FACTORIAL DESIGN FOR A SEQUENTIAL ALLOCATION STUDY

by

Robert Vernon Eckert

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in candidacy for the degree of

MASTER OF SCIENCE

in

Statistics

December 1961

Blacksburg, Virginia

TABLE OF CONTENTS

I. INTRODUCTION

Robert J. Taylor and Herbert A. David (1961), under the sponsorship of the National Cancer Institute, developed a new experimental approach for the screening of prospective drugs to be used in the treatment of cancer. Taylor and David introduce a sequential procedure where weighting functions are used to allocate patients to drugs, and where the weights are based upon the performance of the drugs in immediately previous periods. Taylor (I.B.M. Program) further developed a computer program to simulate the proposed procedure under varying experimental conditions. The purpose of this program is to study various criteria which characterize the effectiveness of the clinical procedure, since the nature of the procedure makes a straight-forward theoretical investigation almost impossible.

The primary aim of the investigation by Taylor and David was to improve upon the equal allocation method (Sobel and Huyett, 1957). With this goal in mind, they chose the number of patients involved in the simulated trials in such a manner as to facilitate comparisons of their results with those obtained using the conventional clinical procedure of

equal allocation. Their study showed that, for a certain range of experimental conditions, the sequential allocation procedure was always as good, and in most cases better, than the equal allocation method.

Taylor and David conclude their report by making suggestions for further research. One such suggestion is that further study of their sequential allocation procedure be carried out by using the simulation approach in a factorial experiment. It is such a factorial experiment which will be presented in this paper.

II. THE EQUAL AND SEQUENTIAL ALLOCATION PROCEDURES

A complete description of the equal and sequential allocation procedures is given by Sobel and Huyett (1957), and Taylor and David (1961) respectively. Three examples will be presented here to explain the procedures with respect to the related experimental conditions.

2.1 Equal Allocation

Equal allocation is the conventional method of selecting patients for testing drugs.

Assume that four drugs are to be tested and that they have a "least favorable configuration" (see Section 3.2.4) of expected probabilities of success, henceforth to be denoted as treatment parameters, of 0.55, 0.45, 0.45, 0.45. These drugs are to be tested in a trial consisting of a total of 288 patients. It can be seen that this trial will involve four treatments (drugs) with a difference in the treatment parameters between the best and second best of 0.10, and an average of the best and second best equal to 0.50. Under equal allocation one-fourth of the patients are assigned to each treatment. A hypothetical result of such a trial is shown in Table 1.

Table 1

A Hypothetical Trial

Equal Allocation

Treatment Number	Treatment Parameter	Number of Patients Allocated	Number of Favorable Responses	Favorable Response Ratio	Rank of Treatment
1	0.55	72	41	0.57	1
2	0.45	72	33	0.46	2
3	0.45	72	30	0.42	3
4	0.45	72	29	0.40	4

Attention is called to the fact that equivalent experimental conditions would be obtained by testing these same drugs using four different trials or periods with 18 patients per treatment. Thus the equal allocation procedure may be thought of as a trial with just one period.

It can be seen from the hypothetical outcome that treatment 1 exhibits the highest favorable response ratio, therefore it ranks first. Notice also that there are a total of 133 favorable responses, which results in a proportion of 0.46 favorable responses for the complete trial.

2.2 Sequential Allocation under Rank Weighting Function

For purposes of simplicity, assume that four drugs with

the same treatment parameters of 0.55, 0.45, 0.45, 0.45, are to be tested in a three period trial consisting of a total of 288 patients. These patients are to be allocated by means of the rank weighting function, which goes into operation after the first period. A hypothetical trial is shown in Table 2.

Table 2

A Hypothetical Trial

Rank Weighting Function

Treatment Number	Treatment Parameter	Cumulative Number of Patients Allocated	Cumulative Number of Favorable Responses	Favorable Response Ratio	Rank of Treatment
		1st Period - Equal Allocation			
1	0.55	24	16	0.67	1
2	0.45	24	11	0.46	3
3	0.45	24	13	0.54	2
4	0.45	24	9	0.38	4
		2nd Period			
1	0.55	62	34	0.55	1
2	0.45	43	19	0.44	3
3	0.45	53	26	0.49	2
4	0.45	34	13	0.38	4
		3rd Period			
1	0.55	100	52	0.52	1
2	0.45	62	27	0.44	4
3	0.45	82	38	0.46	3
4	0.45	44	21	0.48	2

The number of patients assigned per drug for the 1st period using equal allocation is, of course, equal to $\frac{1}{4} \cdot 96$ or 24 patients, as shown in Table 2.

The main interest of this example is the operational procedure of the rank weighting function. The weights of this function for the ith treatment (drug) and the jth period are defined to be,

$$W_{ij} = \frac{(T + 1) - R_{i,j-1}}{\frac{1}{2} T(T + 1)} ,$$

where

T = number of treatments.

R = rank of the ith treatment for the j - 1 period.

The number of patients for the ith treatment and the jth period is then

$$A_{ij} = W_{ij} N ,$$

where N is the total number of patients in the jth period.

For our example consider the treatment ranks of the 1st period, which determine the following allocation weights for the 2nd period.

$$W_{1,2} = \frac{(4 + 1) - 1}{\frac{1}{2}(4)(5)} ,$$

$$W_{1,2} = .40 ,$$

also

$$W_{2,2} = .20 ,$$

$$W_{3,2} = .30 ,$$

$$W_{4,2} = .10 .$$

Thus the 2nd period assignments A_{i2} based on 96 patients are,

$$A_{1,2} = (.40)(96) ,$$

$$A_{1,2} = 38 ,$$

also

$$A_{2,2} = 19 ,$$

$$A_{3,2} = 29 ,$$

$$A_{4,2} = 10 .$$

When added to the respective 1st period assignments of 24 patients the cumulative allocations are 62, 43, 53, and 34, as shown in Table 2.

The hypothetical example shows that treatment 1 has a final response ratio of 0.52, which ranks it first. The total number of favorable responses is 138, which gives a proportion of 0.48 favorable responses per trial.

2.3 Sequential Allocation under 4th Power Weighting Function

Assume, once again, that four drugs with the same treatment parameters of 0.55, 0.45, 0.45, 0.45, are to be tested

in a three period trial consisting of a total of 288 patients. The allocation is to be carried out by means of the 4th power weighting function, which also goes into operation after the first period. This example differs from Example 2.2 by the choice of the weighting function only.

Table 3

A Hypothetical Trial

4th Power Weighting Function

Treatment Number	Treatment Parameter	Cumulative Number of Patients Allocated	Cumulative Number of Favorable Responses	Favorable Response Ratio	Rank of Treatment
		1st Period - Equal Allocation			
1	0.55	24	13	0.54	1.5
2	0.45	24	11	0.46	2
3	0.45	24	13	0.54	1.5
4	0.45	24	10	0.42	3
		2nd Period			
1	0.55	53	27	0.51	2
2	0.45	45	18	0.40	4
3	0.45	53	23	0.43	3
4	0.45	41	22	0.54	1
		3rd Period			
1	0.55	81	44	0.54	1
2	0.45	63	26	0.41	4
3	0.45	72	31	0.43	3
4	0.45	72	34	0.47	2

The main interest of this example is the operational procedure of the 4th power weighting function, which differs from the rank weighting function in the method of allocating patients. The weights for the ith treatment (drug) and the jth period of the pth power weighting function are defined to be,

$$w_{ij}^{(p)} = \frac{w_{ij}^{p}}{\sum_{s=1}^{t} w_{sj}^{p}} .$$

The w_{ij} are the 1st power weights and they are defined as follows,

$$w_{i,j+1} = \frac{1}{t}(1 + d_{ij}) ,$$

where

t = number of treatments,

$d_{ij} = r_{ij} - \bar{r}_{\cdot j}$, the response rate deviations,

where

r_{ij} = <u>cumulative</u> response rate.

$r_{\cdot j} = \sum_{i=1}^{t} \frac{r_{ij}}{t}$ = average <u>cumulative</u> response rate.

For our example consider the favorable response ratios for the 1st period, which determine the following allocations

for the 2nd period.

$$\bar{r}_{\cdot 2} = \frac{(.54 + .46 + .54 + .42)}{4} ,$$

$$\bar{r}_{\cdot 2} = .49 .$$

Hence

$$d_{1,2} = (.54 - .49) ,$$

$$d_{1,2} = .05 ,$$

also

$$d_{2,2} = - .03 ,$$

$$d_{3,2} = .05 ,$$

$$d_{4,2} = - .07 .$$

The 1st power weights then become,

$$w_{1,2} = \frac{1}{4}(1 + .05) ,$$

$$w_{1,2} = .2625 ,$$

also

$$w_{2,2} = .2425 ,$$

$$w_{3,2} = .2625 ,$$

$$w_{4,2} = .2325 .$$

Therefore the 4th power weights are,

$$w_{1,2}^{(4)} = \frac{(.2625)^4}{[(.2625)^4+(.2425)^4+(.2625)^4+(.2325)^4]} ,$$

$$w_{1,2}^{(4)} = .30 ,$$

also

$$w_{2,2}^{(4)} = .22 ,$$

$$w_{3,2}^{(4)} = .30 ,$$

$$w_{4,2}^{(4)} = .18 .$$

The 2nd period assignments are,

$$A_{1,2} = (.30)(96) ,$$

$$A_{1,2} = 29 ,$$

also

$$A_{2,2} = 21 ,$$

$$A_{3,2} = 29 ,$$

$$A_{4,2} = 17 .$$

When added to the respective 1st period assignments of 24 patients the cumulative allocations are 53, 45, 53, and 41, as shown in Table 3.

It can be seen in the hypothetical example that treatment 1 has a final response ratio of 0.54, which ranks first.

The total number of favorable responses turns out to be 135, which gives a proportion of 0.47 favorable responses per trial.

In general, nth power weighting functions can be generated by letting p take on any value. Taylor and David (1961) discuss the possible advantages and disadvantages of using the higher versus the lower values of p .

III. THE EXPERIMENTAL CONDITIONS

The purpose of this paper is to study the effectiveness of different weighting functions in the sequential allocation procedure as outlined in Chapter II. Four response variables will be considered in order to evaluate the operational procedure of these functions under varying experimental conditions. Further elaborations on the practical significance of these variables and factors is given by Taylor and David (1961).

3.1 Response Variables

The response variables under study are,

(i) Relative frequency of correct selection (that is, the relative frequency with which the best treatment ranks first).

(ii) Proportion of patients allocated to the best treatment.

(iii) Proportion of (total) favorable responses per trial.

(iv) Rank of the best treatment.

3.1.1 Relative Frequency of Correct Selection

Of the four variables under consideration, the relative frequency of correct selection of the drug with the highest

treatment parameter is of primary concern. It can be seen that an increase in the ability to correctly select a treatment (drug), when it is in fact the best treatment, is of major importance. It is noteworthy to mention that the improvement of the relative frequency of correct selection was the basis for the development of the sequential allocation procedure.

3.1.2 Proportion of Patients Allocated to the Best Treatment

The proportion of patients allocated to the best treatment is of importance due to the natural desire to cure people. It is evident that the more people assigned to the best treatment, the greater will be the proportion of successes.

3.1.3 Proportion of Favorable Responses per Trial

The proportion of favorable responses per trial is the only variable which takes all the treatments into consideration. This proportion is somewhat analogous to the previous variable and it is obviously of importance in any type of clinical work. It can be seen that any improvement on this ratio would be of interest under practical experimental conditions.

3.1.4 Rank of the Best Treatment

The rank of the best treatment is also of interest in analyzing the results of clinical trials which have been carried out under practical experimental conditions. This variable is highly related to the relative frequency of correct selection, and when taken together they provide excellent information about the treatment which has given the best performance.

3.2 Design Factors

The four response variables will be studied by means of factorial designs with the following experimental conditions or factors,

Allocation

(i) Equal

(ii) Sequential, Rank, 4th, 8th, 16th power weighting function

Number of treatments, 4, 6, 8

Number of periods, 3, 6

Treatment parameters (probabilities of success)

(i) Difference between best and second best treatments, 0.10, 0.20

(ii) Average of best and second best treatments, 0.50, 0.35

These factors are used in two separate factorial designs (see Section 3.3).

3.2.1 Allocation

The two different methods of allocation have been discussed in Chapter II. The sequential allocation procedure is of primary interest, whereas the equal allocation method is introduced simply as a basis for comparison.

3.2.2 Number of Treatments

It is important to mention that the number of treatments considered, in practical clinical trials, is generally controlled by an outside source. The values used in this paper have been chosen in agreement with the number of treatments most frequently investigated in actual practice.

3.2.3 Number of Periods

The selection of the number of periods is optional. The values used in this work have been chosen in an attempt to gain information concerning the most efficient selection.

3.2.4 Treatment Parameters (probabilities of success)

It is assumed that each treatment is associated with a theoretical probability of success. These proportions have been chosen (Sobel and Huyett, 1957) in agreement with the

theory of a "least favorable configuration".

It can be seen that the choice of treatment parameters is almost unlimited. It is for this reason that the theory of a "least favorable configuration" was introduced, in order to help select the most advantageous combination. In brief, this concept is defined to be the configuration which gives the lowest probability of correct selection. It has been shown (Sobel and Huyett, 1957) that the configuration consisting of one best treatment, with all other treatments being equally inferior, gives the lowest probability of correct selection when, for an experiment with k treatments

(i) $p_1 = \frac{1}{2}(1 + d)$,

(ii) $p_i = (p_1 - d)$, $\quad i = 1, 2, \ldots, k$,

(iii) $\frac{p_1 + p_i}{2} = 0.50$,

(iv) $n \to \infty$,

where

p_1 = treatment parameter (that is, probability of success) for the best treatment

p_i = treatment parameter for the other treatments

d = difference between best and other treatments

n = number of replications per experiment.

The following treatment parameters have been selected for study under the "least favorable configuration" theory.

(i) 0.55, 0.45, $\sim$ 0.45, and

(ii) 0.60, 0.40, $\sim$ 0.40.

A 0.10 difference between the best and second best treatments can be seen in parameter combination (i), while a 0.20 difference is evident in (ii). The combinations both have a 0.50 average of the best and second best treatments.

Other treatment parameter configurations to be studied are,

(iii) 0.40, 0.30, $\sim$ 0.30, and

(iv) 0.45, 0.25, $\sim$ 0.25.

A 0.10 difference between the best and second best treatments appears in combination (iii) while a 0.20 difference is shown in (iv). However, both configurations have a 0.35 average of the best and second best treatments.

3.3 Factorial Designs

3.3.1 Design 1 - Equal Allocation

The purpose of this equal allocation design is to supply information, on the response variables, which can be used as a basis for comparison in the discussion of the

sequential allocation results.

The design used is as follows:

Factor	Level
Number of treatments	4, 6, 8
Difference between best and second best treatments	0.10, 0.20
Average of best and second best treatments	0.50, 0.35

The number of patients assigned to each trial and the number of trials per experiment will both be held fixed at 288 and 400 respectively. The purpose of the 400 replications is to provide a true error term for carrying out the necessary tests of hypotheses which have been postulated for each of the factorials. The response output will consist of $3 \times 2 \times 2 = 12$ different experiments.

3.2.2 Design 2 - Sequential Allocation

The purpose of this sequential allocation design is to study the relative effectiveness of the four weighting functions under varying experimental conditions.

The design used is as follows:

Factor	Level
Weighting function	Rank, 4th, 8th, 16th
Number of treatments	4, 6, 8
Number of periods	3, 6
Difference between best and second best treatments	0.10, 0.20
Average of best and second best treatments	0.50, 0.35

The number of patients assigned to each trial and the number of trials per experiment will both be held fixed at 288 and 100 respectively. The response output will consist of 4 x 3 x 2 x 2 x 2 = 96 different experiments.

IV. THE ANALYSIS

The order of procedure used to carry out and analyze this particular experiment is as follows,

(i) generation of data,

(ii) test for homoscedasticity,

(iii) analysis of variance.

4.1 Generation of Data

The combined response output consisting of 108 experiments was obtained on the IBM 650 Computer by the use of Taylor's simulated trials program. Care was taken not to destroy the random number sequence, and no experiments were deleted or altered. The simulation procedure took approximately 250 hours of actual computer time.

Summaries of the 108 experiments were obtained on the IBM 650 Computer by means of a frequency distribution program. Included in these summaries were the means and variances of variables 1, 2, and 4 (see Section 3.1), as well as the distributions of the ranks, patient allocations, favorable responses, and favorable response ratios for each individual treatment in the analysis.

Additional summary work was necessary in order to obtain the means and respective variances for response variable 3, the proportion of favorable responses per trial. This variable is the only one which involves all the treatments collectively, and for this reason it was not fully reduced by the original summarizing program.

4.2 Test for Homoscedasticity

One of the assumptions of a factorial analysis is that of a homogeneous variance. A separate investigation of each of the four response variables was begun, under both Designs 1 and 2, in order to determine if the assumption of a homogeneous variance would be valid. When this condition did not hold, the appropriate variance stabilizing transformations were carried out before continuing with the analysis.

4.2.1 Relative Frequency of Correct Selection

The relative frequency of correct selection is, by definition, a binomial type variable, therefore an arc-sine transformation (O. Kempthorne, 1952) was performed on these observations.

The investigation of the remaining three variables was begun by means of graphical work, and these results were in turn supplemented by regression analyses relating means to variances. The idea behind this particular form of investigation was to

(i) estimate the variability of the variances, and

(ii) establish some relationship between the mean and variance.

4.2.2 Proportion of Patients Allocated to the Best Treatment

Of the four response variables under investigation, the proportion of patients allocated to the best treatment furnished the most interesting results. The graphical work performed on this variable showed that the variances were quite heterogeneous among experiments from the four different weighting functions. Within each function, there was an indication of a mean-variance relationship somewhat similar to that found in the binomial distribution, but these variances were not sufficiently heterogeneous to justify a variance stabilizing transformation.

The higher power weighting functions produced much larger and more variable variances. This phenomenon suggested that a factorial analysis be performed separately on the data for each of the four functions. Such an investigation

would obviously have very little practical significance in comparison with a collective study of the weighting functions, and for this reason the usual combined analysis was carried out. The heterogeneity of the variances should, however, be kept in mind when interpreting the results.

4.2.3 Proportion of Favorable Responses per Trial

The graphical work as well as the regression analysis performed on this variable showed that the variances were relatively homogeneous and unrelated to the means, therefore no stabilizing transformation was necessary.

4.2.4 Rank of the Best Treatment

Both the graphical work and the regression analysis performed on this variable established the existence of a linear relationship between the mean and variance. In order to eliminate the effect of this relationship, a square root transformation (O. Kempthorne, 1952) was applied to the original observations of this variable.

4.3 Analysis of Variance

When the condition of a homogeneous variance had been fulfilled for each of the four response variables, eight distinct factorial analyses were performed on the IBM 650

Computer by means of a standard factorial analysis program. The usual additive model for a factorial experiment, with all interactions present, was assumed to hold. All effects were regarded as fixed (Schultz, 1955), therefore all statistical tests for significance were carried out against the error (within cell) mean square. This error was assumed to be normally and independently distributed with mean zero and variance σ_e^2 .

For an interpretation of the practical significance of the analysis of variance, means corresponding to significant sources of variation were studied by the use of tables and graphs.

V. RESULTS OF THE ANALYSIS

In order to simplify the study of the factorial analyses, the following notation, for the factors in this experiment, will be adopted for tabular and graphical presentations and used throughout the discussion of the four response variables.

Wtfn = weighting function

Treat = number of treatments

Per = number of periods

Diff = difference between best and second best treatments

Aver = average of best and second best treatments

Further simplification has been carried out by means of pooling all interactions of order greater than two. This pooling is justified by the fact that not one of these interactions turned out to be statistically significant for any variable under Design 1 or Design 2. Also, F. ratios of value less than 1 will not be calculated.

All analysis of variance tables are based on transformed values (where applicable), while all tables of means are based on original observations. It is worthwhile to mention that some of the analysis of variance results turn out to be statistically significant under conditions which have no

practical significance. Examples of this situation will be presented and discussed.

5.1 Relative Frequency of Correct Selection

Recall that of the four response variables under study, the relative frequency of correct selection is the most important one. Conclusions for this variable are somewhat conservative due to the fact that all first place ties involving the best treatment were considered as incorrect choices, and hence they did not contribute to the percentage of correct selection for any experiment. This omission was performed mainly to simplify computations, as well as do away with the problem of how to credit ties with respect to the percentage of correct selection. There were, however, only a very small number of such ties.

The relative frequency of correct selection has no within-cell mean square to be used as an error term, therefore the higher order interaction is used in both Designs 1 and 2. An approximate error, for the transformed data, of 0.25 (O. Kempthorne, 1952) yields results comparable to those shown for this variable.

5.1.1 Design 1

The analysis of variance is shown in Table 4.

Table 4

Factor	df	ms	F • Ratio
Treat	2	81.9404	803.34**
Diff	1	324.0436	3,176.90**
Aver	1	2.1224	20.81**
Treat • Diff	2	1.3916	13.64
Treat • Aver	2	.3232	3.17
Diff • Aver	1	.2640	2.59
Higher order interaction	2	.1020	

The statistically significant ratios in the analysis of variance table are (in order of significance),

(i) the main effect due to the difference between the best and second best treatments.

(ii) the main effect due to the number of treatments.

(iii) the main effect due to the average of the best and second best treatments.

The practical significance of the two largest F • ratios can be seen by an examination of Table 5, which shows the average percentages of correct selection for all levels of

these factors.

Table 5

Average Percentage of Correct Selection

Diff \ Treat	4	6	8	Average
.10	75.6	55.6	40.6	57.3
.20	97.0	92.2	77.6	88.9
Average	86.3	73.9	59.1	73.1

It can readily be seen from Table 5 that the 0.20 level of difference between the best and second best treatments produces a percentage which is over 30 points higher than the value obtained for the 0.10 level. This result is noteworthy, but not surprising, due to the fact that a higher percentage of correct selection is to be expected when the difference between the best and second best treatments is doubled.

Table 5 also shows that the main effect due to the number of treatments is such that the percentage of correct selection goes up as the number of treatments goes down. This result is reasonable since there are fewer possibilities of incorrect selection when the number of treatments is the lowest.

The investigation of the means for the main effect due

to the average of the best and second best treatments showed that even though this particular source of variation is statistically significant (see Table 4), it was found that the difference between the means is too small to contribute anything to the results from a practical point of view.

Consider the second order interaction due to the number of treatments, and the difference between the best and second best treatments. This interaction was close enough to statistical significance to justify an investigation of means along with the three significant ratios. This investigation turned out to be worthwhile, because it showed that this particular interaction possessed some practical significance, as shown in Figure 1 (also in Table 5). It turns out that the variation in the percentages of correct selection, due to the three levels of the number of treatments, is not consistent for the two levels of difference between the best and second best treatments.

It is interesting to notice that the percentages are all higher under a difference of 0.20, regardless of which treatment level is taken into consideration. However, the increase in the correct selection becomes greater as the number of treatments is increased.

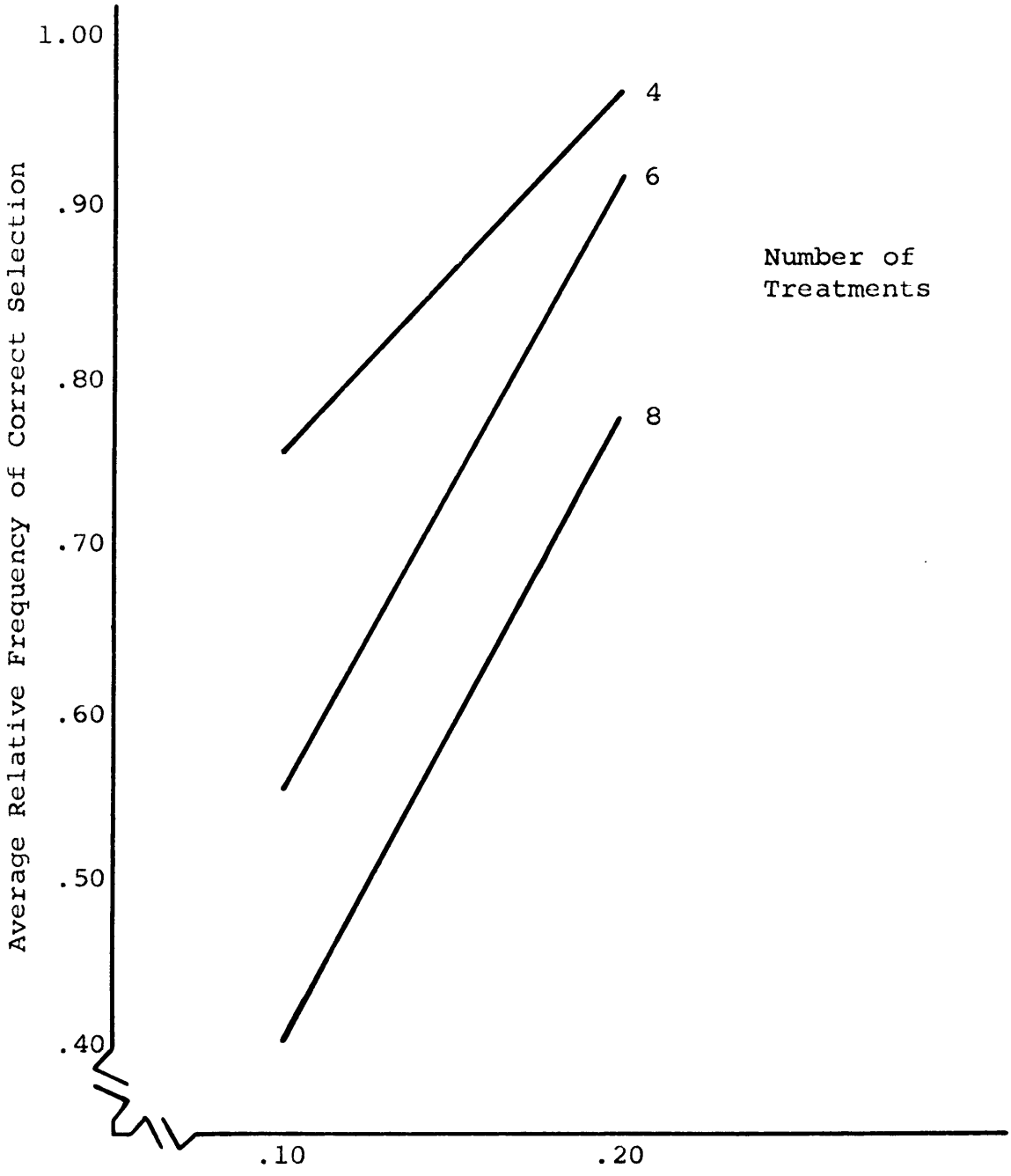

Figure 1

Treat · Diff Interaction (Design 1)

5.1.2 Design 2

The analysis of variance is shown in Table 6.

Table 6

Factor	df	ms	F · ratio
Wtfn	3	.3437	
Treat	2	115.9827	215.46**
Per	1	.3408	
Diff	1	720.5454	1,338.56**
Aver	1	.7814	1.45
Wtfn · Treat	6	.1953	
Wtfn · Per	3	.4325	
Wtfn · Diff	3	.1096	
Wtfn · Aver	3	.1446	
Treat · Per	2	.5649	1.05
Treat · Diff	2	2.1146	3.93*
Treat · Aver	2	.0340	
Per · Diff	1	.0019	
Per · Aver	1	.4420	
Diff · Aver	1	.4079	
Higher order interactions	63	.5383	

The statistically significant ratios in the analysis of variance table are (in order of significance),

(i) the main effect due to the difference between the best and second best treatments.

(ii) the main effect due to the number of treatments.

(iii) the interaction between the number of treatments, and the difference between the best and second best treatments.

Table 7 is presented to facilitate a practical interpretation of the significant F · ratios.

Table 7

Average Percentage of Correct Selection

Diff \ Treat	4	6	8	Average
.10	80.4	64.7	48.5	64.5
.20	98.6	95.1	89.0	94.2
Average	89.5	79.9	68.7	79.4

The conclusions based on Design 2 are almost identical to those of Design 1.

The most interesting aspect of the relative frequency of correct selection is the fact that neither the weighting functions nor the number of periods came out to be statistically significant.

Table 8 points out the fact that there is no obvious trend in the percentages among the weighting functions.

Table 8

Average Percentage of Correct Selection

Wtfn	
Rank	79.4
4th	78.2
8th	80.1
16th	79.8

However, a comparison of Tables 5 and 7 shows that all four functions are better than equal allocation.

It was found that three periods gives a percentage of correct selection of 79.0 as compared with a value of 79.7 for six periods. As with the weighting functions, this variation is much too small to permit any statement concerning the preference of either number.

5.2 Proportion of Patients Allocated to the Best Treatment

5.2.1 Design 1

In the equal allocation method this variable is always

equivalent to $\frac{1}{\text{number of treatments}}$, since each treatment is assigned the same number of patients. Thus an analysis of variance for this proportion is meaningless.

5.2.2 Design 2

The test of homoscedasticity for this variable (see Section 4.2.2) indicated that the variances were quite heterogeneous among the weighting functions. Due to practical considerations, the collective analysis was nevertheless performed, but the heterogeneity of the variances should be kept in mind in the interpretation of the results.

The analysis of variance is shown in Table 9.

Table 9

Factor	df	ms	F • ratio
Wtfn	3	1,376,733.3	1,565.00**
Treat	2	2,268,080.0	2,578.24**
Per	1	184,250.0	209.45**
Diff	1	1,687,280.0	1,918.01**
Aver	1	1,790.0	2.03
Wtfn • Treat	6	11,315.0	12.86**
Wtfn • Per	3	17,776.7	20.21**
Wtfn • Diff	3	228,776.7	260.06**
Wtfn • Aver	3	1,566.7	1.78
Treat • Per	2	825.0	
Treat • Diff	2	210.0	
Treat • Aver	2	35.0	
Per • Diff	1	13,660.0	15.53**
Per • Aver	1	5,540.0	6.30*
Diff • Aver	1	10.0	
Higher order interactions	63	847.6	
Error	9,504	879.7	

The statistically significant ratios in the analysis of variance table are (in order of significance),

(i) the main effect due to the number of treatments.

(ii) the main effect due to the difference between the best and second best treatments.

(iii) the main effect due to the weighting function.

(iv) the interaction between the weighting function and the difference between the best and second best treatments.

(v) the main effect due to the number of periods.

Other effects in the analysis of variance (see Table 9) are also statistically significant, but an investigation of the corresponding means has shown that they are of little or no practical importance.

Consider Table 10 for a practical interpretation of two of the significant F · ratios.

Table 10

Average Proportion of Patients
Allocated to the Best Treatment

Wt Fn / Treat	Rank	4th	8th	16th	Average	Expected
4	.332	.339	.430	.536	.409	.250
6	.226	.238	.305	.410	.295	.167
8	.169	.177	.237	.321	.226	.125
Average	.243	.251	.324	.422	.310	

The main effect due to the number of treatments can be seen in Table 10. The proportion of patients allocated to the best treatment is obviously higher when there are fewer treatments, and lower when there are more treatments. Note, however, that these proportions are all higher than the equal allocation expected values, which are also shown in the table.

The main effect due to the weighting function is also shown in Table 10. It is evident that the proportion of allocation is highest under the 16th power weighting function. The distinguishing characteristic between the different weighting functions is: the higher the power, the more patients are expected to be allocated to the best treatment. This result is therefore not surprising. It is at this point that the heterogeneity of the variances must be considered. The higher power weighting functions allocate more patients to the best treatment on the <u>average</u>, but this proportion will vary widely between individual experiments.

The interaction between the weighting function and the number of treatments, shown in Figure 2 (also in Table 10), is one of those effects which is statistically significant

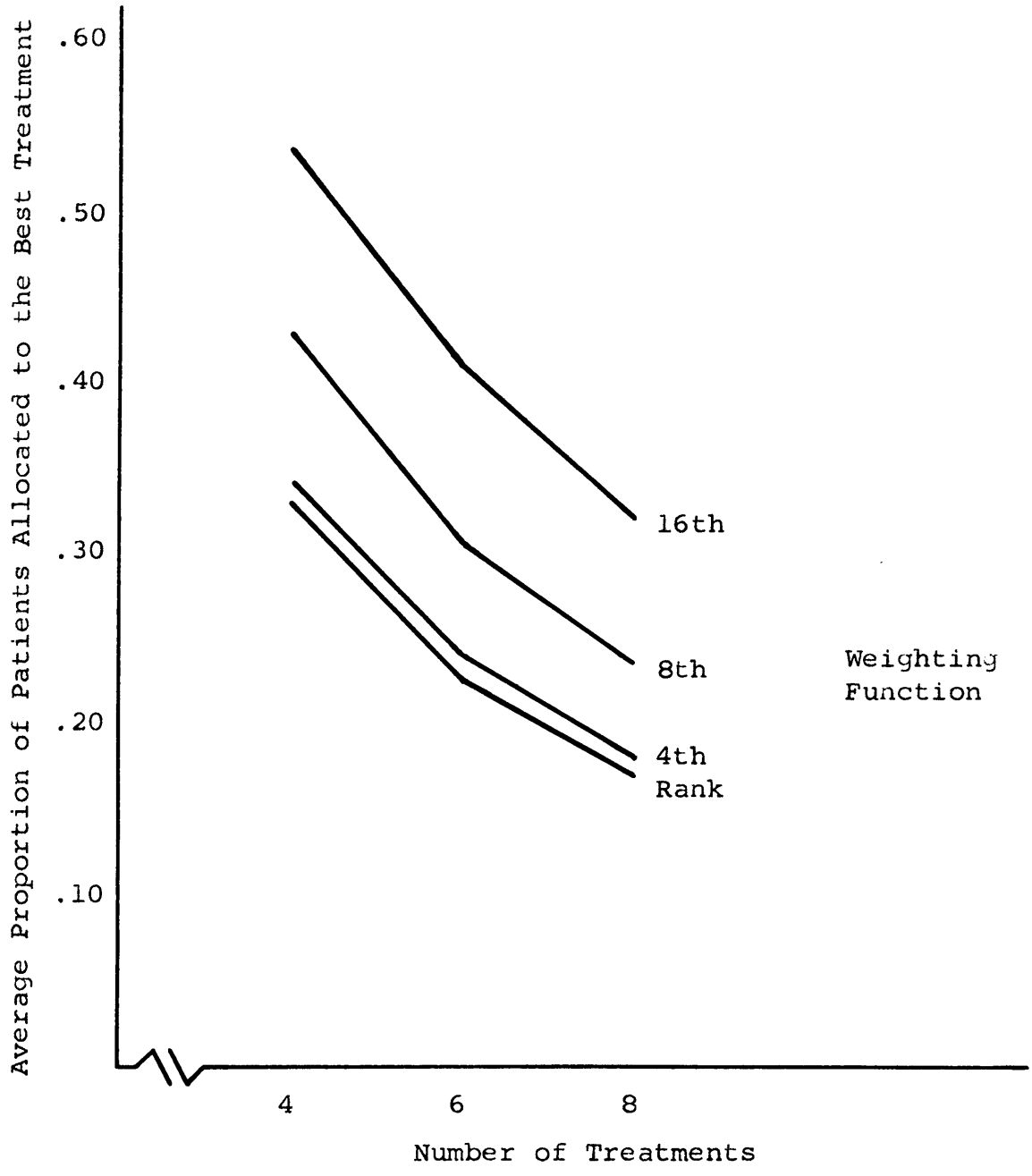

Figure 2

Wtfn · Treat Interaction (Design 2)

(see Table 9), but not practically significant. The existence of a parallel trend among the lines connecting the number of treatments, is apparent for each of the four weighting functions. If this particular interaction were of any practical value, the deviation from parallelism would be much more evident.

The main effect due to the difference between the best and second best treatments is shown in Table 11.

Table 11

Average Proportion of Patients Allocated to the Best Treatment

Diff \ Wtfn	Rank	4th	8th	16th	Average
.10	.229	.226	.268	.332	.264
.20	.256	.276	.379	.512	.356
Average	.243	.251	.324	.422	.310

It can be seen that the 0.20 difference level produced a proportion of allocation which is much greater than that obtained for the 0.10 level. Previous work on the relative frequency of correct selection (see Table 7) has already

shown that the higher percentage of correct selection occurs for a 0.20 level, therefore this result is to be expected.

The interaction between the weighting function, and the difference between the best and second best treatments is shown in Figure 3 (also in Table 11). It is evident that the higher power weighting functions become much more "active" for the 0.20 difference. In other words, the larger the difference, the greater the proportion of allocation to the best treatment under the higher power functions. This result is to be expected from the considerations used in the formulation of the sequential allocation procedure.

The number of periods produced a variation which is statistically significant (see Table 9) for the proportion of patients allocated to the best treatment. Three periods showed a proportion of 0.295 in comparison with a value of 0.325 for six periods. Previous work on the relative frequency of correct selection (see Section 5.1.2) has already shown that a slightly higher percentage of correct selection occurs for the six period level, therefore the highest proportion of allocation is to be expected at the same level.

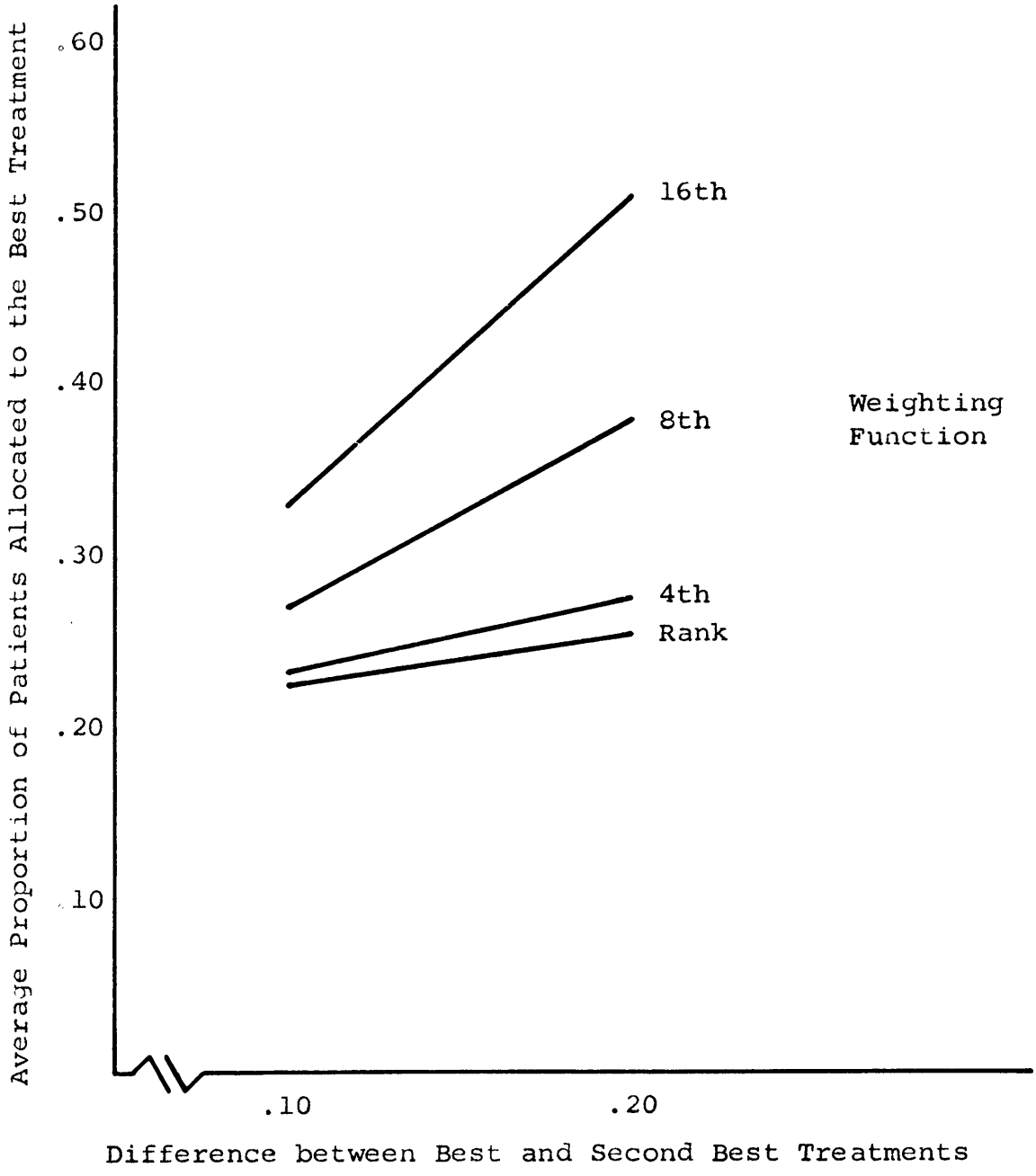

Figure 3

Wtfn · Diff Interaction (Design 2)

5.3 Proportion of Favorable Responses per Trial

5.3.1 Design 1

The analysis of variance is shown in Table 12.

Table 12

Factor	df	ms	F · ratio
Treat	2	12,706.00	539.08**
Diff	1	95,568.80	4,054.68**
Aver	1	2,286,998.00	97,030.04**
Treat · Diff	2	1,259.20	53.42**
Treat · Aver	2	71.20	3.02
Diff · Aver	1	4.80	
Higher order interaction	2	29.60	1.26
Error	4,788	23.57	

The statistically significant ratios in the analysis of variance table are (in order of significance),

(i) the main effect due to the average of the best and second best treatments.

(ii) the main effect due to the difference between the best and second best treatments.

(iii) the main effect due to the number of treatments.

(iv) the interaction between the number of treatments and the difference between the best and second best treatments.

The average of the best and second best treatments produced a variation which is statistically significant (see Table 12). The 0.35 difference level showed a 0.301 proportion of favorable responses in comparison with a value of 0.452 for the 0.50 level. It is evident that the proportion of favorable responses is "controlled" by the overall average of each treatment parameter combination (see Section 3.2.4), therefore it is reasonable to expect the result obtained.

Consider Table 13 for a practical interpretation of the remaining three significant F · ratios.

Table 13

Average Proportion of Favorable Responses per Trial

Diff \ Treat	4	6	8	Average
.10	.399	.391	.386	.392
.20	.375	.359	.350	.361
Average	.387	.375	.368	.377

The main effect due to the difference between the best and second best treatments is shown in Table 13. The highest proportion of favorable responses occurs for the 0.10 difference and this result is reasonable since the expected value combinations (see Section 3.2.4) for this level have a higher overall average than those for the 0.20 difference level.

The main effect due to the number of treatments can also be seen in Table 13. The highest proportion of favorable responses occurs for the lowest number of treatments. It is evident that the overall average of the treatment parameters goes up as the number of treatments goes down (see Section 3.2.4), therefore the higher proportion is to be expected for the fewer number of treatments.

Another example of a statistically significant source of variation with little practical significance is shown in Figure 4 (also in Table 13). The relatively parallel trend, which is the distinguishing characteristic of no interaction, is evident between the lines connecting the differences, therefore this effect is of little practical importance.

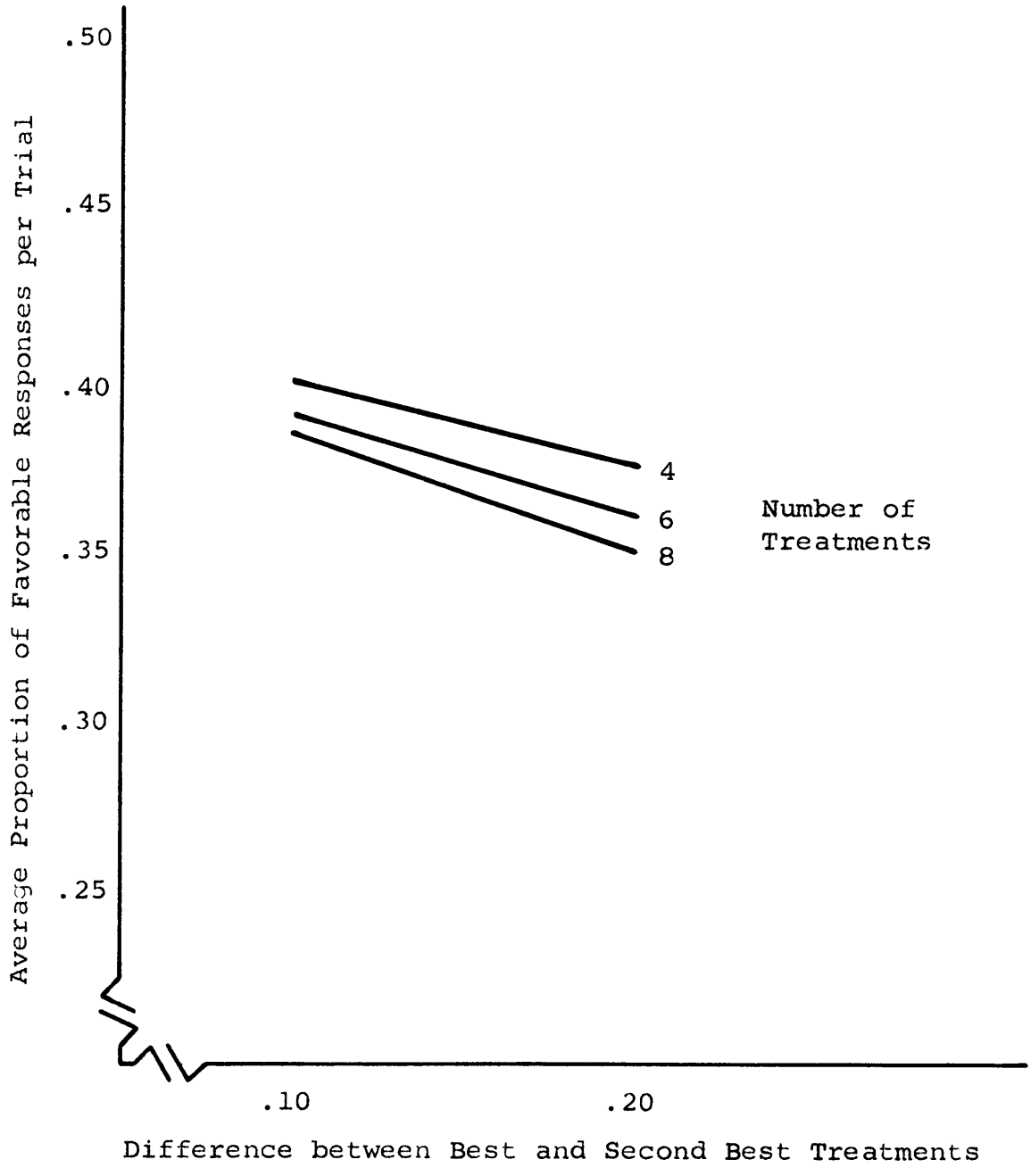

Figure 4

Treat · Diff Interaction (Design 1)

5.3.2 Design 2

The analysis of variance is shown in Table 14.

Table 14

Factor	df	ms	F · ratio
Wtfn	3	42,861.00	131.97**
Treat	2	59,625.00	183.59**
Per	1	5,370.00	16.53**
Diff	1	4,740.00	14.59**
Aver	1	4,452,921.00	13,710.58**
Wtfn · Treat	6	309.67	
Wtfn · Per	3	391.33	1.20
Wtfn · Diff	3	14,938.33	46.00**
Wtfn · Aver	3	63.33	
Treat · Per	2	75.00	
Treat · Diff	2	6,033.50	18.58**
Treat · Aver	2	145.50	
Per · Diff	1	2,110.00	6.50*
Per · Aver	1	164.00	
Diff · Aver	1	101.00	
Higher order interactions	63	86.02	
Error	9,504	324.78	

The statistically significant ratios in the analysis of variance table are (in order of significance),

(i) the main effect due to the average of the best and second best treatments.

(ii) the main effect due to the number of treatments.

(iii) the main effect due to the weighting function.

(iv) the interaction between the weighting function and the difference between the best and second best treatments.

There are other statistically significant sources of variation (see Table 14) in the analysis of variance, but an investigation of the means has shown that these effects are of little or no practical importance.

The conclusions for the main effect due to the average of the best and second best treatments, and the main effect due to the number of treatments are identical with those presented under Design 1 for this variable.

Consider Table 15 for a practical interpretation of the remaining two significant F · ratios.

Table 15

Average Proportion of Favorable Responses per Trial

Diff \ Wtfn	Rank	4th	8th	16th	Average	Equal Allocation
.10	.397	.396	.402	.409	.401	.392
.20	.376	.380	.401	.427	.396	.361
Average	.387	.388	.401	.418	.399	

The main effect due to the weighting function is shown in Table 15. The highest proportion of favorable responses occurs for the higher power weighting functions. Previous work on the proportion of patients allocated to the best treatment (see Table 10) has already shown that the highest proportion of allocation occurs for the higher power functions also, therefore the total proportion of favorable responses is expected to increase.

The main effect due to the difference between the best and second best treatments is also shown in Table 15. This effect is interesting since the difference noticed in Design 1 is not apparent for Design 2.

The interaction between the weighting function and the difference between the best and second best treatments is

illustrated in Figure 5 (also in Table 15). An examination of this graph shows that the advantages of the higher power weighting functions increase for the 0.20 difference between the best and second best treatments. Actually a decrease in the proportion of favorable responses is to be expected in passing from the 0.10 difference level to the 0.20 level, since the overall average of the treatment parameters decreases in the same order.

5.4 Rank of the Best Treatment

5.4.1 Design 1

The analysis of variance is shown in Table 16.

Table 16

Factor	df	ms	F · ratio
Treat	2	21.2037	240.40**
Diff	1	62.0848	703.91**
Aver	1	.3224	3.66
Treat · Diff	2	5.5790	63.25**
Treat · Aver	2	.0772	
Diff · Aver	1	.0608	
Higher order interaction	2	.0001	
Error	4,788	.0882	

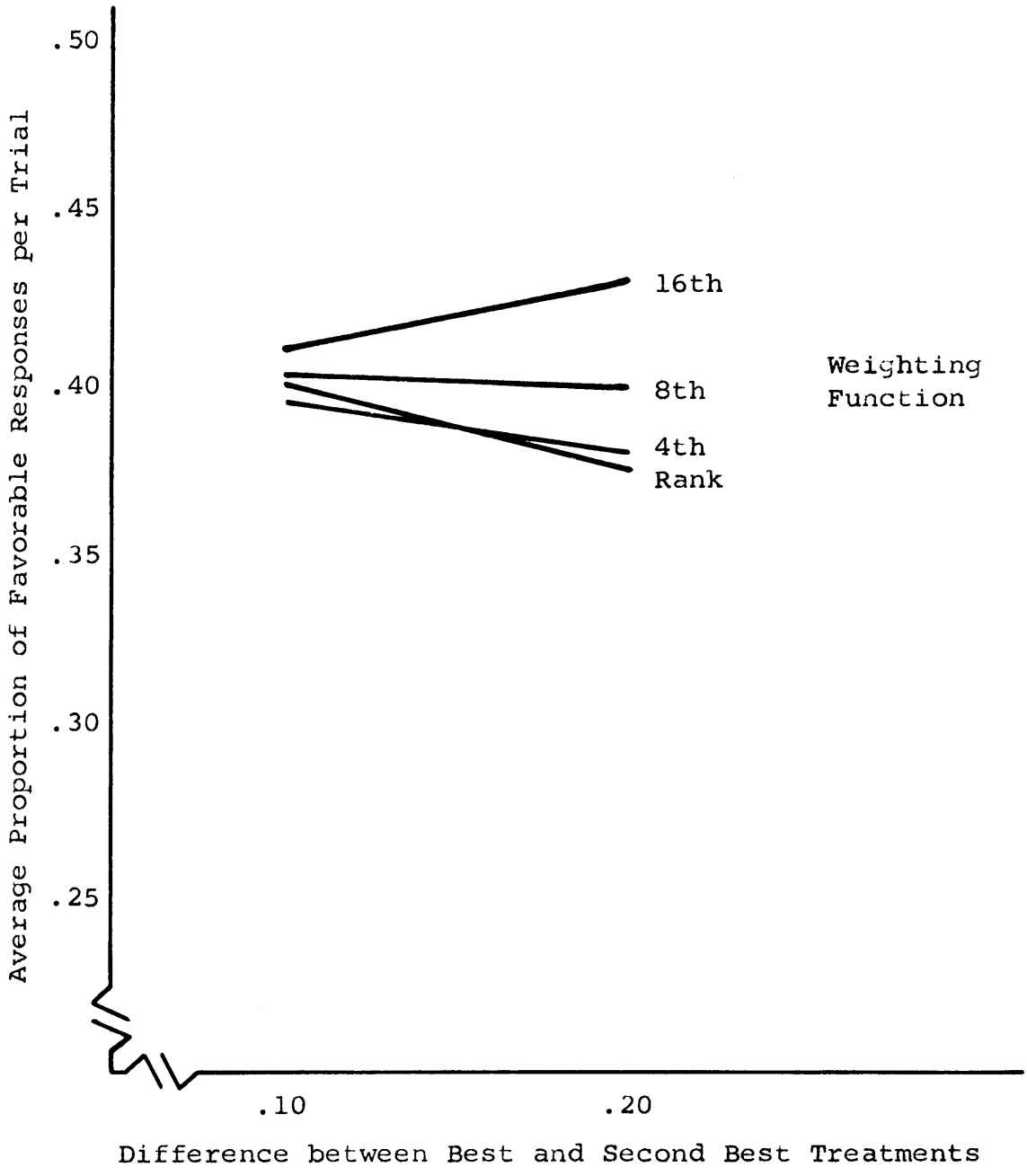

Figure 5

Wtfn · Diff Interaction (Design 2)

The statistically significant ratios in the analysis of variance table are (in order of significance),

(i) the main effect due to the difference between the best and second best treatments.

(ii) the main effect due to the number of treatments.

(iii) the interaction between the number of treatments and the difference between the best and second best treatments.

Consider Table 17 for a practical interpretation of the significant F · ratios.

Table 17

Average Rank of the Best Treatment

Diff \ Treat	4	6	8	Average
.10	1.29	1.77	2.37	1.81
.20	1.03	1.08	1.33	1.15
Average	1.16	1.43	1.85	1.48

The highest rank appears for the 0.20 difference level, which is not surprising since the best treatment is naturally expected to give a better performance under the largest

difference between the best and second best treatments.

The rank of the best treatment goes up as the number of treatments decreases. This result is also reasonable since the rank of the best treatment is expected to be highest when the number of treatments is at a minimum.

The interaction between the number of treatments and the difference between the best and second best treatments is illustrated in Figure 6 (also in Table 17), which shows that the rank of the best treatment is more consistent for the 0.20 difference level, regardless of the treatment level taken into consideration. This result is also quite logical, since a large treatment difference will tend to overshadow other experimental conditions.

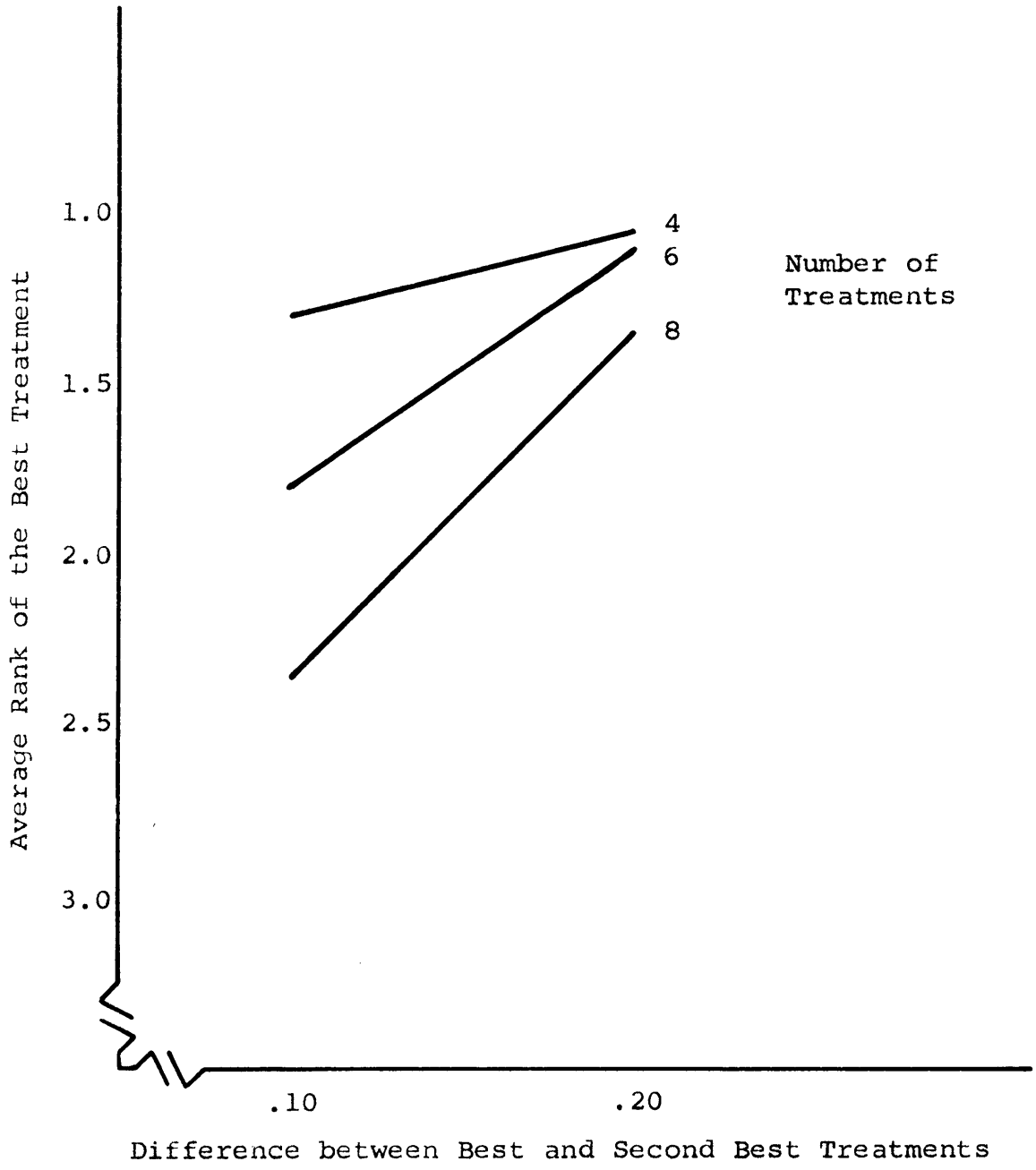

Figure 6

Treat · Diff Interaction (Design 1)

5.4.2 Design 2

The analysis of variance is shown in Table 18.

Table 18

Factor	df		F · ratio
Wtfn	3	.2596	2.36
Treat	2	38.2504	347.42**
Per	1	.0541	
Diff	1	141.3485	1,283.82**
Aver	1	.0076	
Wtfn · Treat	6	.0500	
Wtfn · Per	3	.2859	2.60
Wtfn · Diff	3	.0351	
Wtfn · Aver	3	.1181	1.07
Treat · Per	2	.1221	1.11
Treat · Diff	2	16.2256	147.37**
Treat · Aver	2	.0474	
Per · Diff	1	.2769	2.51
Per · Aver	1	.1225	1.11
Diff · Aver	1	.0004	
Higher order interactions	63	.1246	1.13
Error	9,504	.1101	

The statistically significant ratios in the analysis of variance table are (in order of significance),

(i) the main effect due to the difference between the best and second best treatments.

(ii) the main effect due to the number of treatments.

(iii) the interaction between the number of treatments and the difference between the best and second best treatments.

Consider Table 19 for a practical interpretation of the significant F · ratios.

Table 19

Average Rank of the Best Treatment

Diff \ Treat	4	6	8	Average
.10	1.30	1.76	2.49	1.85
.20	1.02	1.09	1.25	1.12
Average	1.16	1.43	1.87	1.48

The conclusions for this design are almost identical as those for Design 1. It is also of interest to point out (see Table 18) that the rank of the best treatment was not influenced by the weighting function or the number of periods.

VI. SUMMARY OF THE RESULTS

R. J. Taylor and H. A. David (1961) developed a new experimental approach in the screening of prospective drugs to be used in the treatment of cancer. The procedure is a sequential one, where the allocation of patients to drugs is based upon the performance of the drugs in immediately previous periods. Taylor (I.B.M. Program) further developed a computer program to study the effectiveness of the procedure by means of simulated trials.

The purpose of this paper has been to study the sequential allocation procedure further by means of simulation in the form of a factorial experiment. Of primary concern has been the behavior of different weighting functions operating under varying experimental conditions. Response variables have been studied as a means for evaluating the effectiveness of the procedure.

The purpose of cancer research is to gain information concerning the response characteristics of different treatments (drugs) by means of clinical trials. It is evident that the values of the treatment parameters are not known previous to such experimentation, for if they were the

trials would obviously have no meaning. Attention is also called to the fact that there is very little option in the choice of the number of treatments, since this value is usually determined by an outside source related to the particular experimental conditions at the time. This means that the weighting function and the number of periods are the only elements over which the experimenter has full control, and for this reason only these factors will be considered in the summary of the results.

6.1 Relative Frequency of Correct Selection

The use of weighting functions as a method of allocating patients improved the relative frequency of correct selection of the best treatment, but no one function took preference over the others from a practical point of view. Even though the weighting functions did give similar performances, it is of importance to point out that all four of them improved upon the percentage observed for the equal allocation design.

The number of periods did not produce enough variation to be statistically significant, since the percentage of correct selection was about the same for three and six periods.

6.2 Proportion of Patients Allocated to the Best Treatment

The proportion of patients allocated to the best treatment went up as the power of the weighting function increased. This result was to be expected, since the distinguishing characteristic between the weighting functions is that of a greater allocation of patients to the best treatment for each increase in power. All four of the functions improved upon the proportion of allocation observed for the equal allocation design.

The number of periods produced sufficient variation to be statistically significant. Six periods displayed a higher proportion of allocation than three periods.

6.3 Proportion of Favorable Responses per Trial

The proportion of favorable responses per trial showed a definite gain for each increase in the power of the weighting function. This result is reasonable, since it has already been shown that an increase in power results in a greater allocation of patients to the best treatment, therefore a gain is to be expected in the overall proportion. The four functions all improved upon the response ratio as observed for the equal allocation design.

The number of periods produced no significant variation in the proportion of favorable responses.

6.4 Rank of the Best Treatment

An investigation of the results obtained for the rank of the best treatment showed that this study gave strong support to the conclusions previously drawn for the relative frequency of correct selection. As before, the weighting functions displayed no trend, and no one of them was more practical than the rest.

The number of periods produced no significant variation in the rank of the best treatment.

6.5 Conclusions

The sequential allocation procedure improved upon the results of the equal allocation method, but the best weighting function is yet to be determined. The weighting functions turned out to be most important in the allocation of patients to the best treatment, with the allocation going up with each increase in power.

The primary concern of the sequential procedure, that of picking the best treatment (drug) in the fewest number of periods, is yet to be recognized. Trials consisting of

three periods offer definite advantages over the six period arrangement, but a combined investigation of two, three, and four periods is necessary before any final conclusions can be presented.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. R. J. Freund for his sponsorship and helpful guidance throughout the past six months.

Special recognition is given to , of the National Cancer Institute, for his valuable contributions and most important corrections, as well as his continued interest in the problem.

Thanks is given to for her assistance in the operation of the IBM equipment.

Personal gratitude is expressed to for the care with which she typed the final manuscript.

BIBLIOGRAPHY

Kempthorne, O., The Design and Analysis of Experiments, John Wiley and Sons, Inc., New York, (1952) pp. 153-158.

Schultz Jr., E. F., "Rules of Thumb for Determining Expectations of Mean Squares in Analysis of Variance", Biometrics, 11, (June, 1955) pp. 123-135.

Sobel, M. and Huyett, M. J., "Selecting the Best One of Several Binomial Populations", The Bell System Technical Journal, 36, (March, 1957) pp.537-576.

Taylor, R. J. and David, H. A., Sequential Allocation of Patients in Clinical Trials, (June, 1961).

The vita has been removed from the scanned document

To my parents, , in appreciation of their many tangible qualities which makes me proud to call them Mom and Dad.

He saw a star up in the sky,
a vision of tomorrow.
He reflected back at past results,
and shook his head in sorrow.

He looked again into the sky,
another vision seeking.
The past is gone, it's today that counts,
new light in his soul was breaking.

He set his star, he made his plans,
he measured off the distance.
Not much time, but time enough,
by banishing all resistance.

He set about each day to do,
that portion so allotted.
Results inspired his determined course,
failures were quickly blotted.

And day by day he joyfully trod,
the distance soon was traveled.
He stopped, a survey of results he took,
and to himself he marveled.

The price we pay for success today,
determines the reward we get tomorrow.

ABSTRACT

Robert J. Taylor and Herbert A. David developed a new experimental approach in the screening of prospective drugs to be used in the treatment of cancer. The procedure is a sequential one, where the allocation of patients to drugs is based upon the performance of the drugs in immediately previous periods. Taylor further developed a computer program to study the effectiveness of the procedure by means of simulated trials.

The purpose of this paper has been to study the sequential allocation procedure further by means of simulation in the form of a factorial experiment. Of primary concern has been the behavior of different weighting functions operating under varying experimental conditions. Response variables have been studied as a means for evaluating the effectiveness of the procedure. The results of this experiment are comparable with the findings originally presented by Taylor and David.