Computer Optimization of Pharmaceutical Formulations I: General Procedure

JOSEPH B. SCHWARTZ[▲], JOEL R. FLAMHOLZ, and ROBERT H. PRESS

Abstract □ A technique was developed whereby a formulation with optimum properties, according to predetermined criteria, can be selected *via* computer analysis. The results of a statistically designed set of experiments based on five independent variables were used as the data input to the computer. Regression analysis of these data resulted in a set of second-order polynomial predictor equations. Restrictions placed by the pharmacist on the characteristics of interest serve to eliminate undesirable formulations. Ultimately, the formulator can trade off properties to obtain an optimum formulation. Computer-generated graphs further aid the formulator in the understanding of the particular system. An example is presented of the selection of an optimum formulation by this procedure and of its preparation in the laboratory. Predictions and experimental results show excellent agreement.

Keyphrases □ Computer optimization of pharmaceutical formulations—general procedure, examples □ Formulation of pharmaceutical preparations—computer optimization, general procedure and examples □ Pharmaceutical technology—general procedure for computer optimization of pharmaceutical formulations

The development of a pharmaceutical formulation and the associated process involves a number of variables. Mathematically, they can be divided into two groups: (*a*) the independent or formulation and process variables that are controllable factors, and (*b*) the dependent variables that are the responses or characteristics of the resultant drug delivery system. Much of the work in pharmaceutics has been in the pursuit of relationships between the two.

Table I—Experimental Design for Five Factors

Trial	Factor Level in Experimental Units				
	X_1	X_2	X_3	X_4	X_5
1	−1	−1	−1	−1	1
2	1	−1	−1	−1	−1
3	−1	1	−1	−1	−1
4	1	1	−1	−1	1
5	−1	−1	1	−1	−1
6	1	−1	1	−1	1
7	−1	1	1	−1	1
8	1	1	1	−1	−1
9	−1	−1	−1	1	−1
10	1	−1	−1	1	1
11	−1	1	−1	1	1
12	1	1	−1	1	−1
13	−1	−1	1	1	1
14	1	−1	1	1	−1
15	−1	1	1	1	−1
16	1	1	1	1	1
17	−1.547	0	0	0	0
18	1.547	0	0	0	0
19	0	−1.547	0	0	0
20	0	1.547	0	0	0
21	0	0	−1.547	0	0
22	0	0	1.547	0	0
23	0	0	0	−1.547	0
24	0	0	0	1.547	0
25	0	0	0	0	−1.547
26	0	0	0	0	1.547
27	0	0	0	0	0

Pharmaceutical research and development projects are often optimization problems. One experiments by a series of logical steps, carefully controlling the variables one at a time, until a satisfactory formulation results.

It may be necessary to trade off properties during such experimentation, to sacrifice one characteristic to improve another; *e.g.* to accept a softer tablet in order to improve the dissolution profile. Thus, the primary objective may not be to optimize absolutely but to compromise effectively and thereby to produce the best formulation under a given set of restrictions.

The techniques of optimization are well documented in the literature (1–7). Fonner *et al.* (8) demonstrated the applicability of mathematical optimization techniques to pharmaceutical systems. Their application of the Lagrangian method, while able to handle several responses or dependent variables, was generally limited to two independent variables. Experiences in these laboratories showed that a greater number of these variables are of general interest, but as the number of formulation factors is increased, the situation becomes more involved mathematically and graphically impossible. Although various multivariate techniques are available, computers are almost a necessity for their application.

The work reported here was undertaken to determine the feasibility of optimizing a pharmaceutical formulation with the aid of computers in a manner which could be used with relative ease by the development pharmacist. The procedure that resulted appears capable of being carried out by persons unfamiliar with the mathematics of optimization and with no previous computer experience.

This paper covers the general procedure followed to optimize formulations. Subsequent reports will discuss several other aspects of the optimization technique, its application to a practical situation, and the interpretation of the large quantity of data generated by this particular method.

THEORY

The type of project of concern here is one of selecting the most desirable level of ingredients or controllable process factors. That is, it is desired to *quantitate* a formulation that has been *qualitatively* determined.

The technique involves performing a set of statistically designed experiments as described by Box and Wilson (1, 2) and using the resulting data to derive a mathematical model which can be used for the optimization of a formulation. The experimental design is dependent on the number of variables involved in the study; for five independent or formulation variables, the modified half-factorial design shown in Table I requires a total of 27 experiments.

The first 16 experiments represent a half-factorial design for five factors at two levels resulting in $\frac{1}{2} \times 2^5 = 16$ trials. The two levels here are represented as +1 and −1. For the remainder of the study, three additional levels were selected: zero represented the base

Table II—Translation of Experimental Conditions into Physical Units[a]

Factor	−1.547 eu[b]	−1 eu	Base 0	+1 eu	+1.547 eu
X_1 = calcium phosphate–lactose ratio (1 eu = 10 mg.)	24.5/55.5	30/50	40/40	50/30	55.5/24.5
X_2 = compression pressure (1 eu = 0.5 ton)	0.25	0.5	1	1.5	1.75
X_3 = starch disintegrant (1 eu = 1 mg.)	2.5	3	4	5	5.5
X_4 = granulating gelatin (1 eu = 0.5 mg.)	0.2	0.5	1	1.5	1.8
X_5 = magnesium stearate (1 eu = 0.5 mg.)	0.2	0.5	1	1.5	1.8

[a] Physical quantities are rounded off only for convenience. [b] eu = experimental units.

level midway between the above-mentioned levels and the positive and negative 1.547 values represented the extreme values. The reader is referred to the literature for a discussion of the statistical design (1, 2, 5).

The proper name for this design is "a five-factor, orthogonal, central, composite, second-order design," and the type of predictor equation resulting from such a study is a second-order polynomial (with 21 terms) having the following form:

$$Y_i = a_0 + a_1X_1 \ldots + a_5X_5 + a_{11}X_1^2 \ldots + a_{55}X_5^2 + a_{12}X_1X_2 \ldots + a_{45}X_4X_5 \quad \text{(Eq. 1)}$$

where:

Y_i = level of a given response (dependent variable)
a = regression coefficients for second-order polynomial
X_j = level of independent variable

Such an equation is generated for each dependent variable relating it to the set of five independent variables.

EXPERIMENTAL

The formulation selected as a model system for the optimization program was a production formula already in existence. Briefly, the process involves a starch-paste–gelatin granulation of the drug with diluents. After drying and milling, the disintegrant, the glidant, and the lubricant are added.

The five independent formulation variables selected for this particular study were: X_1, diluent ratio; X_2, compressional force; X_3, disintegrant level; X_4, binder level; and X_5, lubricant level. With the exception of these five variables, everything else in the formulation and the processing steps remained constant throughout the study including: (*a*) the level of active ingredient; (*b*) the quantity of starch as paste; (*c*) the glidant level; (*d*) granulating, milling, drying, and blending conditions; and (*e*) the speed of the compressing machine. The translation of the statistical design (Table I) into physical units for the five variables is shown in Table II[1].

Each experiment consisted of a 30,000-tablet batch through the granulating, milling, dry mixing, and lubricating steps. From each batch, 3000 tablets were compressed on a rotary press[2] equipped with 35 sets of 0.6-cm. (0.25-in.) round, flat beveled-edge tooling run without precompression.

Table III—Values for Index of Determination from Regression Analysis

Response	R^2, %
Y_1 Disintegration time	98.08
Y_2 Hardness	98.90
Y_3 Dissolution	97.90
Y_4 Friability	75.38
Y_5 Weight	87.56
Y_6 Thickness	98.15
Y_7 Porosity	99.45
Y_8 Mean pore diameter	61.22

The responses measured on the resulting tablets were: Y_1, disintegration; Y_2, tablet hardness; Y_3, dissolution; Y_4, friability; Y_5, weight uniformity; Y_6, thickness uniformity; Y_7, tablet porosity (pore volume); and Y_8, mean pore diameter. Most of these response variables are properties of general interest to tablet formulators, but the list could be varied to fit the particular formulation.

Disintegration time was determined using the procedure and apparatus outlined in USP XVIII (9). Tablet hardness was determined using an electric hardness tester[3].

Dissolution measurements were made in 0.1 *N* HCl following the USP basket technique with a basket rotation at 100 r.p.m. The dissolution profile was followed by means of a servo-recorder[4], a spectrophotometer[5], and a flow cell.

Tablet friability was measured on a 60-tablet sample in a friabilator[6] using a 4-min. cycle (10).

Weight and thickness uniformity were calculated as the relative standard deviation of the given measurement. For tablet weight, 10 tablets were individually weighed on an analytical balance[7] and 20 tablet thicknesses were determined with a dial comparator[8].

Pore volume was determined[9], and a sample size of 10 tablets was necessary to obtain adequate readings of the mercury volume. Mean pore diameter was calculated from the porosimeter data.

RESULTS AND DISCUSSION

The responses from the 27 experiments were measured and tabulated. The large volume of data is unwieldy to report here and, for the most part, is irrelevant to the present discussion.

Statistical analysis was carried out which included the calculation of mean values for each of the eight responses in each of 27 experiments. Without repeated emphasis, it should be noted that values used for optimization must be statistically significant. The formulator must be confident that the results used as input are truly representative of the particular system.

The sets of data resulting from the statistical analysis were then subjected to computerized[10] regression analysis to determine the fit to a second-order model (Eq. 1). In most cases, the *R*-square value or the index of determination was satisfactory, as shown in Table III.

The ability of the system to predict accurately is only as good as the regression fit of the predictor equation used. The values for friability, weight, and mean pore diameter are less than desirable, and it may be that many pharmaceutical responses of interest do not follow a second-order model.

One advantage, however, of the digital type of system used is that it can be modified to accept other mathematical models, whether they be higher order polynomials, any other empirical relationships, or mathematical models based on first principles. For the present work, it was decided to restrict the model to the

[1] There are several places in the optimization procedure where the developmental pharmacist is indispensible. This is one of them, since only by formulation experience can one select reasonable ranges for those variables determined to be most important.

[2] Stokes model 580, F. J. Stokes Division of Pennwalt Corp., Warminster, Pa.

[3] Ahiba, Gubelin International Corp., Mount Kisco, NY 10549

[4] Heath model EUW-20A.

[5] Gilford model 240, Gilford Instrument Laboratories, Inc., Oberlin, Ohio.

[6] Roche.

[7] Mettler model H20T.

[8] B. C. Ames Co., Waltham, Mass.

[9] Aminco Mercury Intrusion Porosimeter, American Instrument Co., Bethesda, Md.

[10] The entire optimization system (series of programs) was written for use on a CDC 1700 (Control Data Corp., Minneapolis, Minn.).

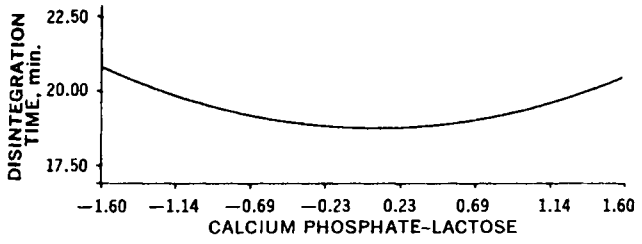


Figure 1—*Plot of disintegration time (minutes) as a function of diluent ratio generated from computer tapes and drawn by plotter. The abscissa represents the experimental range in terms of experimental units (see text).*

second-order polynomials, realizing that predictions for several variables might be inaccurate.

The remainder of the technique involves the analysis of the set of polynomial equations obtained *via* regression.

Optimization—In addition to the computerized statistical analysis and regression analysis already mentioned, programs were needed for the actual optimization. Two major steps were used: (*a*) the feasibility search, and (*b*) the grid search.

The feasibility program is used to locate a set of response constraints that is at the limit of possibilities. That is, given this particular set of restrictions, there *is* a solution (formulation), but if the constraints are tightened, there is no solution. The mechanics are to select several values for the responses of interest and slowly relax them until a possible solution is found. For example, the constraints in Table IV were fed into the computer and were relaxed one at a time until a solution was found. The first choice would have been a formulation with a disintegration time of <1 min., tablet hardness >12 kg., and dissolution >100% in 50 min.[11]. This set of constraints was expected to be unreasonable and, indeed, no solution was found until the values had been systematically relaxed to a disintegration time of 5 min., tablet hardness of 10 kg., and dissolution of 100% at 50 min. Any further relaxation of the constraints would, of course, also yield solutions.

The feasibility program is designed so that it stops after the first possibility; the formulation given may be one of many that could satisfy the constraints.

The second program, the grid search, is essentially a brute force method in which the experimental range is divided into a grid of arbitrary size and methodically searched. Based on an input of the most reasonable constraints (most likely resulting from the feasibility program), the grid search program prints out all points (formulations) in the grid that meet the criteria.

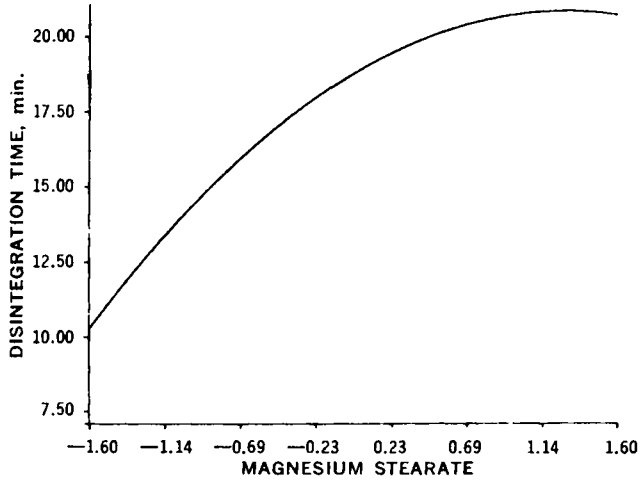


Figure 2—*Plot of disintegration time as a function of lubricant level drawn from computer tapes.*

[11] Values greater than 100%, as well as negative values, appear throughout this work. Because the second-order polynomials describe curves that may be parabolic, hyperbolic, ellipsoid, *etc.*, extrapolation may take one out of the realm of physical reality.

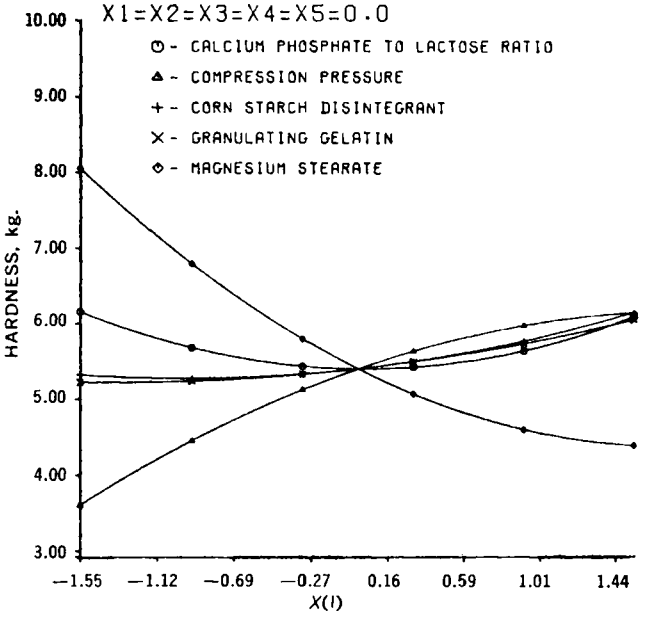


Figure 3—*Composite plot for tablet hardness as a function of each independent variable,* X_i (i = *1, 2, . . . , 5*).

The rationale for the feasibility program becomes apparent, for if one had arbitrarily selected suitable properties or constraints and made a grid search, the printout might contain hundreds of combinations that fit—or none. But, with the results of the feasibility study, it is known within reason the point at which tighter constraints yield no solutions. By running the grid search in this neighborhood, the printout will have only a few solutions from which the most suitable one can be selected.

For example, a grid search utilizing the above constraints (disintegration time ≤ 5 min., tablet hardness ≥ 10 kg., and dissolution ≥100% in 50 min.) produced eight solutions (formulations) that satisfied the requirements. A second grid search where the constraint on tablet hardness was relaxed to 8 kg. yielded 650 solutions. Of course, it is easier to choose one of eight solutions than one of 650 solutions.

In addition to providing a printout of each formulation, the grid search program also gives the corresponding values for the responses. This is the point where one can trade off one response for

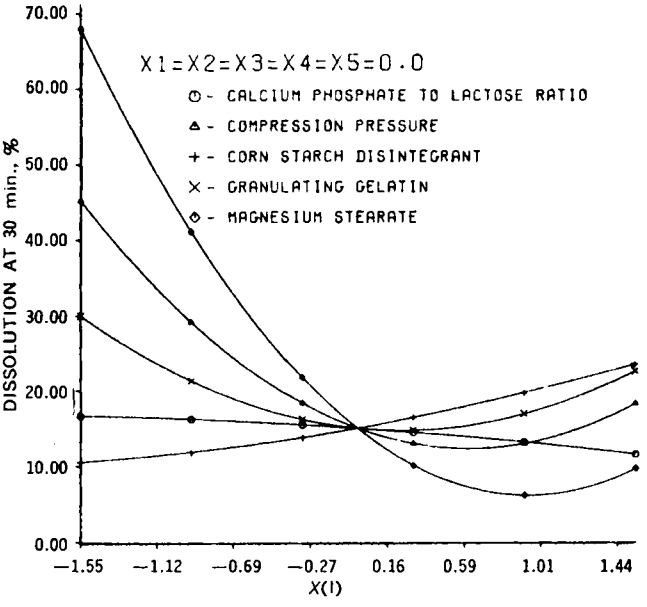


Figure 4—*Composite plot for dissolution response as a function of each independent variable,* X_i (i = *1, 2, . . . , 5*).

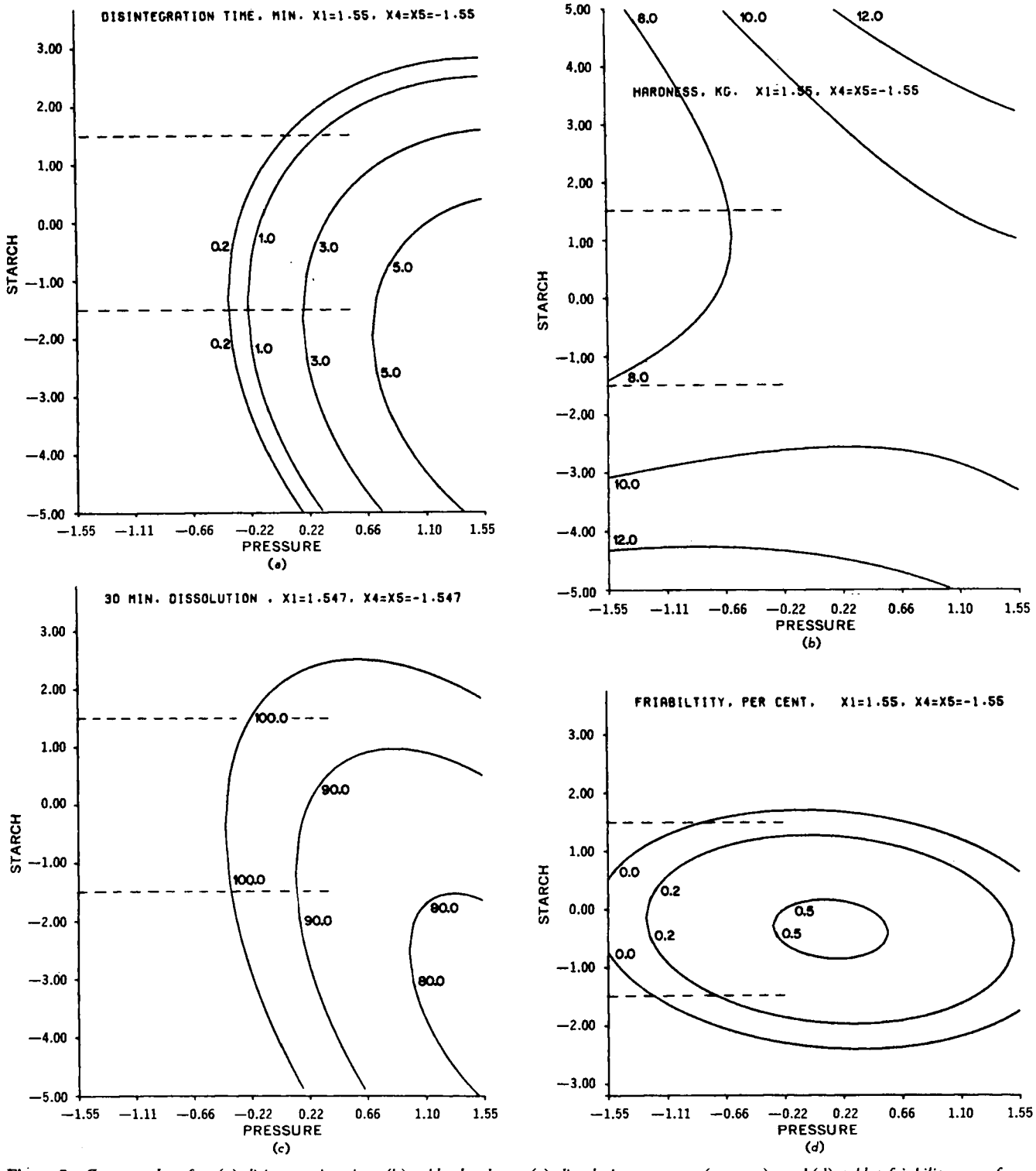


Figure 5—*Contour plots for:* (a) *disintegration time,* (b) *tablet hardness,* (c) *dissolution response (percent), and* (d) *tablet friability as a function of disintegrant level and compressional force. Dotted lines on ordinate denote limits of experimental range (−1.547 to +1.547 eu, see text).*

another or compromise as needed. Thus, the best or most acceptable formulation is selected from the grid search printout to complete the optimization.

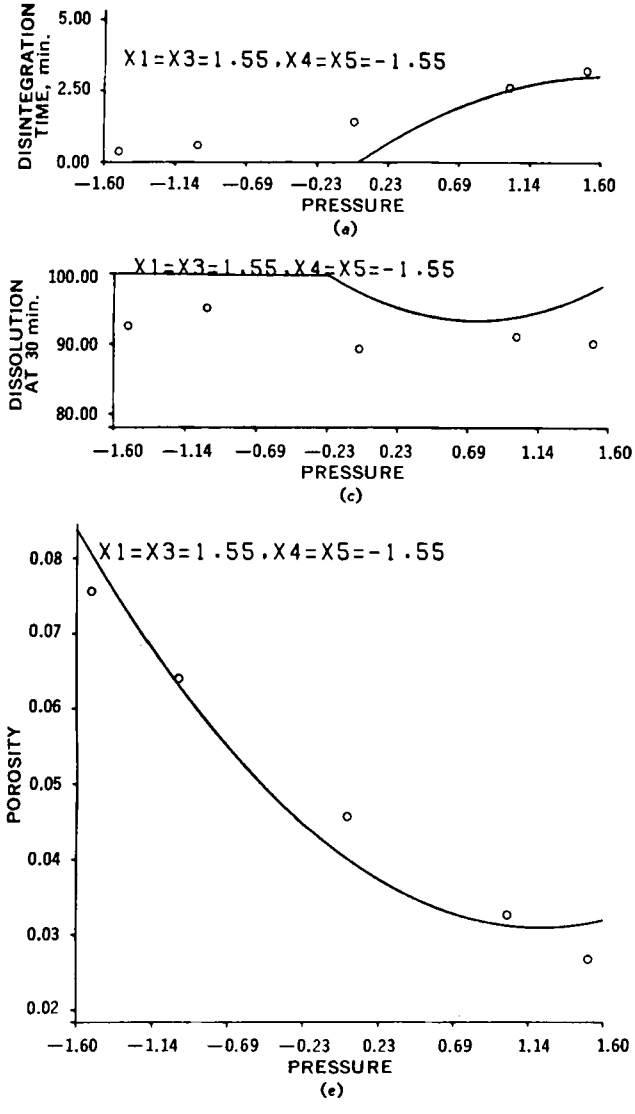
Graphical Approach—Although the computer programs discussed allow one to select the optimum formulation, certain refinements were desirable. Two graphical techniques have proven useful in the optimization procedure.

The first graphical procedure (Fig. 1) is a plot of a given response variable as a function of one of the independent variables. The responses are functions of the five independent variables and can be represented by Eq. 2:

$$Y_i = f(X_1, X_2, X_3, X_4, X_5) \qquad \text{(Eq. 2)}$$

with the full relationship given by Eq. 1. Thus, the relationship between the response and any one variable, *e.g.*, X_1, may be viewed as a partial derivative[12] of Y_i with respect to X_1 while holding all

[12] For convenience, this type of graph is called a "partial derivative" graph, although it is the slope that is actually the partial derivative.



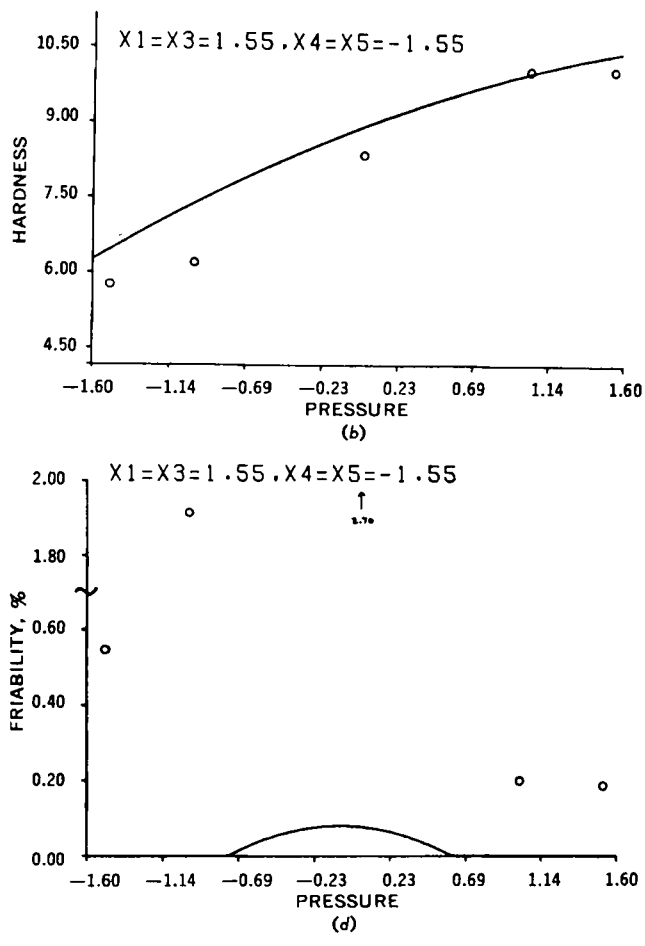


Figure 6—*Plots for several responses as a function of compressional force. The solid lines represent the computer predictions for:* (a) *disintegration time,* (b) *tablet hardness (kilograms),* (c) *dissolution response (percent),* (d) *friability, and* (e) *tablet pore volume; the circles represent experimental points.*

other X's constant. However, because Eq. 1 is a second-order equation with all possible cross-terms, the value of the partial derivatives is still a function of all the independent variables:

$$\left(\frac{\partial Y_i}{\partial X_1}\right)_{X_2X_3X_4X_5} = f(X_1,X_2,X_3,X_4,X_5) \qquad \text{(Eq. 3)}$$

For example, the full partial derivative with respect to X_1 is illustrated in Eq. 4:

$$\left(\frac{\partial Y_i}{\partial X_1}\right)_{X_2X_3X_4X_5} = a_1 + 2a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + a_{14}X_4 + a_{15}X_5 \qquad \text{(Eq. 4)}$$

Thus, the values at which X_2, X_3, X_4, and X_5 are held constant are also important since a change in value will change the graph significantly.

Figure 1 is a plot of disintegration time, Y_1, as a function of the diluent ratio, X_1. In this case, X_2, X_3, X_4, and X_5 are all being held constant at zero or the base level. Note that the X axis is represented in statistical units rather than in physical units; *i.e.*, the experimental range extends from −1.547 to +1.547 experimental units (Table II). The ordinate is represented in the physical units corresponding to the dependent variable plotted. This graph, as well as the others in this series, was drawn from computer tapes[13], so that even the graphical analysis is computerized. The same type of plot can be made for each of the other independent variables, and Fig. 2 shows disintegration time as a function of X_5, the lubricant level.

In fact, one can generate five of these "partial derivative" plots for each of the responses or tablet properties. It is obvious that the paperwork could become overwhelming, so a "composite partial derivative" plot was generated (Figs. 3 and 4). This device enables one to get an overall view of any single response such as tablet hardness. Five of the previous *single* graphs were superimposed by the computer. Each of the five curves represents one of the independent variables, and the X axis represents the experimental field. Keeping the abscissa in terms of experimental units allows this superpositioning. The values at the top of the graph represent the

Table IV—Specifications for Feasibility Search

Variable	Constraint	Experimental Range[a]
Disintegration time, min.	<1 (1)[b] 3 (2) 5 (3)	1.33–30.87
Hardness, kg.	>12 (1) 10 (2) 8 (3)	3.82–11.6
Dissolution, % at 50 min.	>100 (1) 90 (2) 80 (3)	13.3–89.1

[a] It is possible to request values for a response that are more desirable than any data obtained in the set of 27 experiments. [b] (1) = first choice.

[13] By a CalComp plotter, California Computer Corp.

Table V—Optimum Formulation

Variable	Level (in Experimental Units)
Diluent ratio	+1.547
Compressional force	+1.547
Disintegrant level	+1.547
Binder level	−1.547
Lubricant level	−1.547

Table VI—Experimental and Predicted Levels of Responses for Optimum Formulation

Response	Experimental	Predicted
Disintegration time, min.	3.31	3.03
Hardness, kg.	10.08	10.38
Dissolution, % in 30 min.	89.85	98.02
Friability, %	0.18	−0.38[a]
Thickness, mm.	2.31	2.37
Porosity (pore volume), ml./g.	0.0269	0.0318
Mean pore diameter, μ	1.1042	0.5589
Dissolution, % in 50 min.	95.48	117.85[a]
Mean granule diameter, mm.	0.170	0.292
Weight uniformity, *RSD* %	0.74	0.54
Thickness uniformity, *RSD* %	1.14	2.00

[a] See *Footnote 11.*

level at which any variable is being held constant during the partial derivative operation. In this case, all are at zero or the base level.

One could generate an infinite number of these plots because the constant values noted at the top of the graphs can cover the entire experimental range for each variable; therefore, some thought must be given to the selection of the constant levels.

The second type of plot useful in optimizations is the contour plot (Fig. 5), again drawn by computer. Essentially, this allows the representation of a three-dimensional situation in two dimensions and is of special utility when the formulator knows, *a priori*, that two of the variables are of most import.

For the contour plots in Figs. 5*a*–5*d*, the specific response is noted on the graph, as are the fixed values of the three extra variables. Both axes are represented in experimental units this time.

Optimum Formula—Based on the printout of the grid search discussed in a previous section, one of the eight formulations was selected as most suitable. It was the formulation with every variable at its experimental extreme (Table V). Diluent ratio, pressure, and disintegrant were to be at high values; binder and lubricant were to be low.

One might have been able to guess these directions, but it would likely have taken many experiments to locate the exact combination of values. The particular formulation selected as the "optimum" is unlike any one of the 27 prepared as part of the experimental plan.

The formulation was prepared and tested in the same manner as the previous ones. Table VI shows the experimental results and the predicted values for the various responses. Between the initiation of this project and this experiment, several additional responses were added.

The agreement between experimental and predicted values is extremely good, especially for the responses of primary interest: disintegration, hardness, and dissolution.

Experimental *versus* Prediction—The optimum formulation was compressed at various pressures, *i.e.*, as a function of X_2. Some results are illustrated in Figs. 6*a*–6*e*, where the solid line is the computer prediction drawn by the plotter. In general, the agreement is excellent. As noted previously, the values for friability are not quantitative although qualitatively the same convex parabolic shape is present.

SUMMARY AND CONCLUSIONS

The optimization technique presented here has enabled the formulator to select a formulation exhibiting optimum properties. In addition, the large quantity of data generated by the required experimentation and the infinite possibilities of subsequent computerized experimentation and graphical analysis permit the investigator to gain an insight into his or her process.

No previous computer experience and no familiarity with computer languages are required.

The steps necessary for the developmental pharmacist to carry out the optimization program may be summarized as follows:

1. Select a system.
2. Select variables:
   a. Independent.
   b. Dependent.
3. Perform experiments and test product.
4. Submit data for statistical and regression analysis.
5. Set specifications for feasibility program.
6. Select constraints for grid search.
7. Evaluate grid search printout.
8. Request and evaluate:
   a. "Partial derivative" plots, single or composite.
   b. Contour plots.

The last step, which concerns the graphical techniques, may be requested at any time after the regression analysis has been performed and will probably be appropriate at several different stages of a project.

The choice of variables is in the hands of the formulator and there are no restrictions as long as the property can be quantitated. The ultimate goal, of course, would be to relate the formulation variables to responses of proven biological significance, *e.g.*, bioavailability, blood levels, or even therapeutic response.

It is recognized that the stability of a formulation is an important parameter. It is possible, in this method, to include the effects of time and storage conditions by using a stability-indicating property as a response variable.

The technique presented here is not meant to, and cannot, replace the developmental pharmacist. It is simply another tool which can aid in his or her search to further define the pharmaceutical system, and because it can be defined, it can be better controlled.

REFERENCES

(1) G. E. P. Box and K. B. Wilson, *J. Roy. Statist. Soc.*, *B*, **13**, 1(1951).

(2) "The Design and Analysis of Industrial Experiments," O. L. Davies, Ed., Hafner, New York, N. Y., 1967.

(3) W. G. Cochran and G. M. Cox, "Experimental Designs," Wiley, New York, N. Y., 1957, chap. 8A.

(4) L. Cooper and D. Steinberg, "Introduction to Methods o Optimization," Saunders, Philadelphia, Pa., 1970.

(5) L. Virag, *Int. Chem. Eng.*, **10**, 513(1970).

(6) R. A. Barneson, N. F. Brannock, J. G. Moore, and C. Morris, *Chem. Eng.*, **July 27, 1970**, 132.

(7) W. E. Claxton, presented to the Industrial Pharmaceutical Technology Section, APhA Academy of Pharmaceutical Sciences, San Francisco meeting, Mar. 1971.

(8) D. E. Fonner, Jr., J. R. Buck, and G. S. Banker, *J. Pharm. Sci.*, **59**, 1587(1970).

(9) "The United States Pharmacopeia," 18th rev., Mack Publishing Co., Easton, Pa., 1970.

(10) E. G. E. Shafer, E. G. Wollish, and C. E. Engel, *J. Amer. Pharm. Ass., Sci. Ed.*, **45**, 114(1956).

ACKNOWLEDGMENTS AND ADDRESSES

Received December 18, 1972, from *Pharmaceutical Research and Development, Merck Sharp & Dohme Research Laboratories, West Point, PA 19486,* and *Management Information Systems, Merck & Co., Rahway, NJ 07065*

Accepted for publication January 31, 1973.

Presented to the Industrial Pharmaceutical Technology Section, APhA Academy of Pharmaceutical Sciences, Houston meeting, April 1972.

The authors acknowledge the statistical assistance of N. R. Bohidar and the guidance and encouragement of F. A. Restaino, J. E. Allegretti, and L. Silver.

▲ To whom inquiries should be directed.