DRUG DEVELOPMENT AND INDUSTRIAL PHARMACY, 5(2), 175-216 (1979)

SELECTING KEY PHARMACEUTICAL FORMULATION FACTORS BY REGRESSION ANALYSIS

N. R. Bohidar[1], F. A. Restaino[2] and J. B. Schwartz[2]

[1]Biometrics Research Department

Merck Sharp and Dohme Research Laboratories

West Point, Pennsylvania 19486

[2]Pharmaceutical Research and Development

Merck Sharp and Dohme Research Laboratories

West Point, Pennsylvania 19486

ABSTRACT

The use of Selective Regression Analysis to determine which formulation factors are governing product properties is demonstrated. The techniques of using and combining the two procedures called "All Possible Regression" (APR) and "Stepwise Regression" (SWR) are presented and applied to a multivariate pharmaceutical formulation problem.

The technique was successfully applied to a system consisting of 10 response variables (tablet properties). Analysis of the results showed that for this formulation compression pressure and lubricant level exert the greatest effect on the majority of the properties. The results obtained from this method of analysis can

aid the formulator in selectively controlling the product properties of choice while leaving the others undisturbed.

Selective Regression Analysis also provides a basis for understanding the underlying mechanism of the system under consideration.

INTRODUCTION

The development of a drug delivery system involves a large number of process variables. It is of interest to the research pharmacist to know which of these process variables influence directly or indirectly a selected property of a given drug delivery system. This knowledge would provide a basis for controlling the key formulation factors (process variables) to effect a favorable change in the response parameter of interest. For instance, five or more formulation factors were considered in the development of a pharmaceutical tablet formulation in (1). If one is interested in bringing about a desirable change in the dissolution profile of the system considered in (1), then, under the assumption of the existence of a close functional relationship between the five formulation factors and the dissolution rates, it would be perfectly reasonable to ask which of these five factors would play an important role in bringing about effective changes in the response parameter under consideration. One may entertain more specific questions such as, will a change in the lubrication level significantly improve the dissolution profile of the system, or, to what extent a change in the diluent ratio would maintain the same dissolution value as we accepted in the

initial stage of the development of the process? The answer to the latter question is of prime interest in the situation in which one is trying to achieve desirable changes in a given parameter without substantially affecting the values of the other parameters measured in the system.

The interrelation among the process variables and the interrelation between the dependent variable and each of the process variable in the system play a vital role in the selection process of the key formulation factors associated with the system. Due to the inseparable closeness of the variables in a correlated system, it would not be easy to identify these factors without resorting to an objective procedure of elaborate step-by-step numerical examination. There is a statistical method, called "Selective Regression Analysis" which accomplishes this very goal effectively. Out of the several available alternate procedures associated with this analysis, there are two most widely accepted procedures which can be used independently, and if the results of the two independent analyses are combined properly, the combined results would generally yield an unique set of key formulation factors for a given response parameter.

The primary purpose of this paper is to demonstrate the role of "Selective Regression Analysis" in the identification of the key formulation factors associated with a given system. The two procedures associated with the selective regression analysis are called "All Possible Regression" (APR) and "Stepwise Regression" (SWR). The individual results of these two procedures are ex-

plicitly presented, the technique of combining the results of these two procedures are illustrated and the appropriate interpretation of the combined results are elucidated. The outcome of the combination of the results of the two procedures has the property of essentially eliminating the deficiency associated with either one of the two procedures used individually.

In this paper, the two selective regression procedures (APR and SWR) are applied to the data associated with an optimization experiment previously described in (1). The effect of the five formulation factors considered were measured on each of the ten response parameters (tablet properties) based on 27 distinct formulations. The selective regression analysis is applied to each of the ten response variables and a set of key formulation factors is obtained for each of the ten tablet properties. The results of these ten separate analyses are presented and discussed.

The theoretical section is meant to serve as a guide for following the steps associated with APR procedure, SWR procedure and the combination technique. However, computer programs can perform all of the indicated operations for both APR and SWR procedures, and it is only necessary to feed the raw data into the computer system. The combination technique does not require any elaborate computational analysis. A familiarity with the theory is useful for the analysis of the results and for those who wish to modify available programs to suit their own specific needs and formats.

EXPERIMENTAL

The five formulation factors and the ten response parameters measured in each of the 27 tablet formulations considered in the experiment are presented in Table I and Table II, respectively. The experimental procedure and the details with regard to the formulation factors and the response parameters considered in the experiment have been sufficiently elaborated in Ref. 1. Thus, the data set to be subjected to selective regression analysis contains 27 values for each response variable shown in Table II.

THEORY AND PROCEDURE

The structure of the data for a given response parameter is presented in a matrix form in Table III.

TABLE I

FORMULATION FACTORS

Symbolic Designation	Formulation Factors	Units of Measurement
X_1 (CL)	Calcium Phosphate/ Lactose Ratio	milligram/milligram
X_2 (CP)	Compression Pressure	tons
X_3 (SD)	Starch Disintegrant	milligram
X_4 (GG)	Granulating Gelatin	milligram
X_5 (MS)	Magnesium Stearate	milligram

TABLE II
RESPONSE PARAMETERS

Symbolic Designation		Response Variables	Units
Y_1	(DT)	Disintegration Time	Minutes
Y_2	(HD)	Tablet Breaking Strength	Kilograms
Y_3	(DR)	Dissolution (t_{30})	Percent Released in 30 min.
Y_4	(FR)	Friability	Percent Weight
Y_5	(TH)	Thickness Uniformity	RSD, %
Y_6	(PO)	Pore Volume	Milliliters per gram
Y_7	(MP)	Mean Pore Diameter	Micrometers
Y_8	(WT)	Weight Uniformity	RSD,%
Y_9	(TB)	Tablet Breakage	Number of Chipped Tablets
Y_{10}	(GM)	Granulation Mean Diameter	Millimeters

TABLE III
DATA MATRIX

Experiment Number (Formulations)	(CL)	(CP)	(SD)	(GG)	(MS)	A Response Variable
1	X_{11}	X_{21}	X_{31}	X_{41}	X_{51}	Y_1
2	X_{12}	X_{22}	X_{32}	X_{42}	X_{52}	Y_2
3	X_{13}	X_{23}	X_{33}	X_{43}	X_{53}	Y_3
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
27	$X_{1,27}$	$X_{2,27}$	$X_{3,27}$	$X_{4,27}$	$X_{5,27}$	Y_{27}

To accomplish the selective regression analysis for each of the ten parameters considered in the system, ten such data matrices (see Table III), one for each parameter, have to be constructed and fed into the computer either simultaneously or independently depending upon the file manipulation capability available in the software system.

A description of the two procedures mentioned above is given in a step-by-step manner in the following:

APR Procedure

For the purpose of describing the procedure, let X_{ij} ($i = 1, 2, \ldots K$, $j = 1, 2, \ldots N$) denote the numerical value of the i^{th} process variable (independent variable) for the j^{th} experiment (here, tablet formulation) and let K and N represent the number of process variables and the number of experiments considered in the study, respectively. In this case, K = 5 and N = 27 (Table III). Now let Y_j ($j = 1, 2, \ldots N$) denote the numerical value of the response variable (dependent variable) for the j^{th} experiment (here, tablet formulation).

The sum of squares (s_{ii}) and the sum of products (S_{im}, $i \neq m$) associated with the K formulation factors considered are as follows:

$$\text{Sum of Squares} = S_{ii} = \left[\sum_{j=1}^{N} X_{ij}^2 - \left(\sum_{j=1}^{N} X_{ij} \right)^2 N^{-1} \right] \qquad \text{(Eq. 1)}$$

where i = 1, 2, ... K

$$\text{Sum of Products} = S_{im}\ (i \neq m) = \left[\sum_{j=1}^{N} X_{ij}X_{mj} - \left(\sum_{j=1}^{N} X_{ij}\right)\left(\sum_{j=1}^{N} X_{mj}\right) N^{-1}\right] \quad \text{(Eq. 2)}$$

where i = 1, 2, ... K and m = 1, 2, ... K.

There would be a total of K sum of squares and 1/2 K(K-1) sum of products associated with K process variables. There would be 5 sum of squares and 10 sum of products when 5 process variables are considered. The sum of squares and sum of product quantities are arranged in a square matrix form as follows:

$$\begin{bmatrix} S_{11} & S_{12} & S_{13} \ldots & S_{1K} \\ S_{21} & S_{22} & S_{23} \ldots & S_{2K} \\ . & . & . \ . \ . \ . & . \\ . & . & . \ . \ . \ . & . \\ . & . & . \ . \ . \ . & . \\ S_{K1} & S_{K2} & S_{K3} & S_{KK} \end{bmatrix}$$

This matrix is called the information-matrix and is denoted by the expression (X'X). The sum of squares are arranged in the main diagonal of the matrix and the sum of products are placed in their respective off-diagonal positions. Since $S_{pq} = S_{qp} (p \neq q;$ p,q = 1, 2, ... 5), one has only ½ K (K+1) distinct elements in the matrix. If there are K process variables, then the dimension of this matrix is (KxK) with K-rows and K-columns and there are K^2 elements in the matrix. When K = 5, the dimension of (X'X) is(5x5) with 25 elements in the matrix.

The sum of squares and the sum of products associated with Y_j, the response parameter under consideration, are defined as follows:

$$\text{Sum of Square for } Y = SSQ(Y) = \sum_{j=1}^{N} Y_j^2 - \left(\sum_{j=1}^{N} Y_j\right)^2 N^{-1} \quad \text{(Eq. 3)}$$

$$\text{Sum of Products of X and Y} = R_i = \sum_{j=1}^{N} X_{ij}Y_j - \left(\sum_{j=1}^{N} X_{ij}\right)\left(\sum_{j=1}^{N} Y_j\right) N^{-1}$$

$$i=1, 2, \ldots K \quad \text{(Eq. 4)}$$

Now, if the R_i-quantities are arranged in a column denoted by (X'Y), the resulting expression is as follows:

$$(X'Y) = \begin{bmatrix} R_1 \\ R_2 \\ \cdot \\ \cdot \\ \cdot \\ R_K \end{bmatrix} \quad \text{(Eq. 5)}$$

The column in (Eq. 5) is called the vector of the right hand side of the normal equations, which have the following form:

$$(X'X)\,\hat{\beta} = (X'Y) \quad \text{(Eq. 6)}$$

where $\hat{\beta}$'s are the K unknown partial regression coefficients to be estimated from the normal equations. Rearranging (Eq. 6), the normal equations can be expressed as:

$$\hat{\beta} = (X'X)^{-1}(X'Y) \quad \text{(Eq. 7)}$$

where $(X'X)^{-1}$ denotes the inverse of the matrix (X'X). This equation is used to calculate the numerical values of the

β-coefficients as a function of the independent and dependent variables.

The computation of the inverse of (X'X) matrix is illustrated by considering a (2x2) matrix, in the following.

Let

$$(X'X) = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \qquad \text{(Eq. 8)}$$

$$(X'X)^{-1} = \begin{bmatrix} S_{22}/D & -S_{12}/D \\ -S_{21}/D & S_{11}/D \end{bmatrix} \qquad \text{(Eq. 9)}$$

where $D = S_{11}S_{22} - S_{12}S_{21}$, which is called the determinant of the matrix (X'X). The computation of the inverse of a matrix of higher dimension (K >2) is complex and one ought to resort to a computer for its resolution.

An explicit form of (Eq. 7) has the following structure:

$$\begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_K \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \cdots & S_{1K} \\ S_{21} & S_{22} \cdots & S_{2K} \\ \cdot & \cdot \cdots & \cdot \\ \cdot & \cdot \cdots & \cdot \\ \cdot & \cdot \cdots & \cdot \\ S_{K1} & S_{K2} & S_{KK} \end{bmatrix}^{-1} \begin{bmatrix} R_1 \\ R_2 \\ \cdot \\ \cdot \\ \cdot \\ R_K \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{1K} \\ C_{21} & C_{22} & C_{2K} \\ \cdot & \cdot \cdots & \\ \cdot & \cdot \cdots & \\ \cdot & \cdot \cdots & \\ C_{K1} & C_{K2} & C_{KK} \end{bmatrix} \begin{bmatrix} R_1 \\ R_2 \\ \\ \\ \\ R_K \end{bmatrix}$$

where C_{ij}, i = 1, 2, ... K and j = 1, 2, ... K, are the elements of $(X'X)^{-1}$ matrix.

After the solution of the simultaneous equations, the numerical estimates of b_1, b_2 ... b_K are obtained. The regression sum of squares are calculated as follows:

Regression Sum of Squares = Reg SSQ

$$= b_1R_1 + b_2R_2 + . + b_KR_K = \sum_{i=1}^{K} b_iR_i$$

Now a new quantity called the R^2 (associated with the K-variable system) is calculated as follows:

$$\%R^2 = \left[\sum_{i=1}^{K} b_iR_i/SSQ(Y) \right] \times 100 = \frac{\text{Reg SSQ}}{SSQ(Y)} \times 100$$

This quantity represents the proportion of the total variation due to regression. It ranges from 0% to 100%. It provides an index of how well the independent variables, being considered, explain the variation in the dependent variable. Consider the fact that one is studying the effect of two independent variables on a given dependent variable. Suppose that the following results are obtained from the various regression analyses:

Regression Analysis	R^2 Value in %
$(X_1; Y)$	60%
$(X_2; Y)$	10%
$(X_1, X_2; Y)$	61%

From the results, it would be clearly inferred that, X_2-variable does not contribute substantially to the variation in Y. In this situation, the X_1-variable is the predominantly important variable.

The general principle of APR in selecting the key independent variables follows essentially the procedure given above. If there are K independent variables, then one needs to do (2^K-1) regression analyses on the same dependent variable. For K = 5, there are $(2^5-1) = 31$ possible regression analyses. The (2^K-1) regression analyses breakdown into the following sets of regression analyses:

Sets	Symbolized Representation of Each Regression Analysis (In Brackets)	No of Regression Analyses In General	No of Regression Analyses K = 5
Considering One Variable at a Time	$[X_1]$, $[X_2]$.....$[X_K]$	K	5
Considering Two Variables at a Time	$[X_1X_2]$,$[X_1X_3]$..$[X_{K-1}X_K]$	(1/2)K(K-1)	10
Considering Three Variables at a Time	$[X_1,X_2,X_3]$....$[X_{K-2}X_{K-1}X_K]$	(1/6)K(K-1)(K-2)	10
.	.	.	.
.	.	.	.
.	.	.	.
Considering All Variables at a Time	$[X_1,X_2,X_3$ .....$X_K]$	1	1
TOTAL		(2^K-1)	31

A detailed examination of the R^2-values associated with each of the (2^K-1) regression analyses is conducted after arranging them in an ascending order of magnitude within each set. The relative magnitude of the R^2-values (within a set and between sets) plays an important role in the selection process of the key formulation factors.

The procedure for procuring a computer program whose codes have been written in FORTRAN IV language, is provided in reference (2). The program accomplishes all these complex calculations in a matter of a few minutes.

SWR Procedure

This selection procedure heavily depends upon the correlation, partial correlation, linear regression and multiple regression analyses of the data involving statistical tests of significance. It takes into account one variable at a time and selects the set of key factors in a step-wise manner.

It would be easier to explain the theory and procedure by considering a specific situation involving only three independent variables. The theory and the procedure have been developed in such a way that they are applicable to any number of independent variables.

Let the independent variables be denoted by the symbols X_1, X_2, and X_3 and let the dependent variable be denoted by Y. The following steps are considered in the statistical computational procedures leading to the final selection process of the key formulation factors:

STEP I:

Simple correlation coefficients between the dependent variable and independent variables considered are calculated by the following formula:

$$r_{iy} = \frac{\Sigma X_i Y - (\Sigma X_i)(\Sigma Y)\ N^{-1}}{[(\Sigma X_i^2 - (\Sigma X_i)^2\ N^{-1})(\Sigma Y^2 - (\Sigma Y_i)^2\ N^{-1})]^{1/2}}$$

where, r_{iy} is the simple correlation coefficient of X_i (i = 1, 2, 3) and Y.

In this case, there are three such coefficients, symbolically denoted by

$$r_{1y},\ r_{2y} \text{ and } r_{3y}$$

The magnitude of the correlation coefficient varies from -1 to +1. The next step entails the selection of the independent variable whose correlation with the dependent variable has the highest magnitude (ignoring sign). For the sake of this discussion, let r_{3y} be the one with the highest correlation value.

STEP II:

One now proceeds with the linear regression analysis of Y on X_3.

Let $R(b_0)$ represent the sum of squares associated with the mean. Let $R(b_3|b_0)$ represent the corrected regression sum of squares associated with the independent variable X_3. This represents the contribution of X_3 in explaining the variation in Y. The structure of the RANOVA (Analysis of Variance for Regression) is as follows:

RANOVA I

Source	DF	F-test
$R(b_0)$	1	
$R(b_3\|b_0)$	1	*
Residual	(N-2)	
Total	N	

* Implies significance at a pre-chosen level of significance.

Now one goes through the usual test of significance by computing the following quantity,

$$F_{1DF,\ \text{Residual DF}} = \frac{\text{Sum of Square of the "Source" to be Tested}}{\text{Residual Sum of Square/Residual DF}}$$

If this calculated F-value exceeds the tabular value found at the intersection of 1 DF (column) and residual DF (row) at a pre-selected level of significance (see F-table given in reference (3) pages 306 and 307), then that specific source of variation is declared to be significant at the level of significance chosen.

If $R(b_3|b_0)$ is significant at the pre-chosen level of significance, then X_3 is retained for the subsequent analysis associated with the procedure. (If this is not significant, it may be desirable to consider a higher level of significance and continue with the procedure.)

For the purpose of this discussion, let $R(b_3|b_0)$ be significant at the level of significance selected for the study. (It should be noted that this pre-selected level of significance must be maintained

throughout the analysis). Since $R(b_3/b_0)$ is significant, X_3 is retained.

STEP III:

Now, the first order partial correlation between each of the remaining independent variables, here, X_1 and X_2 and the dependent variable, keeping X_3 fixed (that is, adjusted for X_3) are calculated by the following formula:

$$r_{iy.3} = \frac{r_{iy} - r_{i3}r_{y3}}{[(1 - r_{i3}^2)(1 - r_{y3}^2)]^{1/2}}$$

$$i = 1 \text{ and } 2,\ i \neq 3$$

In the case under consideration, there are only two first order partial correlation coefficients,

$$r_{1y.3} \text{ and } r_{2y.3}$$

Select the independent variable whose partial correlation with the dependent variable has the highest magnitude (ignoring sign). For the sake of this discussion, let $r_{1y.3}$ be the one with the highest magnitude.

STEP IV:

Now proceed with the multiple regression analysis of Y on X_1 and X_3, simultaneously.

As before, let $R(b_0)$ represent the sum of squares of the mean. Let $R(b_1|b_0)$ represent the contribution of X_1 to the total variation in Y. Let $R(b_3|b_1,b_0)$ represent the contribution of X_3 after X_1 (alone) has contributed its share of variation to the total variation in Y. Let $R(b_1|b_3,b_0)$ represent the contribution

of X_1 after X_3 (alone) has contributed its share of variation to the total variation in Y. The two possible RANOVA's have the following structure:

RANOVA II-A			RANOVA II-B		
Source	DF	F-test	Source	DF	F-test
$R(b_0)$	1		$R(b_0)$	1	
$R(b_1\|b_0)$	1	*	$R(b_3\|b_0)$	1	*
$R(b_3\|b_1,b_0)$	1	*	$R(b_1\|b_3,b_0)$	1	*
Residual	N-3		Residual	N-3	
Total	N		Total	N	

* Implies significance at a pre-assigned level of significance.

Now one goes through the usual tests of significance of the sources, using the procedure provided in Step II. If $R(b_3|b_1,b_0)$ turns out to be significant in RANOVA II-A, then X_3 is retained. If it is not significant, X_3 is completely eliminated from the subsequent analyses. The logic behind this accept-reject decision rule is as follows. Consider the fact the $R(b_3|b_1,b_0)$ is not significant. This implies that X_1 alone is responsible for a larger proportion of the total variation in Y. In other words, X_3 is a poor contributor in the presence of X_1, even though it was a predominant factor in STEP II of the procedure.

The variable X_1 is retained for the subsequent analyses, if $R(b_1|b_0)$ in RANOVA II-A and $R(b_3|b_1,b_0)$ in RANOVA II-B are both significant. It is eliminated if any one of the tests failed to achieve significance.

For the purpose of this discussion, we have considered the retention of both the variables (see RANOVA II-A and II-B).

STEP V:

Now the second order partial correlation between the remaining independent variable, X_2 and the dependent variable keeping X_1 and X_3 fixed (that is, adjusted for both X_1 and X_3) are calculated by the following second order partial correlation formula:

$$r_{2y.3,1} = \frac{r_{2y.3} - r_{21.3}r_{y1.3}}{[(1 - r^2_{21.3})(1 - r^2_{y1.3})]^{1/2}}$$

For this example, there is only one second order partial correlation coefficient, namely, $r_{2y.3,1}$. This value is computed for the record only, and not for any numerical comparison, since X_2 is the last variable to enter the analysis. In otherwords, X_2, being the last variable, is automatically chosen for the subsequent analyses, regardless of the magnitude of its second order partial correlation coefficient.

STEP VI:

As in STEP IV, proceed with a multiple regression analysis of Y on X_2, X_1 and X_3, simultaneously.

As before, let $R(b_0)$ denote the sum of squares for the mean. Let $R(b_2|b_0)$ represent the individual contribution of X_2 to the total variation in Y. Let $R(b_1|b_2,b_0)$ represent the contribution of X_1 after X_2 (alone) has contributed its share of variation to the total variation in Y. Let $R(b_3|b_1,b_2,b_0)$ represent the contribution of X_3 after X_1 and X_2 have contributed their share of

variation to the total variation in Y. The RANOVA has the following structure:

RANOVA III-A

Source	DF	F-test
$R(b_0)$	1	
$R(b_2\|b_0)$	1	*
$R(b_1\|b_2,b_0)$	1	*
$R(b_3\|b_2,b_1,b_0)$	1	NS
Residual	N-4	
Total	N	

* Implies significance at a pre-assigned level of significance

NS Implies non-significance at that level of significance

Now, one goes through the usual tests of significance of the sources isolated in RANOVA III-A using the procedure outlined in Step II. Since $R(b_3|b_2,b_1,b_0)$ is not significant in RANOVA III-A, X_3 is completely eliminated from the subsequent analyses. The interpretation of this result is that X_3 did not prove to be a true key process variable in the presence of both X_1 and X_2. Note that the following two additional RANOVA's are conducted only to confirm the fact that X_1 and X_2 still continue to be the contributory variables in the absence of X_3.

RANOVA III-B			RANOVA III-C		
Source	DF	F-test	Source	DF	F-test
$R(b_0)$	1		$R(b_0)$	1	
$R(b_2\|b_0)$	1	*	$R(b_1\|b_0)$	1	*
$R(b_1\|b_2,b_0)$	1	*	$R(b_2\|b_1,b_0)$	1	*
Residual	N-3		Residual	N-3	
Total	N		Total	N	

* Implies significance at a pre-assigned level of significance.

The magnitude of $R(b_1|b_2,b_0)$ in RANOVA III-B should be the same as that of $R(b_1|b_2,b_0)$ in RANOVA III-A. If $R(b_1|b_2,b_0)$ in RANOVA III-B and $R(b_2|b_1,b_0)$ in RANOVA III-C are significant, then X_1 and X_2 are retained. Since there are no other variables to enter the analysis, it would be perfectly reasonable to declare that X_1 and X_2 are the two key factors substantially contributing to the variation in the dependent variable.

General Procedure for Combining Results of APR and SWR Analysis (CAS):

The primary purpose of combining the results of APR and SWR analyses is to facilitate a clear interpretation of the findings of the two analyses which are essentially complementary to each other. The combined results when used properly does provide an insight into the specific role played by each of the variables selected by the two procedures. New quantities are derived as a function of the results of the two analyses. The derivation of these quantities do constitute an integral part of the overall analytical procedure. In this section, the procedure for deriving

the quantities has been described in general terms, primarily, symbolically. The appropriate interpretation of these quantities has been examplified in the section on "Results and Discussion" based on actual experimental data.

Consider a multiple regression situation in which there are K independent variables and one dependent variable denoted by Y*. An APR analysis is conducted based on these variables and let A_1, A_2, A_3 and A_4 (for example) denote the four sets of X-variables whose R^2-values are either equal to or close to the maximum R^2-value attainable for the dependent variable Y*. The sets are not necessarily mutually exclusive and the number of X-variables associated with these sets are not necessarily equal.

A SWR analysis is conducted based on the same K independent variables and the dependent variable Y*. Let $S(\alpha_1)$, $S(\alpha_2)$ and $S(\alpha_3)$ (where, $\alpha_3 < \alpha_2 < \alpha_1$) denote the three sets of X-variables selected by the three separate SWR analyses based on α_1, α_2 and α_3 levels of significance, respectively. Usually, in practice, one considers the numerical values of α_1, α_2 and α_3 to be 0.1, 0.5 and 0.01 levels of significance, respectively. The following relationships generally hold for these sets:

1. $S(\alpha_3) \subseteq S(\alpha_2) \subseteq S(\alpha_1)$
2. $S(\alpha_3) \cap S(\alpha_2) \cap S(\alpha_1) = S(\alpha_3)$

where, the symbols $\subseteq$ and $\cap$ imply "contained in" and "intersection" of the sets, respectively. Now consider, in the following, the derivation of the quantities to be interpreted in the next section:

In this derivation, it is assumed that none of the sets A_1, A_2, A_3, A_4, $S(\alpha_1)$, $S(\alpha_2)$ and $S(\alpha_3)$ are null sets (ϕ) meaning that there is at least one member (independent variable) in each set. In other words, the sets are not empty. For different pharmaceutical systems, the number of sets associated with APR and SWR would differ. For this reason, the formulas defined below, when used in practice, must follow a "sequence rule", e.g. suppose A_4 does not exist in APR sets, then A_3 or A_2 or A_1 should be substituted in the formula in that sequence depending upon whether they exist or do not exist, and also in the case of SWR-sets, suppose $S(\alpha_3)$ does not exist, then $S(\alpha_2)$ or $S(\alpha_1)$ should be substituted in the formula in that sequence depending upon whether they exist or do not exist. The general rule should be that, whenever a set is ϕ, meaning that it does not exist, that set should not be used in the formula under consideration.

1. Minimum Central Subset (θ_1), derived by the formula:

$$\theta_1 = A_1 \cap A_2 \cap A_3 \cap A_4 \cap S(\alpha_3)$$

2. Central Subset (θ_2), derived by the formula:

$$\theta_2 = A_1 \cap A_2 \cap A_3 \cap A_4 \cap S(\alpha_1) \cap S(\alpha_2)$$

3. Let the following relationships among R^2-values in the APR analysis hold

$$R^2(A_1) \geq R^2(A_2) \geq R^2(A_3) \geq R^2(A_4)$$

On the SWR analysis, the following relationship must hold:

$$R^2[S(\alpha_3)] \leq R^2[S(\alpha_2)] \leq R^2[S(\alpha_1)]$$

In general, one finds:

$$R^2[S(\alpha_i)] \leq R^2(A_j), \quad i = 1, 2 \text{ and } 3; \; j = 1, 2, 3 \text{ and } 4$$

Now let $\Omega = S(\alpha_1) \cup A_4$ constitute the whole set, where, the symbol U denotes the union of the two sets (on either side of the symbol U).

Complementary Subset (θ_3) is derived by the following formula:

$$\theta_3 = [S(\alpha_1) \cap A_4]^c$$

where the symbol c denotes the complement of the set.

Note: The APR-set (A_1, A_2, A_3 or A_4) which is used to derive the complementary set should have a R^2-value close to the maximum R^2-value attainable. The chosen set should be shown on the table of results. If $R^2[S(\alpha_1)] \simeq R^2(A_4)$, then Set $\theta_3 = \phi$, automatically.

4. Suppose there are two sets associated with APR analysis (say A_1 and A_2) which have the same number of members, and $R^2(A_1) \simeq R^2(A_2)$. Now let $\Omega = A_1 \cup A_2$ be the whole set.

   Interchangeable Subset (θ_4) is derived by the following formula:

   $$\theta_4 = [A_1 \cap A_2]^c$$

   θ_4 contains members of both A_1 and A_2, when $\theta_4 \neq \phi$, the null set (meaning there is not even a single member in the set). Let them be identified by A_1^* belonging to A_1 and A_2^* belonging to A_2. The advantages of detecting interchangeable variables will be elaborated in the next section.

RESULTS AND DISCUSSION

The results of APR, SWR and CAS analyses have been presented in TABLE IV through TABLE XIII, devoting one table to each parameter. It was proposed to present the results of each parameter in four sections to facilitate description and interpretation of the findings of the analysis. Sections 1, 2 and 3 were devoted to the results of APR, SWR and CAS analyses, respectively. The fourth section was primarily devoted to presenting the following regression information associated with each parameter, an explicit form of the regression equation, its R^2-value and the R^2-value for the second order regression function based on the key process variables selected by APR, SWR and CAS analyses.

For each parameter, APR analysis provided several sets (A_1, A_2, A_3, etc.) of process variables (which ranged from 2 to 6 sets depending upon the parameter) chosen on the basis of the closeness of their respective R^2-values to the maximum attainable R^2-value for the parameter under consideration. For instance, four such sets were chosen for dissolution (see TABLE IV) and only a couple of sets were chosen for disintegration time (see TABLE V).

For each parameter, SWR analysis provided three sets of process variables based on the results of $S(\alpha_1 = 0.10)$, $S(\alpha_2 = 0.05$ and $S(\alpha_3 = 0.01)$, respectively. Each parameter, then, has gone through four separate and independent analyses, one APR analysis and three SWR analyses based on the three pre-chosen levels of significance (here, we have considered $\alpha_1 = 0.10$, $\alpha_2 = 0.05$ and $\alpha_3 = 0.01$). It may be noted here that these three levels have been very useful

TABLE IV

RESULTS OF APR, SWR AND CAS ANALYSES

Parameter: Dissolution (t_{30})

Section	Analysis	Sets	Formulation Factors	R^2%
1	APR	A_1	X_1, X_2, X_3, X_4, X_5	66.0
		A_2	X_2, X_3, X_4, X_5	65.6
		A_3	X_1, X_2, X_3, X_5	65.2
		A_4	X_2, X_3, X_5	64.8
2	SWR	SWR(α_1 = 0.10)	X_2, X_5	62.2
		SWR(α_2 = 0.05)	X_2, X_5	62.2
		SWR(α_3 = 0.01)	X_5	51.3
3	CAS	Minimum Central Subset	X_5	
		Central Subset	X_2, X_5	
		Complementary Subset (A_4)	X_3	
		Interchangeable Subset	$X_4 \Leftrightarrow X_1$*	
4	Structural Regression Equation and R^2-values			
	$\hat{Y} = 69.91 - 37.3X_5 - 17.48X_2 + 4.24X_3$			
	R^2 = 64.8% (with X_2, X_3 and X_5)			
	R^2 = 87.2% (for the second order regression function)			

*Provided for completeness only.

TABLE V

RESULTS OF APR, SWR AND CAS ANALYSES

Parameter: Disintegration

Section	Analysis	Sets	Formulation Factors	R^2%
1	APR	A_1	X_1, X_2, X_3, X_4, X_5	83.6
		A_2	X_2, X_3, X_4, X_5	83.5
2	SWR	SWR(α_1 = 0.10)	X_2, X_3, X_4, X_5	83.5
		SWR(α_2 = 0.05)	X_2, X_3, X_4, X_5	83.5
		SWR(α_3 = 0.01)	X_2, X_3, X_4, X_5	83.5
3	CAS	Minimum Central Subset	X_2, X_3, X_4, X_5	
		Central Subset	X_2, X_3, X_4, X_5	
		Complementary Subset	ϕ	
		Interchangeable Subset	ϕ	
4	Structural Regression Equation and R^2-values			
	$\hat{Y} = 1.45 + 9.98X_2 - 2.30X_3 + 6.90X_4 + 6.49X_5$			
	R^2 = 83.5% (with X_2, X_3, X_4 and X_5)			
	R^2 = 91.8% (for the second order regression function)			

ϕ Null set, that is, there are no members (factors) which qualify for the subset considered.

TABLE VI

RESULTS OF APR, SWR AND CAS ANALYSES

Parameter: Hardness (Tablet Breaking Strength)

Section	Analysis	Sets	Formulation Factors	R^2%
1	APR	A_1	X_1, X_2, X_3, X_4, X_5	70.9
		A_2	X_2, X_3, X_4, X_5	70.9
		A_3	X_2, X_3, X_5	68.7
		A_4	X_2, X_4, X_5	68.7
2	SWR	SWR(α_1 = 0.10)	X_2, X_5	66.4
		SWR(α_2 = 0.05)	X_2, X_5	66.4
		SWR(α_3 = 0.01)	X_5	44.8
3	CAS	Minimum Central Subset	X_5	
		Central Subset	X_2, X_5	
		Complementary Subset (A_4)	X_3 or X_4	
		Interchangeable Subset	$X_3 \Leftrightarrow X_4$	
4	Structural Regression Equation and R^2-values			
	$\hat{Y} = 6.09 + 1.65X_2 + 0.53X_4 - 2.35X_5$			
	R^2 = 68.7% (with X_2, X_4 and X_5)			
	R^2 = 83.3% (with the second order regression function)			

TABLE VII

RESULTS OF APR, SWR AND CAS ANALYSES

Parameter: Friability

Section	Analysis	Sets	Formulation Factors	R^2%
1	APR	A_1	X_1, X_2, X_3, X_4, X_5	42.8
		A_2	X_1, X_2, X_4, X_5	42.7
		A_3	X_1, X_3, X_4, X_5	39.9
		A_4	X_1, X_4, X_5	39.8
2	SWR	SWR(α_1 = 0.10)	X_1, X_5	35.6
		SWR(α_2 = 0.05)	X_1, X_5	35.6
		SWR(α_3 = 0.01)	ϕ	
3	CAS	Minimum Central Subset	X_1, X_5	
		Central Subset	X_1, X_5	
		Complementary Subset (A_2)	X_2 and X_4	
		Interchangeable Subset	$X_2 \lessdot\!\!\!\!\approx\!\!\!\!\gtrdot X_3$	
4	Structural Regression Equation and R^2-values			
	$\hat{Y} = 1.54 - 0.018X_1 + 0.125X_2 - 0.149X_4 - 0.256X_5$			
	R^2 = 42.7% (with X_1, X_2, X_4 and X_5)			
	R^2 = 67.3% (for the second order regression function)			

ϕ Null set, that is, there are no members (factors) which qualify for the subset considered.

$\lessdot\!\!\!\!\approx\!\!\!\!\gtrdot$ "almost interchangeable".

TABLE VIII

RESULTS OF APR, SWR AND CAS ANALYSES

Parameter: Weight Uniformity

Section	Analysis	Sets	Formulation Factors	R^2 %
1	APR	A_1	X_1, X_2, X_3, X_4, X_5	34.1
		A_2	X_1, X_2, X_3, X_5	34.0
		A_3	X_1, X_2, X_3, X_5	32.2
		A_4	X_1, X_2, X_5	32.1
2	SWR	$SWR(\alpha_1 = 0.10)$	X_1	22.5
		$SWR(\alpha_2 = 0.05)$	ϕ	
		$SWR(\alpha_3 = 0.01)$	ϕ	
3	CAS	Minimum Central Subset	X_1	
		Central Subset	X_1	
		Complementary Subset (A_4)	X_2 and X_5	
		Interchangeable Subset	$X_3 \dot{\Leftrightarrow} X_4$*	
4	Structural Regression Equation and R^2-values			
	$\hat{Y} = 1.92 - 0.022X_1 + 0.24X_2 - 0.16X_5$			
	$R^2 = 32.1\%$ (with X_1, X_2, and X_5)			
	$R^2 = 60.8\%$ (for a second order regression function)			

ϕ Null set, that is, there are no members (factors) which qualify for the subset considered

$\dot{\Leftrightarrow}$ "almost interchangeable"

* Provided for completeness only.

and provided the necessary information for arriving at the correct conclusions, especially in the pharmaceutical formulation research.

For each parameter, CAS analysis has provided four interpretable subsets which have been constructed by using the formulae given in the CAS procedure above.

The following general questions may be entertained at this point. How do all these results lead to the selection of a set of key process variables? How does one go about determining the most important variable or variables among the key process variables selected? What is the specific role played by each of the key formulation factors selected by the three procedures employed? To answer all these questions by illustrations, consider the results given in TABLE IV. Examining the results of APR analysis in the table (see Section 1), one finds that the maximum attainable R^2-value for this parameter is 66.0% involving all the process variables in the system.

However, using only the following three process variables, X_2, X_3 and X_5 (subset A_4) one obtains a R^2-value of 64.8%, which is only 1.2% less than the maximum attainable. This clearly implies that X_1 and X_4 are not contributing substantially to the variation in the dependent variable under consideration. SWR analysis confirms these findings by reproducing some of the variables associated with subset A_4 in APR analysis. It also independently demonstrates the fact that X_5 and X_2 are the two most important process variables in the system with respect to the dissolution response. However, by using only these two variables in the regression equation, we

can only attain a R^2-value of 62.2% which is 3.8% less than the maximum attainable. The process variables in subset A_4 of APR analysis indicate that the inclusion of X_3 in the regression equation with X_5 and X_2 not only enhances the R^2-value of the equation, but also brings the R^2-value close to the maximum attainable for this parameter. This discussion essentially answers the first of the three questions. The answer to the next question is implicit in the results of SWR analysis. Examinig the results of $S(\alpha_3 = 0.01)$, one comes to a reasonable conclusion that X_5 is the most important formulation factor of the key variables selected. The answer to the third question is found in the results of CAS analysis and requires some elaboration.

In this analysis, "Central Subset" consists of the process variables which are the most important variables among the key formulation factors selected by the procedures. "Minimum Central Subset" provides the most indispensible variables among the process variables associated with "Central Subset". These variables do contribute substantially to the variation in the dependent variable. It is possible that "Minimum Central Subset" may contain the same variables associated with "Central Subset", implying the fact that the variables in "Central Subset" can not be decomposed further into dispensible and indispensible key process variables. In such cases, it may be inferred that all the variables in "Central Subset" would be contributing jointly (and inseparably) to the variation in the dependent variable. This phenomenon can be observed in eight out of the ten parameters studied e.g. in disintegration (Table V) and

friability (Table VII). The exceptions to this phenomenon are found in the results given in Table IV and Table VI, in which case "Minimum Central Subset" is not identical to "Central Subset" and the identification of the most indispensible variable is possible in such cases.

The primary role played by the process variables associated with "Complementary Subset" is to enhance the R^2-value of the regression equation in which these variables appear along with the variables associated with "Central Subset". High R^2-values always lead to increased precision in the estimation of regression parameters and enhance the predictive efficiency of the estimated regression equation.

The interpretation of "Interchangeable Subset" is considered next. The numerical example given in TABLE VI will be used for the illustration of the meaning of this subset. The key factors influencing the tablet hardness are X_2, X_4 and X_5. In otherwords, one can impose changes in these process variables to effect a desirable change in the response variable. However, since CAS analysis shows that X_4 is interchangeable with X_3, one could elect to use X_2, X_3 and X_5 in the regression equation instead of X_2, X_4 and X_5. Controlling in either set would bring about the same desirable change in the response variable. The practical utility of this property can easily be demonstrated by using the results given in TABLE XIV. It may be observed that one of the variables influencing dissolution is X_3. By effecting appropriate changes in X_2, X_3 and X_5, the conditions of dissolution requirements could easily be met.

Since X_4 is not a key formulation factor for dissolution, it would be advisable to impose changes in X_4 (rather than X_3) to bring about desirable changes in hardness if one is interested in bringing about changes in both the parameters considered. By following this strategy, one does not disturb the desirable dissolution profile previously attained.

There may also be economic advantages to this since one of the interchangeable variables may be more economical to use than the other variable. The detection of interchangeable variables are extremely important and CAS analysis has the appropriate tools to detect such variables, if they exist.

In the light of these discussions, the results provided in the tables should be examined in detail. In this discussion, the general principles of interpreting the results and the mechanism of arriving at a correct conclusion have been presented. These tools can now be used to elaborate on the results.

A summary of the salient results of APR, SWR and CAS analyses for each of the ten parameters has been presented in TABLE XIV. It is clearly observed that compression pressure, magnesium stearate, calcium phosphate/lactose ratio were identified as the key factors influencing the formulation in nine, seven and six parameters (out of the ten parameters), respectively. Most of the "interchangeability" was observed between starch disintegrant and granulating gelatin. It may be noted that calcium phosphate/lactose ratio was not identified as a key factor for disintegration, dissolution (t_{30}) or hardness, the most important tablet properties for the system

TABLE IX

RESULTS OF APR, SWR AND CAS ANALYSES

Parameter: Thickness Uniformity

Section	Analysis	Sets	Formulation Factors	R^2 %
1	APR	A_1	X_1, X_2, X_3, X_4, X_5	50.0
		A_2	X_1, X_2, X_3, X_4	49.4
		A_3	X_1, X_2, X_3, X_5	46.7
		A_4	X_1, X_2, X_4	46.0
2	SWR	SWR(α_1 = 0.10)	X_1	37.8
		SWR(α_2 = 0.05)	ϕ	
		SWR(α_3 = 0.01)	ϕ	
3	CAS	Minimum Central Subset	X_1	
		Central Subset	X_1	
		Complementary Subset (A_2)	X_2, X_3 and X_4	
		Interchangeable Subset	$X_3 \dot{\Leftrightarrow} X_5$	
4	Structural Regression Equation and R^2-values			

$$\hat{Y} = 1.03 - 0.0113X_1 + 0.073X_2 + 0.034X_3 + 0.076X_4$$

R^2 = 49.4% (with X_1, X_2, X_3 and X_4)

R^2 = 63.0% (for a second order regression function)

ϕ Null set, that is, there are no members (factors) which qualify for the subset considered.

$\dot{\Leftrightarrow}$ "almost interchangeable"

under consideration (5). It is clearly observed that compression pressure and magnesium stearate were the key factors for each of these three parameters. Corn starch was the key factor for disintegration and dissolution only. Granulating gelatin was identified as a key factor for disintegration alone.

On the basis of these findings, one can begin to analyze the mechanism of action of the key factors in his/her system. For example, the hydrophobic nature of the magnesium stearate (X_5) causes a decrease in dissolution and an increase in disintegration time; its presence also causes a reduction in tablet hardness, probably by preventing the binding of the granules which apparently are cohesive in its absence. This prevention of adequate binding makes it important in friability and table breakage, as well. Compressional Force (X_2) obviously affects tablet hardness and works in opposition to the magnesium stearate. For dissolution and disintegration the compressional force works in the same direction as the magnesium stearate; this is perhaps accomplished by its effect on the tablet pore volume. It is interesting to note that the only key factor for the Pore Volume response (Table X) is the compression pressure, X_2. As the pressure is increased, the pore volume decreases.

The corn starch disintegrant (X_3) is indeed performing its function and is shown as a key factor in disintegration. The fact that it is also a key factor in dissolution seems to indicate that in this formulation and by the test methods utilized, disintegration and dissolution are closely related.

TABLE X

RESULTS OF APR, SWR AND CAS ANALYSES

Parameter: Porosity

Section	Analysis	Sets	Formulation Factors	R^2%
1	APR	A_1	X_1, X_2, X_3, X_4, X_5	87.7
		A_2	X_1, X_2, X_4, X_5	87.7
		A_3	X_1, X_2, X_3, X_5	87.7
		A_4	X_1, X_2, X_3, X_4	87.7
		A_5	X_2, X_3, X_4, X_5	87.1
2	SWR	SWR(α_1 = 0.10)	X_2	87.1
		SWR(α_2 = 0.05)	X_2	87.1
		SWR(α_3 = 0.01)	X_2	87.1
3	CAS	Minimum Central Subset	X_2	
		Central Subset	X_2	
		Complementary Subset	ϕ	
		Interchangeable Subset	ϕ	
4	Structural Regression Equation and R^2-value			
	$\hat{Y} = 0.078 - 0.31X_2$			
	R^2 = 87.1% (with X_2)			

ϕ Null set, that is, there are no members (factors) which qualify for the subset considered.

TABLE XI

RESULTS OF APR, SWR AND CAS ANALYSES

Parameter: Tablet Breakage

Section	Analysis	Sets	Formulation Factors	R^2 %
1	APR	A_1	X_1, X_2, X_3, X_4, X_5	45.7
		A_2	X_1, X_2, X_4, X_5	45.4
		A_3	X_1, X_3, X_4, X_5	42.4
		A_4	X_1, X_4, X_5	42.1
2	SWR	SWR(α_1 = 0.10)	X_1	31.6
		SWR(α_2 = 0.05)	X_1	31.6
		SWR(α_3 = 0.01)	X_1	31.6
3	CAS	Minimum Central Subset	X_1	
		Central Subset	X_1	
		Complementary Subset (A_2)	X_2, X_4 and X_5	
		Interchangeable Subset	$X_2 \dot{\Longleftrightarrow} X_3$	
4	Structural Regression Equation and R^2-values			
	$\hat{Y} = 0.61 - 0.0091X_1 + 0.061X_2 - 0.082X_4 - 0.071X_5$			
	R^2 = 45.4% (with X_1, X_2, X_4 and X_5)			
	R^2 = 73.8% (with a second order regression function)			

$\dot{\Longleftrightarrow}$ "almost interchangeable".

TABLE XII

RESULTS OF APR, SWR AND CAS ANALYSES

Parameter: Granular Mean Diameter

Section	Analysis	Sets	Formulation Factors	R^2%
1	APR	A_1	X_1, X_2, X_3, X_4, X_5	62.2
		A_2	X_1, X_2, X_4, X_5	62.0
		A_3	X_1, X_3, X_4, X_5	61.4
		A_4	X_1, X_4, X_5	61.3
2	SWR	SWR(α_1 = 0.10)	X_1, X_4	56.8
		SWR(α_2 = 0.05)	X_1, X_4	56.8
		SWR(α_3 = 0.01)	X_1, X_4	56.8
3	CAS	Minimum Central Subset	X_1, X_4	
		Central Subset	X_1, X_4	
		Complementary Subset (A_4)	X_5	
		Interchangeable Subset	$X_2 \Leftrightarrow X_3$*	
4	Structural Regression Equation and R^2-values			
	$\hat{Y} = 0.44 - 0.0036X_1 + 0.05X_4 + 0.025X_5$			
	R^2 = 61.3% (with X_1, X_4 and X_5)			
	R^2 = 73.7% (with a second order regression function)			

* Provided for completeness only.

TABLE XIII

RESULTS OF APR, SWR, AND CAS ANALYSES

Parameter: Mean Pore Diameter

Section	Analysis	Sets	Formulation Factors	R^2%
1	APR	A_1	X_1, X_2, X_3, X_4, X_5	65.0
		A_2	X_1, X_2, X_3, X_5	64.8
		A_3	X_1, X_2, X_4, X_5	64.3
		A_4	X_1, X_2, X_3, X_4	64.2
		A_5	X_1, X_2, X_5	64.1
		A_6	X_1, X_2, X_3	64.0
2	SWR	SWR(α_1 = 0.10)	X_2	62.2
		SWR(α_2 = 0.05)	X_2	62.2
		SWR(α_3 = 0.01)	X_2	62.2
3	CAS	Minimum Central Subset	X_2	
		Central Subset	X_2	
		Complementary Subset (A_6)	X_1 and X_3	
		Interchangeable Subset	$X_3 \Longleftrightarrow X_4, X_5$	
4	Structural Regression Equation and R^2-values			
	$\hat{Y} = 0.307 - 0.000615X_1 - 0.091X_2 + 0.0047X_3$			
	R^2 = 64.0% (with X_1, X_2 and X_3)			
	R^2 = 88.7% (with a second order regression function)			

TABLE XIV

SUMMARY OF RESULTS OF APR, SWR AND CAS ANALYSES

PARAMETER	CALCIUM PHOSPHATE/LACTOSE RATIO X_1	COMPRESSION PRESSURE X_2	STARCH (CORN) DISINTEGRANT X_3	GRANULATING GELATIN X_4	MAGNESIUM STEARATE X_5
1. Disintegration		*	●	*	*
2. Dissolution (t_{30})		●	*		*
3. Hardness		●		(*)	*
4. Weight	*	●			*
5. Friability	●	●		(*)	*
6. Thickness	*	*	(*)	●	
7. Porosity		●			
8. Tablet Breakage	●	(*)		●	*
9. Granular Mean Diameter	*			●	*
10. Mean Pore Diameter	*	●	(*)		
TOTAL	6	9	4	6	7

* Represents key formulation factor for the parameter

(*) Indicates an interchangeable variable (see text and TABLES for details)

The gelatin in the granulating solution (X_4) is a key factor in the granule diameter and increases it as expected. Its negative contribution to disintegration may also reflect this binding capacity.

The calcium phosphate/lactose ratio (X_1), (although it exhibits little or no effect on the dissolution, disintegration or tablet hardness responses), is a key factor in friability and tablet breakage. Apparently, the granules formed with a higher level of lactose have better cohesive and binding properties or are more

elastic than those with a higher level of calcium phosphate. This may be related to its effect on granule diameter since larger and perhaps stronger granules are produced at higher lactose levels.

It should be noted that those factors which are demonstrated to be key factors in one formulation may not be so in another because of the physico-chemical interactions between all components in a given system. These results are applicable only to the system studied and each system will require its own analysis.

CONCLUSION

In this study, it was clearly shown that selective regression consisting of APR, SWR and CAS analyses played an important role in identifying the key formulation factors substantially contributing to the variation in each of the ten parameters considered in the study. Imposing controls on these selected key factors, one would be able to effect a favorable change in the response variable. It is interesting to note that compression pressure and magnesium stearate were identified as the two most important key process variables governing the variation in almost all the parameters in the system. Both factors are known to have the intrinsic property of impeding acceleration of the physico-chemical reaction associated with the process. It is noted that there were only three key factors associated with dissolution and hardness. However, there were four key factors associated with disintegration.

This information is vital in that one would be able to achieve economy in cost and time in controlling only those key formulation factors which are substantially contributing to the variation in the response variable.

ACKNOWLEDGEMENTS

The authors wish to thank Mr. N. Tonkonoh for carrying out the computations in the T.S.O system, and Miss Lil Filandino for the excellent job of typing the manuscript.

REFERENCES

(1) J. B. Schwartz, J. R. Flamholz and R. H. Press, JOURN. PHARM. SCI., 62, 1165 (1973).

(2) G. M. Furnival, TECHNOMETRICS, 13(2), 403 (1971).

(3) N. R. Draper and H. Smith, APPLIED REGRESSION ANALYSIS, John Wiley & Sons, Inc., New York, New York, pp. 163-195 (1968).

(4) A. Ralston and H. S. Wilf, eds., MATHEMATICAL METHODS FOR DIGITAL COMPUTERS, Vol. 1, John Wiley & Sons, Inc., New York, New York, pp. 191-203 (1960).

(5) N. R. Bohidar, F. A. Restaino and J. B. Schwartz, JOURN. PHARM. SCI., 64, 966 (1975).