Power of Analysis of Variance for Assessing Batch-Variation of Stability Data of Pharmaceuticals

Sumie YOSHIOKA,*[,a] Yukio ASO,[a] Shigeo KOJIMA,[a] and Alain LI WAN PO[b]

National Institute of Health Sciences,[a] 1–18–1 Kamiyoga, Setagaya, Tokyo 158, Japan and The School of Pharmacy, The University of Nottingham,[b] University Park, Nottingham, NG7 2RD, U. K.

Received May 27, 1996; accepted June 25, 1996

The present study investigates the effect of batch-variation on the evaluation of stability data obtained from three batches of pharmaceutical products. Using stability data generated by the Monte Carlo method, the power of analysis of variance for assessing batch-variation was determined as functions of significance level and assay error. The results obtained from degradation data simulated for a product exhibiting zero-order degradation of 0.2%/month indicate that a significance level of approximately 0.25 is necessary in order to obtain a minimum detectable difference of 25% if the beta error is fixed at 20%. Assay methods with smaller margins of error are recommended, as is repeating the assay in order to reduce assay error down to approximately 0.5% standard deviation. In addition, the validity of the 0.25 significance level, which is recommended in the ICH (International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use) Tripartite guideline, is discussed.

Key words stability; batch-variation; shelf-life; analysis of variance; significance level

The shelf-life of pharmaceutical products is generally estimated based on stability data obtained from three batches of the product. If significant variation is observed in the stability of these three batches, the shelf-life should be determined based on the stability data of the most unstable batch. The ICH Tripartite guideline of stability testing for new drug substances and products[1] describes that in the absence of significant variation among the batches, the stability data from the three batches can be combined in order to estimate a shelf-life that applies to all of the batches. This batch-combination analysis yields a longer estimate of shelf-life (the 95% lower confidence limit of the time until drug product degrades below the lower specified value) because of a larger number of data to be considered. Whether the batch-combination analysis can be applied to a specific product is determined by performing an analysis of variance (ANOVA) on the stability among batches. The ICH guideline recommends the use of a 0.25 significance level for the ANOVA, but data are not available regarding how strictly the batch-variation of stability of pharmaceuticals is assessed by ANOVA using a 0.25 significance level.

The present paper describes the relationship between the significance level and the power of the ANOVA. Using data generated by the Monte Carlo method assuming that a pharmaceutical exhibits zero-order degradation of 0.2%/month, the power of the test (probability at which batch-variation is taken as significant) was determined as a function of significance level. In addition, the shelf-life estimates obtained by the batch-combination analysis are compared with those obtained by ordinary analysis.

Experimental

Stability data were generated using the Monte Carlo method under the assumption that degradation of the product can be described by zero-order kinetics. The slope of the degradation curve was assumed to be 0.2%/month for two batches, and a 0%, 10% or 25% larger slope was assumed for the remaining batch. Theoretical drug contents at 0, 3, 6, 9, 12, 18, 24 and 36 months, expressed as a percentage of the content at time zero, were obtained from these degradation curves. Random numbers selected from a normal distribution with a mean of zero and a standard deviation ranging from 0.4 to 2.0% were added to the theoretical data in order to obtain experimental data with an assay error. Stability data are usually considered to contain assay errors of not less than 0.5%.

Using the degradation data simulated for three batches (drug content c_{ji} at time t_{ji} for batch j and data number i), the uniformity of regression (slope and intercept) was tested by ANOVA based on F distribution with 4 degrees of freedom for the numerator and $\phi_1+\phi_2+\phi_3$ degrees of freedom for the denominator, as represented by Eq. 1.

$$F=\frac{(Q_2-Q_1)/4}{Q_1/(\phi_1+\phi_2+\phi_3)}\geq F(4,\ \phi_1+\phi_2+\phi_3;\ \alpha) \tag{1}$$

where α is the significance level, and

$$Q_1=S_{e1}+S_{e2}+S_{e3}$$

$$Q_2=S_T(cc)-b_TS_T(tc)$$

$$\phi_1=n_1-2 \quad \phi_2=n_2-2 \quad \phi_3=n_3-2.$$

n_1, n_2, and n_3 are the number of data for batch 1, 2, and 3, respectively. S_{e1}, S_{e2}, S_{e3}, $S_T(cc)$, and $b_TS_T(tc)$ were calculated according to the following equations:

$$S_j(tt)=\sum_{i=1}^{n_j}(t_{ji}-\bar{t}_j)^2 \quad (j=1, 2, 3)$$

$$S_j(cc)=\sum_{i=1}^{n_j}(c_{ji}-\bar{c}_j)^2$$

$$S_j(tc)=\sum_{i=1}^{n_j}(t_{ji}-\bar{t}_j)(c_{ji}-\bar{c}_j)$$

$$b_j=S_j(tc)/S_j(tt)$$

$$S_{ej}=S_j(cc)-b_jS_j(tc)$$

$$\bar{t}=\frac{1}{n_1+n_2+n_3}\sum_{j=1}^{3}\sum_{i=1}^{n_j}t_{ji}=\frac{n_1\bar{t}_1+n_2\bar{t}_2+n_3\bar{t}_3}{n_1+n_2+n_3}$$

$$\bar{c}=\frac{1}{n_1+n_2+n_3}\sum_{j=1}^{3}\sum_{i=1}^{n_j}c_{ji}=\frac{n_1\bar{c}_1+n_2\bar{c}_2+n_3\bar{c}_3}{n_1+n_2+n_3}$$

$$S_T(tt)=\sum_{j=1}^{3}\sum_{i=1}^{n_j}(t_{ji}-\bar{t})^2$$

$$S_T(cc)=\sum_{j=1}^{3}\sum_{i=1}^{n_j}(c_{ji}-\bar{c})^2$$

* To whom correspondence should be addressed.

$$S_T(tc)=\sum_{j=1}^{3}\sum_{i=1}^{n_j}(t_{ji}-\bar{t})(c_{ji}-\bar{c})$$

$$b_T=S_T(tc)/S_T(tt)$$

ANOVA was performed using significance levels from 0.01 to 0.5. The power of the test was calculated for each significance level as the fraction of 500 data sets that show a statistically significant difference. For each set of data, the shelf-life was calculated using Woolfe's equation (Eq. 2)[2] according to two methods: the batch-combination method in which the data of all three batches were used to estimate shelf-life, and the batch-separation method in which only the data of the most unstable batch were used. The lower specified drug content of the product was assumed to be 90%, and the shelf-life (t_{90}) was calculated as t_L at time t when the regression curve intersected 90%.

$$t_L=t+\left\{(t-\bar{t})g\pm\frac{t'\sqrt{V_e}}{b}\sqrt{\frac{(t-\bar{t})^2}{S_{tt}}+\left(1+\frac{1}{n}\right)(1-g)}\right\}\frac{1}{1-g} \quad (2)$$

$$g=\frac{t'^2V_e}{b^2S_{tt}}$$

where n, b, t, V_e, S_{tt}, and t' are the number of data, slope, the mean of t_i ($i=1\cdots n$), mean squares, sum of squares, and Student's t value, respectively.

Generation of experimental data, ANOVA,[3] and calculation of shelf-life were performed using BASIC programs.

Results and Discussion

In order to consider how strictly the batch-variation of stability data is assessed by ANOVA using a 0.25 significance level, the power of the ANOVA was calculated as a function of significance level, using degradation data simulated for a pharmaceutical product having a shelf-life commonly observed for pharmaceuticals. A zero-order degradation curve with a slope of 0.2%/month was assumed to yield a shelf-life of approximately 40 months. Figure 1 shows the power of ANOVA on the batch-variation of the simulated data with 10 and 25% differences in slope, and 0.5% and 1.0% assay errors. For the purpose of assessing a difference, beta error (probability at which the difference is taken as not significant when it is) is usually considered to be 20%. At an assay error of 0.5%, the minimum detectable difference is approximately 25% when the beta error and significance level are fixed at 20% and 0.25, respectively. In contrast, increasing the assay error from 0.5% to 1% decreased the power markedly, resulting in a low power of approximately 50%, even at a significance level of 0.25. The effect of assay error on the power when one of the batches shows a 25% larger degradation slope is shown in Fig. 2. The power decreases markedly with increasing assay error, and approaches that of the case when no variation is observed among the batches, when the assay error exceeds 1%.

Shelf-life was calculated from data simulated such that one batch showed a 25% larger degradation slope than the others, and data simulated such that no variation existed among batches. The results are shown in Fig. 3. The shelf-life was estimated using the data of all three batches in the batch-combination method. In contrast, the

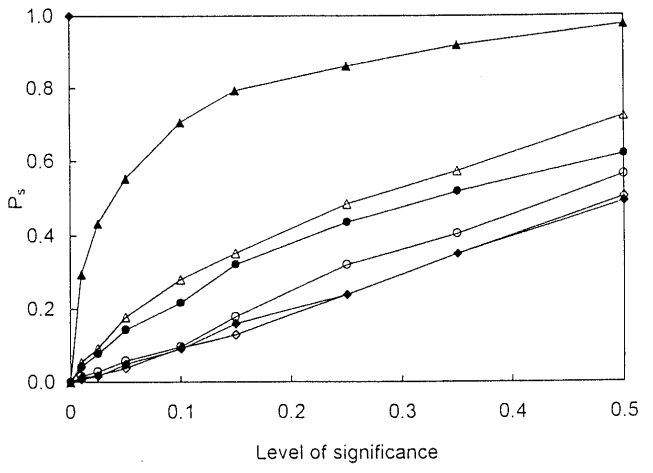

Fig. 1. Effect of Significance Level on the Power of ANOVA, P_s (Probability at Which Batch-Variation is Taken as Significant)

◇ Slope of degradation curve: 0.2%/month for all batches, assay error (standard deviation, S.D.): 1.0%; ◆ 0.2%/month for all batches, S.D. 0.5%; ○ 0.22%/month for one batch and 0.2%/month for the other two batches, S.D. 1.0%; ● 0.22%/month for one batch and 0.2%/month for the other two batches, S.D. 0.5%; △ 0.25%/month for one batch and 0.2%/month for the other two batches, S.D. 1.0%; ▲ 0.25%/month for one batch and 0.2%/month for the other two batches, S.D. 0.5%.

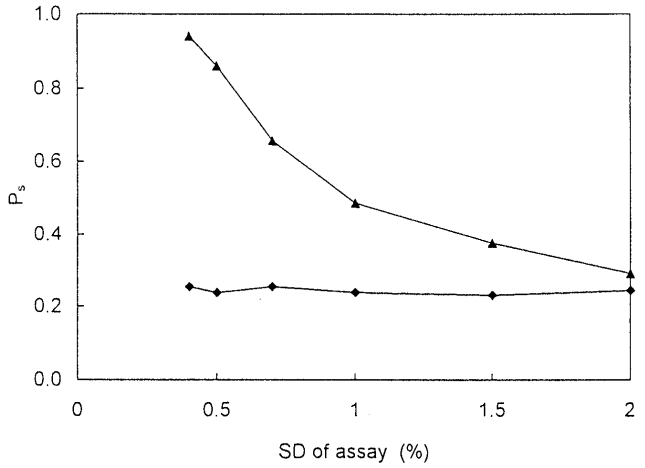

Fig. 2. Effect of Assay Error on the Probability at Which Batch-Variation is Taken as Significant

Slope of degradation curve: 0.2%/month for all batches (◆), and 0.25%/month for one batch and 0.2%/month for the other two batches (▲). Significance level: 0.25.

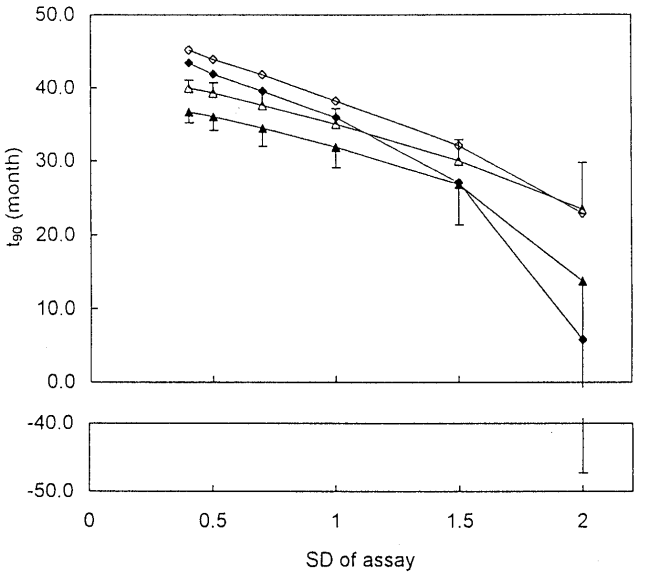

Fig. 3. Effect of Assay Error on Shelf-Life (t_{90}) Estimated by Batch-Combination Analysis (◇△) and by Batch-Separation Analysis (◆▲)

Slope of degradation curve: 0.2%/month for all batches (◇◆), and 0.25%/month for one batch and 0.2%/month for the other two batches (△▲). Significance level: 0.25.

shelf-life was estimated using only the data of the most unstable batch in the batch-separation method. The shelf-life estimated using either method decreases as the assay error increases. The standard deviation of the shelf-life estimates increases with increasing assay error. Shelf-life estimated from the data showing batch variation is shorter than that estimated from the data showing no batch variation at assay errors of not larger than 0.5%, whereas no significant difference is observed at larger assay error levels.

At a 25% difference in slope, the shelf-life estimate obtained by the batch-combination method is approximately 4 months longer than that obtained by the batch-separation method at an assay error of 0.5%. Similar degrees of difference were observed in the mean of estimates between the two methods in the assay error range up to 1.5%, though the standard deviation increases with increasing assay error. This 4-month difference seems to be important enough for shelf-lives of pharmaceuticals. The batch-combination method may provide an estimate of 4 months longer than real shelf-life. If this difference is assumed to be a minimum detectable difference, the significance level should be approximately 0.25 when the beta error and assay error are 20% and 0.5%, respectively, as shown in Fig.1. The 0.25 significance level recommended in the ICH guideline proposes, at a probability of 80%, to estimate shelf-life by the batch-separation method for a pharmaeutical with 25% batch-variation.

Conclusion

The power of the ANOVA for assessing batch-variation of stability data was determined as functions of significance level and assay error. The results obtained with degradation data simulated for a pharmaceutical exhibiting zero-order degradation of 0.2%/month indicate that a significance level of approximately 0.25 is necessary in order to obtain a minimum detectable difference of 25% if the beta error is fixed at 20%. It is recommended to employ assay methods with small error and to repeat the assay in order to reduce the assay error down to approximately 0.5%.

References

1) ICH (International Conference on Harmonisation of Technical Requirements for the Registration of Pharmaceuticals for Human Use) Harmonised Tripartite Guideline for Sability Testing of New Drug Substances and Products (October 1993).
2) Woolfe A. J., Worthington H.E.C., *Drug Dev. Communication*, **1**, 185—210 (1974—1975).
3) Winer B. J., "Statistical Principle in Experimental Design," McGraw-Hill Kogakusha, Ltd., Tokyo, 1971, p. 431.