Bootstrap Re-sampling Technique to Evaluate the Optimal Formulation of Theophylline Tablets Predicted by Non-linear Response Surface Method Incorporating Multivariate Spline Interpolation

Hiroaki ARAI,*[,a] Tatsuya SUZUKI,[a] Chosei KASEDA,[b] Koichi OHYAMA,[c] and Kozo TAKAYAMA[c]

[a] *Pharmaceutical Technology Research Laboratories, Daiichi Pharmaceutical Co., Ltd.; 1–16–13 Kita-kasai, Edogawa-ku, Tokyo 134–8630, Japan:* [b] *Research and Development Headquarters, Yamatake Corporation; 1–12–2 Kawana, Fujisawa, Kanagawa 251–8522, Japan: and* [c] *Department of Pharmaceutics, Hoshi University; 2–4–41 Ebara, Shinagawa-ku, Tokyo 142–8501, Japan.* Received December 6, 2006; accepted January 26, 2007; published online January 29, 2007

Optimal solutions of theophylline tablet formulations were derived from three types of experimental datasets, composed of different numbers of data-points using the response surface method incorporating multivariate spline interpolation (RSM^S). The reliability of these optimal solutions was evaluated by a bootstrap re-sampling technique. Different levels of three causal factors were used as factors of response surface analysis: the lactose/cornstarch ratio (X_1), the amount of carmellose calcium (X_2), and the amount of hydroxypropylcellulose (X_3). The target responses were the dissolution ratio of theophylline for the first 15 min (Y_1) and the hardness (Y_2) of each of the prepared tablets. Similar optimal solutions were estimated in three different sizes of datasets. A bootstrap re-sampling with replacements from the original dataset was applied, and optimal solutions for each bootstrap dataset were estimated. The frequency of the distribution of the optimal solution generated by the bootstrap re-sampling technique demonstrated almost normal distribution. The average and standard deviation of the optimal solution distribution were calculated as evaluation indices reflecting the accuracy and reproducibility of the optimal solution. It was confirmed that the accuracy was sufficiently high, irrespective of the dataset size; however, the reproducibility worsened with a decrease in the number of the experimental datasets. Consequently, it was considered that the novel evaluation method based on the bootstrap re-sampling technique was suitable for evaluating the reliability of the optimal solution.

Key words response surface method; optimization; bootstrap; cross-validation; multivariate spline interpolation; theophylline

In recent years, the ICH Q8 guidelines[1] have necessitated the establishment of a science-based rationale and a design space in pharmaceutical formulation development. The characteristics of a drug product are influenced by a number of parameters related to pharmaceutical formulation and manufacturing conditions. In the past, the process of formulation development was based on the approach of trial and error, in order to simultaneously satisfy the specifications for multiple properties. Therefore, optimization was a wasteful process for pharmaceutical formulations. However, in recent times, scientific approaches such as the response surface method (RSM) and the artificial intelligence (AI) technique have been used for resolving optimization problems.[2—7] RSM is useful for seeking acceptable solutions, however predictions based on the quadratic polynomial method are sometimes limited and the results obtained occasionally exhibit poor estimation.[8,9] To overcome these difficulties posed by the quadratic polynomial method, Takayama *et al.* successfully generated a smooth surface using an optimization technique that incorporated the multivariate spline interpolation (MSI) approach.[10]

Concurrently, it is important to evaluate the reliability of the optimal solutions estimated by RSM. The leave-one-out cross-validation (LOOCV) method is effective for versatile evaluation of the response surface.[11] The LOOCV method can estimate the generalization error of a given model.[12] However, the reliability of optimal solutions estimated by certain response surfaces cannot be directly evaluated using the LOOCV method. Therefore, we applied a novel method, namely, a bootstrap (BS) re-sampling technique,[12—14] to evaluate the reliability of the optimal solutions predicted by RSM incorporating MSI (RSM^S). In this study, we used an experimental dataset of theophylline tablets prepared by the fluidized-bed granulation method.

Theoretical

MSI Architecture MSI has been effectively used as a tool to interpolate altimeter data in the field of geophysics.[10,15] The basic concept of MSI can be considered as the transformation problems of elastics (Fig. 1).[16] In this problem, the experimental data were compared to the variations at each point, and the thin-plate spline minimized the elastic strain energy (Eq. 1).

$$E = \int_s (\nabla^2 f(x))^2 \, ds \tag{1}$$

where ∇^2 is the Lapracean and $\int_s ds$ is the integration over all ranges.

Green functions were used for solving Eq. 1, and the biharmonic spline was expressed as below.

$$y = \sum_{i=1}^{n} \alpha_i g(d_i) \tag{2}$$

where n is the number of data points, d_i is the standardized Euclidian distance between x corresponding to number i data point and the optimal x, $g(d_i)$ is the green function whose variables are d_i. The definition of this function is changed by the dimensions of the input variable (x). α_i is a coefficient calculated by a linear matrix operation.

Multi-Objective Optimization In the optimization problem for practi-

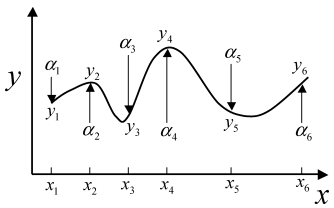


Fig. 1. A Function y That Passes through the Data Points y_i Located at x_i Is Found by Applying Point Forces a_j to a Thin Elastic Beam

* To whom correspondence should be addressed. e-mail: araihg51@daiichipharm.co.jp

cal pharmaceutical formulations, several response variables should be incorporated into a single equation in order to consider all the responses simultaneously. For this purpose, a general transformation that was based on the distance between the predicted value of each response and the individually obtained optimum one was employed by Takayama *et al.*[8—10)]

$$S(X)=\left\{\sum\left(\frac{FD_k(X)-FO_k(X)}{S.D._k}\right)^2\right\}^{1/2} \quad (3)$$

where $S(X)$ is the distance function generalized by the standard deviation $S.D._k$ of the observed values for each response variable, $FD_k(X)$ is the optimum value of each response variable optimized individually over the experimental region and $FO_k(X)$ is the estimated value of all the responses given in the same set of causal factors, *i.e.*, X. To solve the multi objective optimization problem under inequality and/or equality constraints, Eq. 3 is transformed to the unconstrained problem by adding a penalty function as follows.

$$T(X,r)=\left\{\sum\left(\frac{FD_k(X)-FO_k(X)}{S.D._k}\right)^2\right\}^{1/2} + r^{-1}\sum\phi_i\{G_i(X)\}^2+r^{-1}\sum\{H_j(X)\}^2 \quad (4)$$

when $G_i(X)<0$, $\phi_1=1$; when $G_i(X)\geqq 0$, $\phi_1=0$.

In Eq. 4, $T(X,r)$ is the transformed unconstrained objective function, $G_i(X)$ is the inequality constraint, $H_j(X)$ is the equality constraint, r is a perturbation parameter ($r>0$) of $T(X,r)$, and ϕ_i is a step function by which the objective function $S(X)$ is penalized because the value increases abruptly when the values of $G_i(X)$ are negative or when the $H_j(X)$ values deviate from zero. An optimal solution is estimated as the point $X(r)$; this gives a minimum of $T(X,r)$ when the value of r is sufficiently close to zero.

BS for Parameter Estimation The BS method that was introduced by Efron[17)] is a simulation technique based on the empirical distribution of the observed sample.[12—15,17)] Let $x=(x_1,\ldots,x_n)$, an n-sample with an unknown distribution function F, depending on an unknown real parameter θ. The problem involves evaluating this parameter θ by a statistic $\hat{\theta}=s(x)$ from the sample x and in evaluating the estimate accuracy, although the distribution F is unknown. In order to evaluate this estimate accuracy, B samples are built from the initial sample x by re-sampling. These samples are called BS samples and are denoted by x^{*b}.

A BS sample $x^{*b}=(x_1^{*b},\ldots,x_n^{*b})$ is built by a random re-sampling with replacement from the initial sample x. The distribution function of a BS sample x^{*b} is $\hat{F}$, *i.e.*, the empirical distribution of x. A BS replicate of estimator $\hat{\theta}=s(x)$ will be $\hat{\theta}^{*b}=s(x^{*b})$. For example, for the mean of the sample x, the estimator is $s(x)=(1/n)\sum_{i=1}^{n}x_i$, and a BS replicate will be $s(x^{*b})=(1/n)\sum_{i=1}^{n}x_i^{*b}$.

Then, the BS estimate of the standard deviation of $\hat{\theta}$ denoted by $\hat{\sigma}_{boot}(\hat{\theta})$ is given by Eqs. 5 and 6 as follows.

$$\hat{\sigma}_{boot}(\hat{\theta}^*)=\left[\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}^{*b}-\hat{\theta}^*(\cdot))^2\right]^{1/2} \quad (5)$$

$$\hat{\theta}^*(\cdot)=\frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{*b} \quad (6)$$

Experimental

Materials Theophylline (Shiratori pharmaceutical Co., Ltd., Japan), lactose (200-mesh grade, DMV international, Netherlands), cornstarch (Nihon shokuhin kako Co., Ltd., Japan), carmellose calcium (Gotoku Chemical Co., Ltd., Japan), hydroxypropylcellulose (Nippon soda Co., Ltd., Japan) and magnesium stearate (Nitto kasei kogyo k.k., Japan) were all of grades conforming to the current *Japanese Pharmacopoeia* (JP).

Preparation Method of Sample Tablets The formulations of theophylline tablets are listed in Table 1. The lactose/cornstarch ratio (X_1), the amount of carmellose calcium (X_2), and the amount of hydroxypropylcellulose (X_3) were selected as causal factors. The orthogonal array design for three factors and three levels was applied to prepare the test formulations. Theophylline was milled using an impact mill (TASM-1CS, Tokyo atomizer Co., Ltd., Japan), and lactose was sieved through a 60-mesh screen. The milled theophylline (volume mean diameter about 23.6 μm), sieved lactose, cornstarch, and carmellose calcium were blended in a polyethylene bag for

Table 1. Formulation of Theophylline Tablets for Orthogonal Array Experimental Design

Experiment number	Factor level			Formula (mg/tablet)						
	X_1	X_2	X_3	THEO[a)]	LAC[b)]	CS[c)]	CCa[d)]	HPC[e)]	Mg-St[f)]	Total
1	−1	−1	−1	100.0	26.4	61.6	6.0	4.0	2.0	200.0
2	−1	−1	0	100.0	25.8	60.2	6.0	6.0	2.0	200.0
3	−1	−1	1	100.0	25.2	58.8	6.0	8.0	2.0	200.0
4	−1	0	−1	100.0	25.2	58.8	10.0	4.0	2.0	200.0
5	−1	0	0	100.0	24.6	57.4	10.0	6.0	2.0	200.0
6	−1	0	1	100.0	24.0	56.0	10.0	8.0	2.0	200.0
7	−1	1	−1	100.0	24.0	56.0	14.0	4.0	2.0	200.0
8	−1	1	0	100.0	23.4	54.6	14.0	6.0	2.0	200.0
9	−1	1	1	100.0	22.8	53.2	14.0	8.0	2.0	200.0
10	0	−1	−1	100.0	44.0	44.0	6.0	4.0	2.0	200.0
11	0	−1	0	100.0	43.0	43.0	6.0	6.0	2.0	200.0
12	0	−1	1	100.0	42.0	42.0	6.0	8.0	2.0	200.0
13	0	0	−1	100.0	42.0	42.0	10.0	4.0	2.0	200.0
14	0	0	0	100.0	41.0	41.0	10.0	6.0	2.0	200.0
15	0	0	1	100.0	40.0	40.0	10.0	8.0	2.0	200.0
16	0	1	−1	100.0	40.0	40.0	14.0	4.0	2.0	200.0
17	0	1	0	100.0	39.0	39.0	14.0	6.0	2.0	200.0
18	0	1	1	100.0	38.0	38.0	14.0	8.0	2.0	200.0
19	1	−1	−1	100.0	61.6	26.4	6.0	4.0	2.0	200.0
20	1	−1	0	100.0	60.2	25.8	6.0	6.0	2.0	200.0
21	1	−1	1	100.0	58.8	25.2	6.0	8.0	2.0	200.0
22	1	0	−1	100.0	58.8	25.2	10.0	4.0	2.0	200.0
23	1	0	0	100.0	57.4	24.6	10.0	6.0	2.0	200.0
24	1	0	1	100.0	56.0	24.0	10.0	8.0	2.0	200.0
25	1	1	−1	100.0	56.0	24.0	14.0	4.0	2.0	200.0
26	1	1	0	100.0	54.6	23.4	14.0	6.0	2.0	200.0
27	1	1	1	100.0	53.2	22.8	14.0	8.0	2.0	200.0

a) Theophylline. *b*) Lactose. *c*) Cornstarch. *d*) Carmellose calcium (ECG-505). *e*) Hydroxypropylcellulose (HPC-L). *f*) Magnesium stearate.

2 min. The mixture was granulated with an approximately 6% (w/v) aqueous hydroxypropylcellulose solution in a fluid-bed granulator (MFL.01, Vector Corporation, U.S.A.) under constant operational conditions as shown in Table 2. The granules were lubricated with magnesium stearate and were blended in a bin blender (Turbula unit type T2C, Willy A. Bachofen AG, Switzerland) for 5 min. The final blend was compressed into tablets by a universal testing machine (Autograph AG-5000B, Shimadzu Co., Ltd., Japan) at a compression force of approximately 7.85 kN.

Determination of the Response Variables The dissolution ratio of theophylline for the first 15 min (Y_1) and hardness (Y_2) were selected as the response variables that were to be valuated in the resulting tablets. Both variables were represented as the mean of three determinations.

a) Dissolution: Dissolution testing was performed using the paddle method according to the recommendations of the current JP; it was performed at 50 rpm in 900 ml of water at 37 °C. The dissolved theophylline was assayed by an automated flow-through UV spectrophotometric method at 243 nm with a 10-mm-long cell (Automated dissolution apparatus, Toyama Sangyo Co., and Shimadzu Co., Ltd., Japan).

b) Hardness: The hardness of the resulting tablets was measured using a hardness tester (Tablet tester type 6D, Dr. Schleuniger Pharmatron AG, U.S.A.).

Preparation of the Original Dataset Three types of experimental datasets were prepared according to an orthogonal array experimental design that consisted of three causal factors and two response variables with 27 theophylline tablet formulations.

In this paper, the smallest, medium, and the largest sizes of the datasets are named L_9 Original dataset, L_{18} Original dataset, and L_{27} Original dataset, respectively. These datasets are shown in Table 3.

Evaluation Method for an Optimal Solution Based on the BS Technique The concept of the BS evaluation method is shown in Fig. 2. The method of evaluation for an optimal solution based on the BS technique has been described as follows.[18,19]

Step 1. The BS dataset corresponding to the respective original datasets (L_9, L_{18}, and L_{27}) is generated by BS re-sampling repeated n (9, 18, or 27) times to form an ensemble comprising n results.

Step 2. Step 1 is repeated until B times, and B units of the BS dataset are generated.

Step 3. The optimal solution is calculated as X^*optim ($X^{*1}optim$, $X^{*2}optim$,…, $X^{*B}optim$) and Y^*optim ($Y^{*1}optim$, $Y^{*2}optim$,…, $Y^{*B}optim$) for each BS dataset, and the distribution of the optimal solution is generated.

Step 4. The optimal solution and standard deviation of the BS analysis are calculated according to Eqs. 5 and 6.

The accuracy of the original optimal solution, which was calculated from the original dataset, can be evaluated by comparing it to be BS optimal solutions. If the BS optimal solution deviates from the optimal original solution, it is considered that the optimal original solution has low reliability with regard to accuracy. In addition, the precision of the optimal original solution can also evaluated by the BS standard deviation. A large BS standard deviation indicates poor precision of the optimal original solution.

In this study, the frequency numbers B of BS re-sampling were set at 50, 100, 200, 300, 400, and 500.

The Evaluation Index of the Response Surface The optimum solutions and response surfaces for each size of the original datasets (L_9, L_{18}, and L_{27}) were calculated by RSM^S. LOOCV was performed for each response surface. R^2 values that are often used as a conventional evaluation index of the response surface were calculated using Eq. 7.[11]

$$R^2 = 100 \times \left(1 - \frac{SSE}{SST}\right) \tag{7}$$

where SSE is the sum of the squared error between the predicted and the

Table 2. Operational Conditions for Preparing Granules in the Fluid-Bed Granulator

Parameters	Set values
Fluid nozzle/needle size	Fine
Air cap	Medium
Spray pressure (kPa)	34.5
Spray rate of solution (g/min)	1
Air flow (l/min)	100
Inlet air temperature (°C)	60
Target exhaust air temperature (°C)	35

Table 3. Original Dataset of Theophylline Tablet Formulations

		Experiment number	Factor level			% Dissolved theophylline at 15 min, Y_1	Hardness (N), Y_2
			X_1[a] (%)	X_2[b] (%)	X_3[c] (%)		
		1	30	3.0	2.0	88.5	87.3
		5	30	5.0	3.0	76.9	93.2
		9	30	7.0	4.0	51.9	106.9
		12	50	3.0	4.0	47.0	118.7
		13	50	5.0	2.0	97.4	89.2
		17	50	7.0	3.0	75.1	100.0
		20	70	3.0	3.0	61.4	112.8
	L_9	24	70	5.0	4.0	49.1	144.2
		25	70	7.0	2.0	94.2	117.7
		2	30	3.0	3.0	59.4	80.4
		6	30	5.0	4.0	47.6	110.8
		7	30	7.0	2.0	90.6	86.3
		10	50	3.0	2.0	94.3	93.2
		14	50	5.0	3.0	69.7	114.7
		18	50	7.0	4.0	61.9	117.7
		21	70	3.0	4.0	41.0	142.2
	L_{18}	22	70	5.0	2.0	92.8	112.8
		26	70	7.0	3.0	83.7	129.5
		3	30	3.0	4.0	41.0	98.1
		4	30	5.0	2.0	89.1	108.9
		8	30	7.0	3.0	74.9	93.2
		11	50	3.0	3.0	75.6	103.0
		15	50	5.0	4.0	58.0	74.5
		16	50	7.0	2.0	92.9	93.2
		19	70	3.0	2.0	94.1	104.0
L_{27}		23	70	5.0	3.0	75.4	122.6
		27	70	7.0	4.0	55.0	137.3

a) Lactose/cornstarch ratio (%lactose). *b*) % Carmellose calcium. *c*) % Hhydroxypropylcellulose.

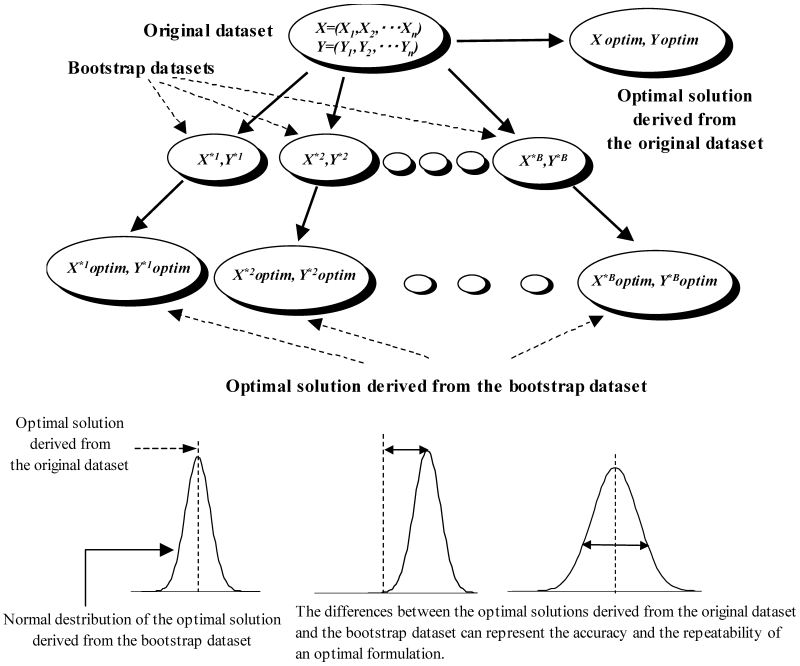
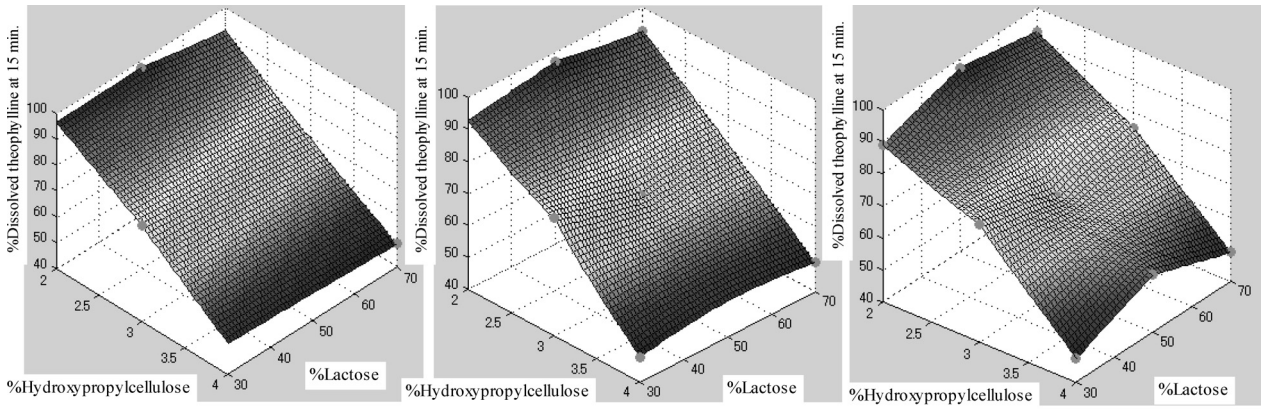


Fig. 2. The Evaluation Process for Estimating the Accuracy and Precision (Repeatability) of an Optimal Formulation Based on the Bootstrap Re-sampling Technique



(a) L_9 design (b) L_{18} design (c) L_{27} design

Fig. 3. The Response Surface of the Dissolution Ratio for the First 15 min (Y_1) as a Function of the Lactose/Cornstarch Ratio (% Lactose, X_1) and the Hydroxypropylcellulose Ratio (X_3) at a Constant Carmellose Calcium Ratio (5.0%) Using the Three Different Sizes of Datasets

measured values. *SST* is the sum of the squared error between each measured value and the average of the measured value.

Evaluation Indices of Accuracy and Reproducibility for Optimal Solutions The δ and CV_B values that are used as evaluation indices of accuracy and repeatability of the optimal solution were calculated using Eqs. 8 and 9.

$$\delta = \frac{|F - F_{B.m}|}{F} \times 100 \tag{8}$$

$$CV_B = \frac{SD_B}{F_{B.m}} \times 100 \tag{9}$$

where F is the original solution of a specified property, $F_{B.m}$ is the BS optimal solution of the same property corresponding to the original solution, and SD_B is the BS standard deviation.

Software In this study, we used the software dataNESIA developed by Yamatake Corporation (Japan) for generating the RSM^S and the BS re-sampling; this software consists of a multi-dimensional interpolating program and a nonlinear optimization program.[20]

Results and Discussion

Prediction of Response Variables and Simultaneous Optimization The dissolution ratio of theophylline for the first 15 min (Y_1) and the hardness (Y_2) of each formulation are shown in Table 3. The dissolution profiles and hardness varied among the formulations. The response surfaces of Y_1 and Y_2 were generated by RSM^S as functions of three causal factors; the lactose/cornstarch ratio (X_1), the amount of carmellose calcium (X_2), and the amount of hydroxypropylcellulose (X_3). Representative examples of the response surfaces of the dissolution ratio using the different sizes of the datasets are shown in Fig. 3. Since these response surfaces were similar to each other, it was suggested that a reasonable response surface was estimated by using RSM^S in spite of the size of the dataset. A conventional LOOCV method was applied to evaluate the reliability of each response surface. The R^2 values defined in Eq. 7 for Y_1 and Y_2 were calculated for

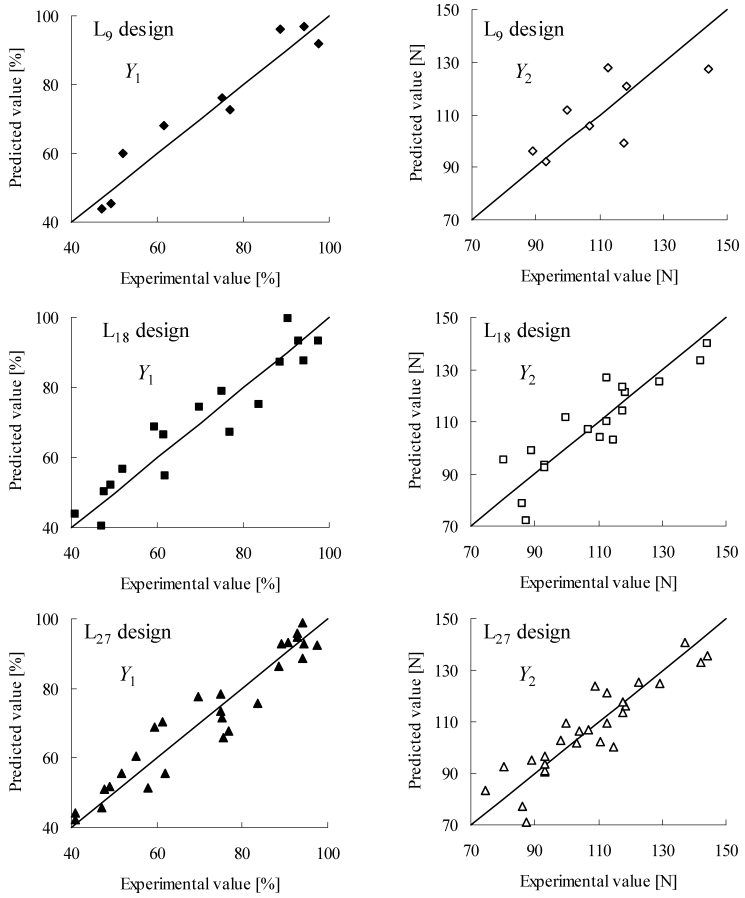


Fig. 4. The Relationships between Experimental and Predicted Values of Y_1 and Y_2

Y_1: % dissolved theophylline at 15 min. Y_2: hardness.

the L_9, L_{18}, and L_{27} datasets. These results are shown in Fig. 4 and Table 5. With the exception Y_2 in L_9 design, all R^2 values exhibited sufficiently high values more than 80; this suggested that all the response surfaces, exception Y_2 in L_9 design, were reasonably stable. Thus, it was confirmed that the obtained response surfaces were hardly affected by the individual data points. The considerably low R^2 values of Y_2 that were predicted by the L_9 dataset suggested that the data numbers of the L_9 dataset were insufficient to generate a stable response surface.

Simultaneous optimal solutions for X_1, X_2, and X_3 and the predicted values for Y_1 and Y_2 calculated for the three sizes of the datasets are shown in Table 4. Calculations were performed according to the standardized Euclidian distance function as defined in Eq. 4 under the restrictions of the experimental region. Interestingly, the optimal solutions exhibited similar results regardless of the size of the dataset; therefore, we can say that RSM^S is applicable even in the case of a small dataset size. However, the reliability of optimal solutions cannot be evaluated quantitatively by means of the R^2 values alone because the values only indicate the stability of the response surface. Therefore, a novel method is required for evaluating the reliability of the optimal solution.

Evaluation of the Optimal Formulation by the BS Method BS datasets for three different sizes of the original datasets were generated by BS re-sampling that was set at a frequency of 50, 100, 200, 300, and 500. The results of the optimal formulations and predicted responses are shown in Tables 6—8. The BS optimal solutions as well as the standard deviations were almost constant despite altering the frequency of re-sampling, suggesting that a re-sampling frequency of more than 50 was sufficient to estimate the stability of the optimal formulations.

Table 4. Optimized Formulations and Predicted Responses Estimated by Three Different Sizes of Datasets

Factors and responses	Experimental design		
	L_9 design	L_{18} design	L_{27} design
X_1: Lactose/cornstarch ratio (% lactose)	66.54	66.84	66.80
X_2: % Carmellose calcium	5.79	5.95	5.84
X_3: % Hydroxypropylcellulose	2.60	2.74	2.66
Y_1: % Dissolved theophylline at 15 min	80.30	80.99	83.5
Y_2: Hardness (N)	106.0	112.1	109.9

Table 5. R^2 Values of % Dissolved Theophylline at 15 min (Y_1) and Hardness (Y_2) as Calculated by Using the Results Obtained from the Leave-One-Out Cross Validation Method

Experimental design	R^2 value		
	L_9 design	L_{18} design	L_{27} design
% Dissolved theophylline at 15 min, (Y_1)	91.7	99.1	91.4
Hardness, (Y_2)	47.6	99.2	81.8

Table 6. Bootstrap Optimal Solutions and Bootstrap Standard Deviations Generated by Different Frequencies of Bootstrap Re-sampling in the L_9 Orthogonal Array Experimental Design

Re-sampling frequency	Optimized formulations			Predicted responses	
	X_1 [b] (%)	X_2 [c] (%)	X_3 [d] (%)	Y_1 [e] (%)	Y_2 [f] (N)
$N=0$ [a]	66.54	5.79	2.60	80.30	114.3
$N=50$	63.83 (5.941)	5.66 (0.634)	2.64 (0.165)	79.85 (3.757)	114.3 (6.237)
$N=100$	64.52 (6.061)	5.53 (0.589)	2.60 (0.188)	80.33 (4.211)	114.7 (5.227)
$N=200$	64.59 (5.890)	5.60 (0.554)	2.64 (0.200)	79.67 (3.716)	114.5 (4.894)
$N=300$	64.31 (6.091)	5.63 (0.583)	2.63 (0.193)	80.03 (3.894)	114.6 (5.296)
$N=400$	64.72 (5.606)	5.61 (0.588)	2.63 (0.192)	79.78 (4.060)	114.7 (5.070)
$N=500$	64.52 (5.846)	5.61 (0.603)	2.62 (0.196)	80.04 (4.076)	114.4 (5.325)

( ): bootstrap standard deviation. *a*) Obtained from the original dataset. *b*) Lactose/cornstarch ratio (% lactose). *c*) % Carmellose calcium. *d*) % Hydroxypropylcellulose. *e*) % Dissolved theophylline at 15 min. *f*) Hardness.

Table 7. Bootstrap Optimal Solutions and Bootstrap Standard Deviations Generated by Different Frequencies of Bootstrap Re-sampling in the L_{18} Orthogonal Array Experimental Design

Re-sampling frequency	Optimized formulations			Predicted responses	
	X_1 [b] (%)	X_2 [c] (%)	X_3 [d] (%)	Y_1 [e] (%)	Y_2 [f] (N)
$N=0$ [a]	66.84	5.95	2.74	80.99	121.0
$N=50$	66.82 (1.420)	5.74 (0.327)	2.69 (0.166)	81.30 (3.497)	119.8 (2.560)
$N=100$	66.72 (0.976)	5.85 (0.268)	2.71 (0.125)	81.31 (2.550)	119.8 (5.569)
$N=200$	66.66 (1.086)	5.85 (0.305)	2.70 (0.103)	81.75 (1.593)	120.0 (2.618)
$N=300$	66.66 (1.126)	5.83 (0.297)	2.69 (0.091)	81.72 (1.466)	119.5 (2.540)
$N=400$	66.60 (1.542)	5.84 (0.299)	2.70 (0.117)	81.55 (2.694)	119.7 (2.756)
$N=500$	66.78 (1.032)	5.82 (0.293)	2.69 (0.089)	81.84 (1.399)	119.7 (2.560)

( ): bootstrap standard deviation. *a*) Obtained from the original dataset. *b*) Lactose/cornstarch ratio (% lactose). *c*) % Carmellose calcium. *d*) % Hydroxypropylcellulose. *e*) % Dissolved theophylline at 15 min. *f*) Hardness.

Table 8. Bootstrap Optimal Solutions and Bootstrap Standard Deviations Generated by Different Frequencies of Bootstrap Re-sampling in the L_{27} Orthogonal Array Experimental Design

Re-sampling frequency	Optimized formulations			Predicted responses	
	X_1 [b] (%)	X_2 [c] (%)	X_3 [d] (%)	Y_1 [e] (%)	Y_2 [f] (N)
$N=0$ [a]	66.80	5.84	2.66	83.50	118.6
$N=50$	66.90 (0.860)	5.73 (0.276)	2.65 (0.084)	83.11 (1.150)	118.1 (2.628)
$N=100$	67.95 (0.818)	5.71 (0.284)	2.66 (0.136)	82.82 (2.686)	118.2 (2.638)
$N=200$	66.88 (0.844)	5.74 (0.266)	2.65 (0.075)	83.23 (1.409)	118.1 (2.010)
$N=300$	66.76 (0.837)	5.78 (0.266)	2.65 (0.085)	83.33 (1.232)	118.2 (2.128)
$N=400$	66.88 (0.859)	5.73 (0.274)	2.64 (0.082)	83.29 (1.302)	117.9 (2.118)
$N=500$	66.89 (0.811)	5.75 (0.264)	2.66 (0.111)	83.08 (2.033)	118.2 (2.354)

( ): bootstrap standard deviation. *a*) Obtained from the original dataset. *b*) Lactose/cornstarch ratio (% lactose). *c*) % Carmellose calcium. *d*) % Hydroxypropylcellulose. *e*) % Dissolved theophylline at 15 min. *f*) Hardness.

Table 9. 95% Confidence Intervals of Optimal Solutions Generated by Three Different Sizes of Datasets When Bootstrapping Was Repeated Approximately 300 Times

Re-sampling frequency	Optimized formulations			Predicted responses	
	X_1 [c] (%)	X_2 [d] (%)	X_3 [e] (%)	Y_1 [f] (%)	Y_2 [g] (N)
L_9 dataset					
Original solution [a]	66.54	5.79	2.60	80.30	114.3
Bootstrap solution [b]	64.31 (6.091)	5.63 (0.583)	2.63 (0.193)	80.03 (3.894)	114.6 (5.296)
95% confidence interval	41.11—69.05	4.30—6.68	2.35—3.25	72.52—85.50	103.9—123.9
L_{18} dataset					
Original solution [a]	66.84	5.95	2.74	80.99	121.0
Bootstrap solution [b]	66.66 (1.126)	5.83 (0.297)	2.69 (0.091)	81.72 (1.466)	119.5 (2.540)
95% confidence interval	64.43—68.63	5.17—6.28	2.50—2.82	78.61—84.58	113.6—123.4
L_{27} dataset					
Original solution [a]	66.80	5.84	2.66	83.50	118.6
Bootstrap solution [b]	66.76 (0.837)	5.78 (0.266)	2.65 (0.085)	83.33 (1.232)	118.2 (2.128)
95% confidence interval	64.96—68.42	5.22—6.15	2.47—2.79	80.89—85.73	113.9—122.0

( ): bootstrap standard deviation. *a*) Obtained from the original dataset. *b*) Bootstrap re-sampling frequency, approximately 300. *c*) Lactose/cornstarch ratio (% lactose). *d*) % Carmellose calcium. *e*) % Hydroxypropylcellulose. *f*) % Dissolved theophylline at 15 min. *g*) Hardness.

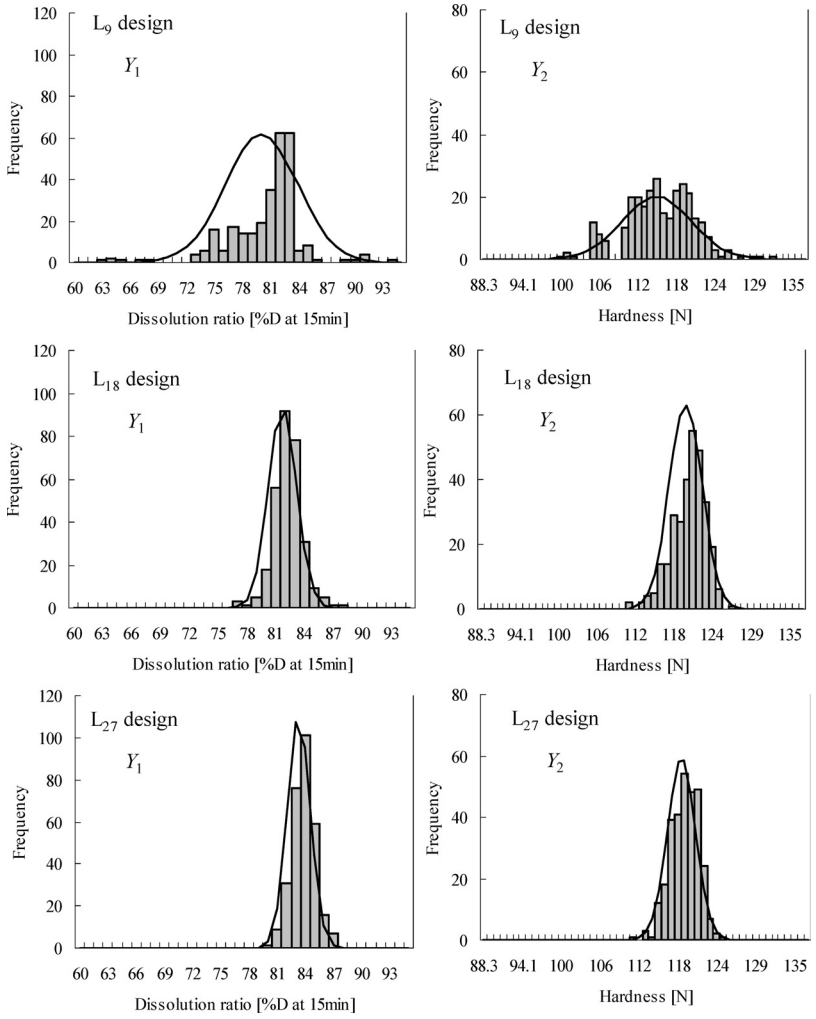


Fig. 5. Distributions of the Optimal Solutions Generated by Bootstrap Re-sampling (N=300) of the Three Different Sizes of Datasets
Y_1: % Dissolved theophylline at 15 min. Y_2: hardness.

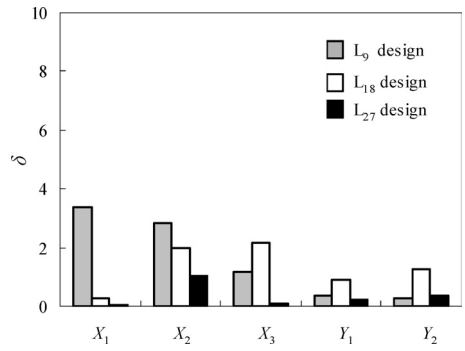


Fig. 6. Comparison of the δ Index for Accuracy of the Optimal Solution between the Three Different Sizes of Datasets

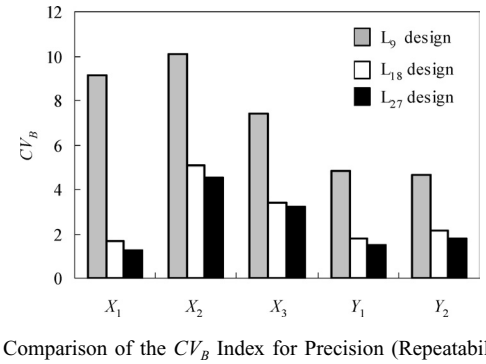


Fig. 7. Comparison of the CV_B Index for Precision (Repeatability) of the Optimal Solution between the Three Different Sizes of Datasets

The distributions of the optimal solutions generated by a re-sampling frequency of 300 are shown in Fig. 5. These distributions were almost symmetrical, and they were regarded as the normal distribution. The peaks of the individual distributions exhibit the BS optimal solution. These results supported the hypothesis that the BS method that is based on the central limit theorem can be applied to evaluate optimal solutions.

The 95% confidence intervals of optimal solutions were calculated for three different sizes of datasets by the percentile method.[21] The results are shown in Table 9. The ranges of the 95% confidence intervals calculated for the L_{18} and L_{27} datasets were sufficiently narrow for the practical studies of formulations. On the other hand, the 95% confidence interval calculated for the L_9 dataset was wider than that calculated for the L_{18} and L_{27} datasets.

Indices corresponding to the accuracy (δ) and reproducibility (CV_B) of the optimal solution as defined in Eqs. 8 and 9 are shown in Figs. 6 and 7, respectively. As the results

of evaluation for optimal solutions, δ values as an accuracy index were sufficiently small regardless of size of dataset. On the other hand, we observed a decrease in the CV_B values a reproducibility index of the optimal solution with an increase in the size of the dataset. In this formulation experiment, the precision of the optimal solution improved with an increase in the number of experimental data, although the accuracy of the optimal solution was ensured even with a small dataset size.

Conclusions

Based on the results of this study, it was considered that a novel evaluation method based on the BS re-sampling technique was suitable for evaluating the accuracy and precision of the optimal solution estimated by RSM^S.

Acknowledgements The authors are very grateful to Mr. Hiroaki Tsutui and Mr. Junji Nishimura, Yamatake Corporation, for their valuable advice in the data analysis. The authors are also grateful to Ms. Michiko Iketani, Daiichi Pharmaceutical Co., Ltd., for her assistance in the experimental work.

References

1) ICH draft guideline, "Pharmaceutical Development Q8," version 4.3, November 2005.
2) Rekhi G. S., Nellore R. V., Hussain A. S., Tillman L. G., Malinowski H. J., Augsburger L. L., *J. Control. Rel.*, **59**, 327—342 (1999).
3) Paterakis P. G., Korakianiti E. S., Dallas P. P., Rekkas D. M., *Int. J. Pharm.*, **248**, 51—60 (2002).
4) Huang Y. B., Tasai Y. H., Yang W. C., Chang J. S., Wu P. C., *Biol. Pharm. Bull.*, **27**, 1626—1629 (2004).
5) AIKhatib H. S., Sakr A., *Pharm. Dev. Technol.*, **8**, 87—96 (2003).
6) Marengo E., Cavalli R., Rovero G., Gasco M. R., *Pharm. Dev. Technol.*, **8**, 299—309 (2003).
7) Matsumura M., Nakagami H., Yamao T., Takayama K., Nagai T., *Chem. Pharm. Bull.*, **42**, 1902—1908 (1994).
8) Takayama K., Morva A., Fujikawa M., Hattori Y., Obata Y., Nagai T., *J. Control. Rel.*, **68**, 175—186 (2000).
9) Wu P. C., Obata Y., Fujikawa M., Li C. J., Higashiyama K., Takayama K., *J. Pharm. Sci.*, **90**, 1004—1014 (2001).
10) Takayama K., Obata Y., Morishita M., Nagai T., *Pharmazie*, **59**, 392—395 (2004).
11) Bourquin J., Schmidli H., Hoogevest P. V., Leuenberger H., *Eur. J. Pharm. Sci.*, **6**, 287—300 (1998).
12) Ueda N., Nakano R., *IEEE Int. Conf. Neural Netw. Proc.*, **1**, 101—104 (1995).
13) Dupret G., Koda M., *Eur. J. Oper. Res.*, **134**, 141—156 (2001).
14) Zhang J., *Neural Netw.*, **12**, 927—938 (1999).
15) Sandwell D. S., *Geophys. Res. Let.*, **14**, 139—142 (1987).
16) Takayama K., *Pharm. Tech. Jpn.*, **18**, 1687—1694 (2002).
17) Efron B., Tibshirani R. J., "An Introduction to the Bootstrap," Chapman and Hall, New York, 1993.
18) Oyama K., Takayama K., *Yakuzaigaku*, **65** (Suppl.), 90 (2005).
19) Arai H., Suzuki T., Kaseda C., Takayama K., *Yakuzaigaku*, **65** (Suppl.), 280 (2005).
20) Kaseda C., *Kagakukougaku*, **68**, 315—317 (2004).
21) Higgins J. J., "An Introduction to Modern Nonparametric Statistics," Duxbury Press, Boston, 2003.