1 | 623.82 | 623.82 | | |
| Between | 5 | 25.41 | 5.08 | $5\sigma_b^2 + \sigma_w^2$ | 0.986 |
| Within | 28 | 4.27 | 0.15 | $\sigma_w^2$ | 0.153 |
| Total | 34 | 653.51 | | | 1.139 |

*Table 8*

*Parameter Estimates for the Simulated Data*

| Variable | *d.f.* | Estimate | S.E. | *t* value |
|---|---|---|---|---|
| Intercept | 1 | 9.29 | 0.768 | 12.10 |
| Slope | 1 | 1.05 | 0.095 | 11.08 |

(2.571)(0.095), or 1.05 ± 0.24. An incorrect interval based on model C is 1.05 ± (2.030)(0.040), or 1.05 ± 0.08. The width of the model C confidence interval is only one-third of the correct one.

## SIMULATION STUDIES

### Confidence Intervals for $\beta_1$

It is of interest to compare the confidence interval in equation 9 to the correct interval (equation 5) in a simulation study where the observations are generated from the correct model A, but the confidence bound is computed both ways. Under the simulation a confidence interval for $\beta_1$ using $z$-scores can be computed using the formula for $(\hat{\beta}_1)$ from equation 3 since true values of the variance components are known:

$$\hat{\beta}_1 \pm z_{1-\alpha/2}\sqrt{\frac{\left(\sigma_b^2 + \frac{\sigma_w^2}{k}\right)}{S_{xx}}} \qquad (12)$$

It is also of interest to compute the experimental confidence coefficient of three methods (equations 5, 9, and 12) where equation 12 serves as "control." If $\rho = 0$, or equivalently if $\sigma_b^2 = 0$, then model A reduces to model C and the statistical test for the slope gains power in terms of more *d.f.* for the denominator ($nk - 2$ instead of $n - 2$). It will be interesting to find out through simulation the "cutoff" point, if any, where model C analysis is preferred to model A analysis.

Data were generated from model A when $\beta_0 = 0$, $\beta_1 = 1$, $n = 7$, and $k = 5$. For each data set, confidence intervals were generated using equations 5, 9, and 12. Note that the interval in equation 12 requires known variance components and hence is not useful in practice, but is included here in simulation for comparison. It acts as a "control," i.e., it should hold the nominal confidence coefficient very well. Once the intervals are computed, they are checked to determine if they contain the true parameter value within the interval. The above steps are replicated 1000 times for each $\rho$, with $\rho$ changed from 0 to 1 by 0.1 increments. We set $\sigma_b^2 + \sigma_w^2 = 1$ without loss of generality, and varied $\sigma_b^2$ from 0 to 1. Table 9 summarizes the simulation results. Note that the interval based on equation 5 holds the nominal 95% confidence very well. It is as good as "control." However, the interval from equation 9 has experimental confidence coefficients ranging from 0.48 ($\rho = 1$) to 0.95 ($\rho = 0$). Unless intraclass correlation is very small ($\rho < 0.1$), the intervals based on model C are unrealistically narrow. For a moderate intraclass correlation of 0.5 or more, model C coverage probability is only 0.67 or less.

### Effect on Stability

Stability analysis of single batches of a pharmaceutical product are frequently conducted. These studies follow model A, but are typically analyzed using model C, the completely independent model. Simulation studies were conducted to determine the effect of erroneous application of model C analysis to model A data. Data were generated from model A when $\beta_0 = 100$, $\beta_1 = -0.25$, $n = 7$ and $k = 5$. For each data set, stability was determined using the lower 95% one-sided confidence interval from equations 5, 9, and 12, as described by Peace (1). Again, note that the interval in equation 12 requires known variance components and hence is not useful in practice but is included here in simulation for comparison. It acts as a "control." Once the shelf lives are computed, each is checked against the

*Table 9*

*Simulation Results Summarizing Coverage Probabilities of Confidence Intervals for the Three Models*

| Correlation (r) | Model B (equation 5) | Model C (equation 9) | Control (equation 12) |
|---|---|---|---|
| 0.0 | 0.954 | 0.951 | 0.949 |
| 0.1 | 0.958 | 0.913 | 0.954 |
| 0.2 | 0.941 | 0.856 | 0.965 |
| 0.3 | 0.957 | 0.802 | 0.946 |
| 0.4 | 0.951 | 0.746 | 0.942 |
| 0.5 | 0.948 | 0.688 | 0.945 |
| 0.6 | 0.954 | 0.666 | 0.958 |
| 0.7 | 0.944 | 0.626 | 0.959 |
| 0.8 | 0.955 | 0.582 | 0.954 |
| 0.9 | 0.942 | 0.505 | 0.949 |
| 1.0 | 0.938 | 0.483 | 0.949 |

*Note*: Simulations were conducted using SAS IML with the RANNOR standard normal random number generator.

control. The above steps are replicated 1000 times for each $\rho$, with $\rho$ changed from 0 to 1 by 0.1 increments. We set $\sigma_b^2 + \sigma_w^2 = 1$ without loss of generality, and varied $\sigma_b^2$ from 0 to 1. Table 10 summarizes the simulation results. Model B results in shelf-life estimates that are conservative, but on average are within 5% of the control shelf lives. Model C, unless $\rho << 0.1$, results in shelf lives that overestimate the true shelf life. As $\rho$ increases toward 1.0, shelf live estimates increase to ≈30% overestimation.

## CONCLUSIONS

When a straight line is fit to data replicated several times at each $x$-value (model A), we have seen that practitioners either use averages obtained at each $x$-value (model B), or treat every replicate at the same $x$-value as independent (model C). If the experiment is balanced in the sense that we have the same number of replicates at each $x$-value, the average model (model B) was shown to be equivalent to the proposed correct mixed linear model with nested error structures (model A), except that estimation and inference about individual between and within variance components cannot be conducted under model B. However, if studies such as HPLC calibration assays are being conducted and interval estimation of the slope and/or the fitted value is desired, the independent model (model C) is irrelevant. A simulation study showed that unless the intraclass correlation is very close to zero (far less than 0.1), the independent model gives unrealistically narrow confidence intervals (i.e., tends to magnify the type I error) and therefore should not be used. The average model (model B) held the nominal confidence coefficient of the slope parameter very well in our simulation study. In the analysis of pharmaceutical stability data from an individual batch, the independent model (model C) is again irrelevant. A simulation study showed that unless the intraclass correlation is very close to zero (far less than 0.1), model C results in unrealistic overestimation (as much as 30%) of the expiration period of drugs. The simulation study also showed that the average model (model B) is acceptably conservative in the sense that it underestimates the true expiration period slightly (less than 5%).

## APPENDIX

### Standard Error of $\hat{\beta}_1$ under Model A

Recall $\hat{\beta}_1 = \Sigma_i(x_i - \bar{x})(\bar{y}_i - \bar{\bar{y}})/S_{xx}$. Then

$$v(\hat{\beta}_1) = v(\Sigma_i(x_i - \bar{x})(\bar{y}_i - \bar{\bar{y}})/S_{xx}) =$$

$$= \left(\frac{1}{S_{xx}}\right)^2 \left(\Sigma_i(x_i - \bar{x})^2 v(\bar{y}_i - \bar{\bar{y}}) + 2\sum\sum_{i<j} (x_i - \bar{x})(x_j - \bar{x}) \operatorname{cov}(\bar{y}_i - \bar{\bar{y}})(\bar{y}_j - \bar{\bar{y}})\right)$$

*Table 10*

*Simulation Results Summarizing Shelf Lives from Models B and C*

| ($\rho$) | Model B (equation 5) | Model C (equation 9) | Control (equation 12) | $\left(\frac{B}{Control}\right) \times 100$ | $\left(\frac{C}{Control}\right) \times 100$ |
|---|---|---|---|---|---|
| 0.0 | 32.4 | 33.7 | 33.9 | 96.4 | 100.3 |
| 0.1 | 31.7 | 34.0 | 33.2 | 97.0 | 104.0 |
| 0.2 | 30.8 | 34.2 | 32.5 | 96.3 | 106.9 |
| 0.3 | 30.1 | 34.3 | 31.8 | 96.0 | 109.3 |
| 0.4 | 29.4 | 34.6 | 31.3 | 95.3 | 112.2 |
| 0.5 | 28.9 | 34.6 | 30.8 | 95.2 | 114.1 |
| 0.6 | 28.6 | 35.1 | 30.5 | 95.8 | 117.6 |
| 0.7 | 28.3 | 35.4 | 30.2 | 95.9 | 120.0 |
| 0.8 | 27.9 | 35.6 | 29.7 | 95.9 | 122.3 |
| 0.9 | 27.6 | 36.1 | 29.5 | 95.9 | 125.5 |
| 1.0 | 27.5 | 36.8 | 29.4 | 96.8 | 129.4 |

*Note*: Simulations were conducted using SAS IML with the RANNOR standard normal random number generator.

$$= \left(\frac{1}{S_{xx}}\right)^2 \left(\Sigma_i (x_i - \bar{x})^2 \frac{n-1}{n} v(\bar{y}_i) + 2\sum\sum_{i<j} (x_i - \bar{x})(x_j - \bar{x})\left(-\frac{1}{n} v(\bar{y}_i)\right)\right)$$

$$= \left(\frac{1}{S_{xx}}\right)^2 \left(\frac{v(\bar{y}_i)}{n}\right)\left((n-1)\Sigma_i (x_i - \bar{x})^2 - 2\sum\sum_{i<j} (x_i - \bar{x})(x_j - \bar{x})\right)$$

$$= \left(\frac{1}{S_{xx}}\right)^2 \left(\frac{v(\bar{y}_i)}{n}\right)\left(n\Sigma_i (x_i - \bar{x})^2 - \left(\Sigma_i (x_i - \bar{x})\right)^2\right)$$

$$= \left(\frac{1}{S_{xx}}\right)^2 v(\bar{y}_i)\Sigma_i (x_i - \bar{x})^2 = \left(\frac{1}{S_{xx}}\right)^2 v(\bar{y}_i) S_{xx}$$

$$= \frac{v(\bar{y}_i)}{S_{xx}} = \frac{\sigma_b^2 + \frac{\sigma_w^2}{k}}{S_{xx}}$$

### Standard Error of under Model B

From model A, the variance of the mean $y_{ij}$ is $v(\bar{y}_i) = v(b_i) + v(\bar{w}_i)$, where

$$\bar{w}_i = \frac{1}{k}\sum_j w_{ij}.$$

So

$$v(\bar{y}_i) = \sigma_b^2 + \frac{\sigma_w^2}{k},$$

and thus

$$v(\hat{\beta}_1) = \frac{v(\bar{y}_i)}{S_{xx}} = \frac{\sigma_b^2 + \frac{\sigma_w^2}{k}}{S_{xx}}$$

### Expected Mean Squares of Residuals from Model C Under the True Model A

Let

$$\text{SSE} = \sum_{ij} (y_{ij} - \hat{y}_i)^2, \quad \text{SSB} = k\sum_i (\bar{y}_i - \hat{y}_i)^2,$$

and

$$\text{SSW} = \sum_{ij} (y_{ij} - \bar{y}_i)^2$$

Then

$$E(\text{SSE}) = E(\text{SSB}) + E(\text{SSW})$$
$$= (n-2)(k\sigma_b^2 + \sigma_w^2) + n(k-1)\sigma_w^2 \text{ (from Table 1)}$$
$$= (n-2)k\sigma_b^2 + (nk-2)\sigma_w^2.$$

Therefore,

$$E(\hat{\sigma}_d^2) = \frac{(n-2)k}{nk-2}\sigma_b^2 + \sigma_w^2 = m\sigma_b^2 + \sigma_w^2$$

where

$$m = m(n,k) = \frac{(n-2)k}{nk-2}$$

### The ratio of the expected squared length of the two confidence intervals from models A and C

$$r = \frac{t^2_{1-\alpha/2,(n-2)} E(\text{MSB})}{t^2_{1-\alpha/2,(n-2)} E(\text{MSB})}, \text{ from (4) and (9),}$$

$$= \frac{F_{1-\alpha,(1,n-2)}(k\sigma_b^2 + \sigma_w^2)}{F_{1-\alpha,(1,nk-2)}(m\sigma_b^2 + \sigma_w^2)}, \text{ from Tables 1 and 3.}$$

## ACKNOWLEDGMENTS

The authors would like to thank Drs. Paul McAllister and John Peterson for their helpful review, which resulted in considerable improvements.

## REFERENCES

1. K. Peace, in Biopharmaceutical Statistics for Drug Development, Marcel Dekker Inc., New York, 1988, p. 228.
2. Guideline for Submitting Documentation for the Stability of Human Drugs and Biologics, Center for Drugs and Biologics FDA Department of Health and Human Services, 1987.
3. T. E. Norwood, Drug Dev. Ind. Pharm., 12, 553 (1986).
4. W. A. Fuller and G. E. Battese, J. Am. Stat. Assoc., 68, 626 (1973).
5. G. A. Milliken and M. Albohali, Am. Stat., 38, 298 (1984).
6. S. Puntnanen and G. P. H. Styan, Am. Stat., 43, 153 (1989).
7. F. S. Acton, Analysis of Straight-Line Data, Wiley, New York, 1959.
8. D. A. Harville, J. Am. Stat. Assoc., 72, 320 (1977).
9. SAS Technical Report P-229, SAS/STAT Software: Changes and Enhancements, Release 6.07, SAS Institute, Inc., Cary, North Carolina, 1992.