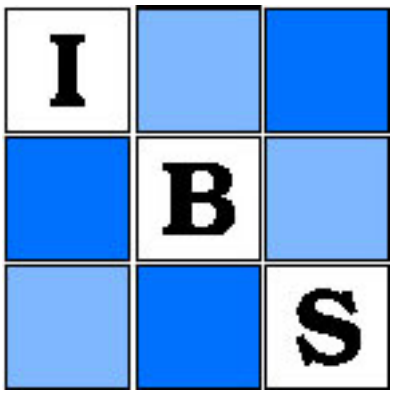

BIOMETRICS 29, 293-300
June 1973

MULTIPLE-STAGE PROCEDURES FOR DRUG SCREENING

J. R. SCHULTZ, F. R. NICHOL, G. L. ELFRING, AND S. D. WEED

Research Laboratories, The Upjohn Company, Kalamazoo, Michigan 49001, U.S.A.

SUMMARY

Completely general methods are developed to obtain the average sample size and the probability of accepting the hypothesis $p = p_0$ for binomial probabilities in multiple stage sampling procedures. An example illustrating the use of a multiple-stage plan of this type for a drug screen is considered. Data on the actual performance of the screen is also given.

1. INTRODUCTION

Direct methods for calculating the operating characteristic (OC) and average sample number (ASN) of closed sequential tests for the mean of a binomial population have been given by Aroian [1968]. These methods were also used by Wilson and Burgess [1971], who specifically considered stages with more than one observation. In general, multiple-stage plans are specified by the number of units examined at each stage, the number of stages, and the acceptance points and the rejection points associated with each stage. The purpose of this report is to give a completely general formulation for obtaining the properties of these plans and to present data on the actual performance of a multiple-stage drug screen.

2. CALCULATION OF OC AND ASN

We consider a sequence of independent random variables $X_1 , X_2 , \cdots$ such that $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$. The sum of these variables will be denoted by S and a specific realization of the sum by s. The objective is to describe a multiple-stage procedure for testing the hypothesis $H_0 : p = p_0$ against the alternative $H_1 : p = p_1$ where $p_1 > p_0$.

Multiple-stage plans are specified by the maximum number of stages, K, the number of units examined at each stage, $(n_1 , n_2 , \cdots , n_K)$, the set of acceptance points, $(a_1 , a_2 , \cdots , a_K)$ and a corresponding set of rejection points, $(r_1 , r_2 , \cdots , r_K)$. The acceptance and rejection points form the boundaries of a test region in which $r_i > a_i$. At the terminal stage $a_K = r_K - 1$ and the maximum number of units required is N_K . When the gth stage is reached, the test statistic is $\sum_{k=1}^{g} s_k$ where s_k is the number of positive

responses observed in stage k. The sequential procedure is then as follows:

$$\text{if } \sum_{k=1}^{g} s_k \leq a_g \text{ , stop sampling and accept } H_0$$

$$\text{if } \sum_{k=1}^{g} s_k \geq r_g \text{ , stop sampling and reject } H_0$$

$$\text{if } a_g < \sum_{k=1}^{g} s_k < r_g \text{ , continue to stage } g + 1.$$

For any given stage, g, with n_g experimental units the sum of the random variables for that stage is S_g and the probability that $S_g = v$ for a specified value of p is given by the binomial distribution function

$$B_g(v, p) = \binom{n_g}{v} p^v (1 - p)^{n_g - v}. \tag{1}$$

Furthermore, the probability that $(S_1 + S_2 + \cdots + S_g) = m$ is given by

$$C_g(m, p) = \sum_{j=E}^{F} C_{g-1}(j, p) B_g(m - j, p) \tag{2}$$

where $E = \max [a_{g-1} + 1, m - n_g]$ and $F = \min [r_{g-1} - 1, m]$.
Thus, the probability that $(S_1 + S_2 + \cdots + S_g) = m$ at the gth stage is the probability that $(S_1 + S_2 + \cdots + S_{g-1}) = u$ times the probability $S_g = v$ summed over all eligible u, v such that $u + v = m$. This allows $C_g(m, p)$ to be calculated recursively. The eligible values of u and v are determined by the accept and reject points.

The probability of accepting H_0 at stage g is then

$$L_g(p) = \sum_{m=a_{g-1}+1}^{a_g} C_g(m, p) \tag{3}$$

and the probability of accepting H_0 through the final stage is given by

$$L(p) = \sum_{g=1}^{K} L_g(p). \tag{4}$$

Similarly, the probability of rejecting H_0 at stage g is

$$R_g(p) = \sum_{m=r_g}^{r_{g-1}+n_g-1} C_g(m, p) \tag{5}$$

and the cumulative probability of rejecting H_0 is then

$$R(p) = \sum_{g=1}^{K} R_g(p). \tag{6}$$

The average sample number is given by

$$\text{ASN} = \sum_{g=1}^{K} N_g[L_g(p) + R_g(p)] \tag{7}$$

where $N_g = (n_1 + n_2 + \cdots + n_g)$. The variance of the sample number is

$$\text{Var (sample number)} = \sum_{g=1}^{K} N_g^2[L_g(p) + R_g(p)] - (\text{ASN})^2. \tag{8}$$

These plans can easily be modified to place all the acceptance probability in the terminal stage. There is then only one acceptance point at $a_K = r_K - 1$ and the procedure for the gth stage is as follows:

$$\text{if} \quad \sum_{k=1}^{g} s_k \geq r_g\,, \quad \text{stop sampling and reject } H_0$$

$$\text{if} \quad \sum_{k=1}^{g} s_k < r_g\,, \quad \text{continue to stage } g + 1.$$

In this case

$$C_g(m, p) = \sum_{j=E^*}^{F^*} C_{g-1}(j, p)B_g(m - j, p) \tag{9}$$

where $E^* = \max [0, m - n_g]$ and $F^* = \min [r_{g-1} - 1, m]$
and the total probability of accepting H_0 is

$$L(p) = L_K(p) = \sum_{m=0}^{a_K} C_K(m, p). \tag{10}$$

The probability of rejecting H_0 at stage g is given by (5), the total probability of rejection is given by (6) and the average sample number is

$$\text{ASN} = N_K L(p) + \sum_{g=1}^{k} N_g R_g(p). \tag{11}$$

The present formulation allows complete flexibility in specification of the test region, the size of each stage and the total number of stages. Furthermore, the procedure is particularly adaptable to efficient computer application.[1] This facilitates the examination of alternative multiple-stage plans to find the most suitable design.

3. EXAMPLE AND DISCUSSION

In preliminary studies of a test for antiviral agents in experimentally infected mice, it was found that 6.0% of 583 animals treated with a positive standard and 84.6% of 625 control animals died during the seven day ob-

[1] An interactive computer program written in FORTRAN IV for time-sharing application will be provided upon request.

servation period. Using this data as a guide, it was determined that when the true probability of death (p) associated with a particular compound under these conditions is 0.25 or less the material would be of interest and when p is 0.60 or greater the compound would be of no interest. It was further specified that the probability of accepting an interesting compound as defined above should be no less than 0.95 and the probability of accepting an uninteresting compound should be no greater than 0.05. Another consideration in this application was that group sizes of ten animals would result in the most efficient use of facilities and labor.

The properties of alternative plans with ten animals per stage were then examined by the methods given in section 2. A Wald region was constructed to guide the initial selection of a plan and its characteristics were obtained for $p = 0.25$ and 0.60. The test region was then modified on the basis of these results until the above requirements were met. Since it was anticipated that most of the compounds would possess little or no activity, particular attention was given to the ASN for values of $p \geq 0.60$. The plan selected for the screen is given in Table 1. With this three-stage design a test substance

TABLE 1

THREE-STAGE SCREENING PLAN

	Accumulated Deaths		
Stage (g)	a_g	r_g	n_g
1	ND[+]	6	10
2	ND	10	10
3	12	13	10

[+] No Decision

is declared active when H_0 is accepted or it is declared inactive when H_0 is rejected. Drugs may be declared inactive at any stage but must pass through all three stages to be declared active. Table 2 gives the probability of accepting H_0 , $L(p)$, the ASN and Var (sample number) for values of $p = 0.05(0.05)0.80$.

Each screening run consisted of up to 60 groups of 10 mice each. Animals in six of these groups were treated with vehicle only and another six groups were treated with a positive standard. These twelve groups were used to monitor the test system and the remaining groups were used to test new compounds. A total of 1,548 unselected compounds were classified in 37

TABLE 2

PROPERTIES OF THE THREE-STAGE PLAN

P	L(p)	ASN	Var(Sample Number)
0.05	0.9999	30.00	0.00
0.10	0.9998	30.00	0.06
0.15	0.9983	29.97	0.57
0.20	0.9902	29.86	2.69
0.25	0.9597	29.52	8.51
0.30	0.8802	28.77	20.25
0.35	0.7322	27.42	38.13
0.40	0.5298	25.39	58.09
0.45	0.3218	22.79	72.45
0.50	0.1589	19.93	74.67
0.55	0.0617	17.16	64.36
0.60	0.0181	14.78	47.14
0.65	0.0038	12.94	29.90
0.70	0.0005	11.65	16.68
0.75	0.0000	10.81	8.14
0.80	0.0000	10.33	3.31

runs. The number of compounds classified at each stage is given in Table 3. All but 57 of the drugs were declared inactive in the first stage; eight of these were subsequently declared active in the third stage. Excluding the vehicle and standard groups, a total of 16,240 mice were used giving an average sample size of 10.5.

The responses of animals treated with vehicle and positive standard

TABLE 3

NUMBER OF COMPOUNDS CLASSIFIED AT EACH STAGE

Stage	No. Compounds Classified		No. Mice Used
	Accepted	Rejected	
1	-	1491	14,910
2	-	38	760
3	8	11	570
Total	8	1540	16,240

were continuously followed as testing proceeded. The values of p obtained from the preliminary tests were used to construct p-charts which were maintained in the laboratory. For each run, the number of deaths in the 60 control animals and in the 60 animals treated with positive standard were recorded on these charts. The objective of these graphs was to facilitate recognition of possible long-term trends. The performance of the screen was also followed by considering each of the 12 groups as a "first stage" test. The probability of rejecting H_0 is given for each stage in Table 4. A compound is rejected in the first stage when six or more of the animals die. Table 4 shows that this would be expected in 99% of the groups treated with vehicle if the value of p remains at 0.85. Of the 222 vehicle treated groups, 218

TABLE 4

PROBABILITY OF REJECTING H_0 AT EACH STAGE FOR SELECTED VALUES OF p

Stage	p			
	.10	.25	.60	.85
1	0.0001	0.0197	0.6331	0.9901
2	0.0000	0.0084	0.2560	0.0098
3	0.0000	0.0121	0.0928	0.0000

(98.2%) had at least six dead animals. None of the groups treated with the positive standard had more than five deaths. Table 4 also shows that such an event would occur only very rarely at the level of activity observed with this material in the preliminary studies. Thus, the test system classified these groups as predicted from the preliminary estimates for vehicle and standard treatments. Any effects resulting from run to run variation in the values of p or deviations from the binomial model were not of sufficient magnitude to influence the screen's ability to appropriately classify compounds of these two types.

Procedures in which testing is carried out in stages have been used extensively in both acceptance sampling and drug screening. Armitage and Schneiderman [1958], Schneiderman [1961] and Roseberry and Gehan [1964] have considered multiple stage procedures for the case in which the variable of interest is normally distributed. Other procedures have been developed for the normal case which consider the costs of testing (Davies [1963]) and the costs of making wrong decisions (Dunnett [1961]).

The similarity between drug screening and acceptance sampling has been discussed by Davies [1958]. A test procedure can be designed for either application by considering two points on the OC curve and examining plans on the basis of these specifications. For screening, these points can be defined relative to activity of standard treatments. Thus, information giving the proportion of animals which respond to a positive standard and the proportion which respond to a negative standard is required. Some flexibility in setting error rates is usually necessary because it may not be possible to construct a practical screening plan meeting exact specifications. In screening applications the number of experimental units examined at each stage is often determined by practical laboratory restrictions. A tentative plan is specified on the basis of a Wald region or from previous knowledge and its properties obtained by the methods given in section 2. These results are then used to modify the test region until a suitable plan is found. This procedure is rapidly accomplished with interactive computer facilities.

Multiple-stage designs retain much of the advantage in reduced sample size obtained with fully sequential procedures, particularly when most of the agents to be tested have little or no activity. A single-stage plan consisting of 26 animals with accept and reject points of 10 and 11, respectively, gives values of $L(p)$ which are essentially identical to those in Table 2. The single stage plan would have required 24,000 additional animals to classify the 1,548 compounds which were examined in the three-stage screen.

PROCEDURE A PLUSIEURS DEGRES POUR TRI DE DROGUE

RESUME

On développe des méthodes absolument générales pour obtenir la taille moyenne de l'échantillon et la probabilité d'accepter l'hypothèse $p = p_0$ pour des probabilités binomiales dans des procédures d'échantillonnage à plusieurs degrés. On considère un exemple qui illustre l'utilisation d'un plan à plusieurs degrés de ce type pour un tri de drogue. On donne aussi des éléments sur la performance actuelle du tri.

REFERENCES

Armitage, P. and Schneiderman, M. [1958]. Statistical problems in a mass screening program. *Ann. N. Y. Acad. Sci. 76*, 896–908.

Aroian, L. A. [1968]. Sequential analysis, direct method. *Technometrics 10*, 125–32.

Davies, O. L. [1958]. The design of screening tests in the pharmaceutical industry. *Bull. Intern. Statist. Inst. 36*, 226–41.

Davies, O. L. [1963]. The design of screening tests. *Technometrics 5*, 481–9.

Dunnett, C. W. [1961]. Statistical theory of drug screening. In: *Quantitative Methods in Pharmacology*. North-Holland, Amsterdam, Holland.

Roseberry, T. D. and Gehan, E. A. [1964]. Operating characteristic curves and accept-reject rules for two and three stage screening procedures. *Biometrics 20*, 73–84.

Schneiderman, M. [1961]. Statistical problems in the screening search for anticancer drugs by the National Cancer Institute of the United States. In: *Quantitative Methods in Pharmacology*, North-Holland, Amsterdam, Holland.

Wilson, E. B. and Burgess, A. R. [1971]. Multiple sampling plans viewed as finite Markov chains. *Technometrics 13*, 371–83.

Received March 1972, *Revised August* 1972

Key Words: Binomial distribution; Sequential designs; Exact methods; Screening data.