PAIRED SEQUENTIAL ANALYSIS OF CLINICAL TRIAL DATA: A SIMULATION STUDY OF TWO SEQUENTIAL DESIGNS BY BROSS

by

W.F. Page, Ph.D.

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Memio Series No. 1070

May 1976

PAIRED SEQUENTIAL ANALYSIS OF CLINICAL TRIAL DATA: A SIMULATION STUDY OF TWO SEQUENTIAL DESIGNS BY BROSS

By W. F. Page, Ph.D.*
Department of Biostatistics
University of North Carolina

In a clinical trial testing the difference in "cure rates" between treatments, sequential analysis can be used to analyze the data as they come in, thus making it possible to stop the experiment as soon as a statistical decision is reached. One of the simplest methods of analyzing two treatments is to plot the data on a "sequential chart". Bross [2] provides two such sequential charts, of his own design, in graphical form, whereas Armitage [1] presents these same two in tabular form. In this paper, the two Bross sequential charts are studied using simulation methods. More detailed information is provided for these two sequential designs, and then the simulation results are checked against the performance specifications furnished by Bross.

Background and Methods

In this sequential analysis method the outcome data are analyzed in pairs, one patient per pair on each treatment (although the randomization to treatment group need not be in blocks of two), and the outcome of each pair is used to update a sequential chart. A look at Figure 1 will help illustrate the basic idea of a sequential chart. If treatment A

---

*Support for this research was provided through a training grant furnished to the Department of Biostatistics, University of North Carolina at Chapel Hill by the Burroughs Wellcome Co., Research Triangle Park, N. C.

produces a cure and treatment B doesn't, one places an "X" in the box to the right of the last box, but if treatment B produces a cure and treatment A doesn't, the "X" is placed instead in the next square above. If both treatments produce cures or both produce failures, no "X" is plotted. The results of this plotting is a zig-zag path of "X's" moving above and to the right in single square increments.

Figure 1 about here

After an "X" is plotted, the experimenter is faced with one of two situations, either a "barrier" on the chart has been crossed, or no barriers have been crossed. If a barrier has been crossed, the chart will indicate which statistical decision should be made (i.e., either treatment A is superior, treatment B is superior, or the two do not differ significantly); if no barrier is crossed, the experiment continues with the plotting of another pair. Because this sequential chart is designed so that the parade of "X's" must eventually cross one of the barriers, it is called "closed".

Ideally, a statistician would create a new chart for each trial to be analyzed, tailoring that chart to the expected proportion of cures for the two treatments (which are denoted by P_1 and P_2, respectively, throughout this paper), and the desired type I and type II error. In practice, however, this somewhat tedious task is seldom undertaken, and instead a ready-made chart is used. One of the goals of this paper is to provide more detailed power and sample size information on the two ready-made

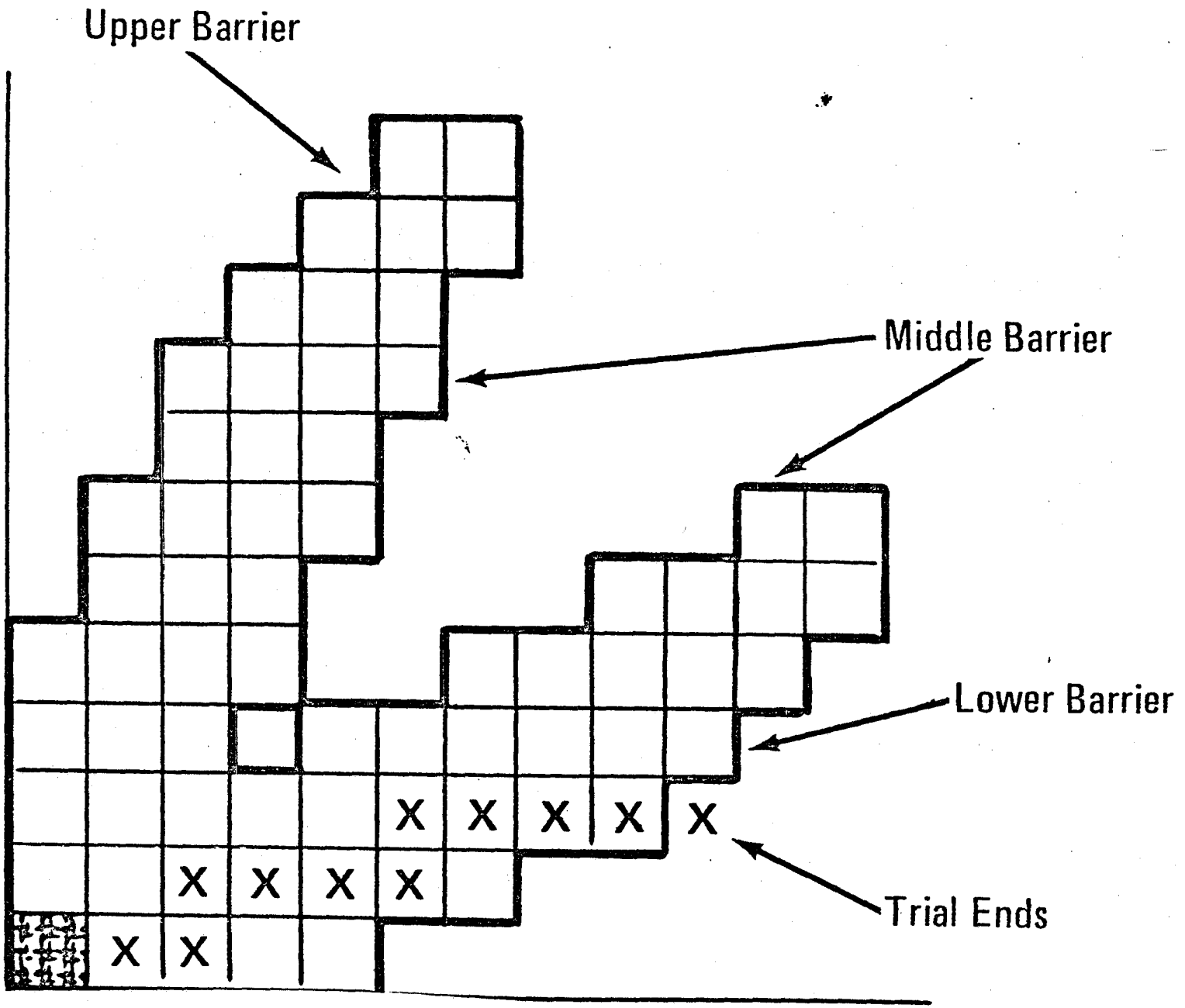


Figure 1 - HYPOTHETICAL SEQUENTIAL CHART

Bross sequential designs, in the hope that more statisticians will use these charts.

Because the Bross charts were derived by a "cut-and-try" method, a theoretical derivation of their properties turns out to be difficult. Hence computer simulation was chosen as the tool to use in studying these designs. Sequential pairs were generated whose outcomes were charted on each of the sequential plans. Each string of random pairs was followed to completion (i.e., some barrier is crossed), and 1,000 such strings were generated for each choice of P_1 and P_2, $P_1 < P_2$, in multiples of 0.10.

One must be cautious in presenting and interpreting simulation results. Because these simulation results depend upon the choice of "seed" for the random number generator, the number of trials performed, etc., one must be sure to treat them as just one "typical" convenient sample outcome. There is also the question of whether the simulation output should be "smoothed", to conform to known theoretical constraints. We have distinguished two situations here: in the comparison of simulation results with performance specifications, the actual raw values have been tabulated (Table 3 and Figures 3A, 3B); but in all other cases, the results have been smoothed.

Power of the Bross Sequential Plans

Since the type I error is fixed in both the sequential plans (for plan A, $2\alpha = .20$ and for plan B, $2\alpha = .10$), the properties of interest become

power and average number of pairs needed to reach a sequential decision (average sample size, for short). In considering power, one must keep in mind that there are three possible outcomes for each trial, and hence one can compute the power of one- and two-tailed tests of hypotheses.

For a two-tailed test the null hypothesis $H_o : P_1 = P_2$ is tested against the alternative hypothesis $H_a : P_1 \neq P_2$. This means that H_o is rejected if the charted path crosses either the upper or lower barrier, the probability of this being the sum of the two probabilities. When $P_1 = P_2 = .5$, for example, the probabilities from Table I are .1065 and .1065, which give .2130 when added (this is actually the type I error since H_o is true in this example). If, say, $P_1 = .5$ and $P_2 = .7$, the probability of rejecting H_o is .9005 + .0005 = .9010, the power of the test. For all cases where $P_1 = P_2$, both "tails" (probabilities of crossing the upper or lower barriers) have been adjusted to have the same size.

TABLES 1A and 1B about here

For the one-tailed test, the null hypotheses, have the form $H_o : P_1 \leq P_2$ or $H_o : P_1 \geq P_2$ so that the alternative hypotheses are one-sided $H_a : P_1 > P_2$ or $H_a : P_1 < P_2$, respectively. For specificity, take $H_o : P_1 \leq P_2$, so that $H_a : P_1 > P_2$. In terms of the sequential design, the null hypothesis is thereby accepted when the path merely enters the lower section of the chart i.e., once it is impossible for the upper boundary to be crossed; it is not necessary that the path of "X's" cross the middle or lower

TABLE 1A

Proportion of Plan A trials which reach, respectively, the three sequential decisions $P_1 < P_2$, $P_1 = P_2$ and $P_1 > P_2$, by various values of P_1 and P_2*.

P_2									
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9950	.8930	.1065
.90	.0000	.0000	.0000	.0000	.0000	.0000	.0050	.1065	.7870
	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0005	.1065
	1.0000	1.0000	1.0000	1.0000	.9985	.9545	.6590	.1065	
.80	.0000	.0000	.0000	.0000	.0015	.0450	.3400	.7870	
	.0000	.0000	.0000	.0000	.0000	.0005	.0000	.1065	
	1.0000	1.0000	1.0000	.9900	.9005	.5550	.1065		
.70	.0000	.0000	.0000	.0100	.0990	.4380	.7870		
	.0000	.0000	.0000	.0000	.0005	.0070	.1065		
	1.0000	1.0000	.9900	.8900	.4990	.1065			
.60	.0000	.0000	.0100	.1080	.4900	.7870			
	.0000	.0000	.0000	.0020	.0110	.1065			
	1.0000	.9985	.9005	.4990	.1065				
.50	.0000	.0015	.0990	.4900	.7870				
	.0000	.0000	.0005	.0110	.1065				
	1.0000	.9545	.5550	.1065					
.40	.0000	.0450	.4380	.7870					
	.0000	.0005	.0070	.1065					
	.9950	.6590	.1065						
.30	.0050	.3400	.7870						
	.0000	.0010	.1065						
	.8930	.1065							
.20	.1065	.7870							
	.0005	.1065							
	.1065								
.10	.7870								
	.1065								
	.10	.20	.30	.40	.50	.60	.70	.80	.90

P1

* Table entries are symmetric; we have chosen $P_1 < P_2$.

TABLE 1B

Proportion of Plan B trials which reach respectively three sequential decisions $P_1 < P_2$, $P_1 = P_2$, and $P_1 > P_2$ by various values of P_1 and P_2*.

P_2									
	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9945	.8305	.0482
.90	.0000	.0000	.0000	.0000	.0000	.0000	.0055	.1695	.9036
	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0482
	1.000	1.000	1.000	1.000	.9960	.9310	.5380	.0482	
.80	.0000	.0000	.0000	.0000	.0040	.0690	.4615	.9036	
	.0000	.0000	.0000	.0000	.0000	.0000	.0005	.0482	
	1.0000	1.0000	1.0000	.9890	.8555	.4170	.0482		
.70	.0000	.0000	.0000	.0110	.1445	.5790	.9036		
	.0000	.0000	.0000	.0000	.0000	.0040	.0482		
	1.0000	1.0000	.9890	.8340	.3680	.0482			
.60	.0000	.0000	.0110	.1660	.6295	.9036			
	.0000	.0000	.0000	.0000	.0025	.0482			
	1.0000	.9960	.8555	.3680	.0482				
.50	.0000	.0040	.1445	.6295	.9036				
	.0000	.0000	.0000	.0025	.0482				
	1.0000	.9310	.4170	.0482					
.40	.0000	.0690	.5790	.9036					
	.0000	.0000	.0040	.0482					
	.9945	.5380	.0482						
.30	.0055	.4615	.9036						
	.0000	.0005	.0482						
	.8305	.0482							
.20	.1695	.9036							
	.0000	.0482							
	.0482								
.10	.9036								
	.0482								
	.10	.20	.30	.40	.50	.60	.70	.80	.90

P_1

* Table entries are symmetric; we contrain $P_1 \leq P_2$.

barrier for H_o to be accepted. Hence, H_a is accepted only when the upper barrier is crossed, and that probability is simply the single "upper tail" probability. When $P_1 = P_2 = .5$, for example, the probability from Table IA is simply .1065, which is half the previous two-tailed value. If $P_1 = .5$ and $P_2 = .7$, the probability is .0005 for rejecting $H_o : P_1 \leq P_2$, and it is .9005 for rejecting $H_o : P_1 \geq P_2$. Because the labeling of P_1 and P_2 are arbitrary, Tables IA and IB impose the condition $P_1 \leq P_2$ and then furnish both upper and lower tails.

Sample Size for the Bross Sequential Plans

The "average sample size", here the mean number of pairs to reach a sequential decision, is a statistic of great interest in designing a trial. It should be noted that, even though the Bross designs are closed (and so there is a maximum number of pairs needed to reach a decision), only the discordant pairs are charted. Hence, if both members of a pair were cures or failures, that pair does not add information to the trial, whereas it does add to the sample size.

Bross' article includes some comparisons of the median sequential sample size with the usual fixed sample size. On the average, the sequential sample sizes are smaller, but this does not rule out the possibility that a sequential trial may run longer than the fixed sample size trial. The standard deviation of the sample size distribution has been included in Tables 2A and 2B along with the mean sequential sample size, to provide more information on the "spread" of the distribution of sequential sample sizes. However, even this statistic does not really answer the question of how probable it is that a long sequential trial will occur.

TABLE 2A

Number of Pairs Needed to Reach Sequential Decision for Plan A:

Mean ± Standard deviation by values of $P_1 \leq P_2$

P_2	.10	.20	.30	.40	.50	.60	.70	.80	.90
.90	8.893 ±1.84	10.369 ±2.78	12.453 ±4.21	15.338 ±5.97	19.987 ±8.89	28.000 ±13.81	46.601 ±24.31	90.494 ±42.02	163.683 ±52.01
.80	10.369 ±2.78	12.419 ±4.15	15.819 ±6.69	21.508 ±10.58	30.837 ±15.57	47.739 ±23.29	73.179 ±28.46	91.318 ±29.19	
.70	12.453 ±4.21	15.819 ±6.69	21.757 ±10.70	31.476 ±16.20	46.768 ±21.51	61.337 ±22.92	70.265 ±21.85		
.60	15.338 ±5.97	21.508 ±10.58	31.476 ±16.20	45.979 ±20.62	57.825 ±19.96	60.945 ±18.75			
.50	19.987 ±8.89	30.837 ±15.57	46.768 ±21.51	57.825 ±19.96	58.933 ±17.52				
.40	28.000 ±13.81	47.739 ±23.29	61.337 ±22.92	60.945 ±18.75					
.30	46.601 ±24.31	73.179 ±28.46	70.265 ±21.85						
.20	90.494 ±42.02	91.318 ±29.19							
.10	163.683 ±52.01								

P_1

* Table entries are symmetric; we have chosen $P_1 \leq P_2$.

TABLE 2B

Average number of pairs to reach sequential decisions for Plan B: Means ± standard deviations were determined by stimulations and smoothed. Entries tabulated by P_1 and P_2*.

P_2	.1	.2	.3	.4	.5	.6	.7	.8	.9
.9	10.060 (±1.87)	11.745 (±2.86)	13.813 (±4.18)	17.386 (±6.26)	23.090 (±10.24)	33.121 (±16.51)	55.007 (±29.71)	112.627 (±52.18)	167.162 (±60.05)
.8	11.745 (±2.86)	14.439 (±4.68)	18.016 (±7.52)	24.384 (±11.95)	37.411 (±19.64)	58.878 (±28.87)	84.351 (±34.25)	93.228 (±33.61)	
.7	13.813 (±4.18)	18.016 (±7.52)	25.504 (±12.91)	37.908 (±19.51)	57.606 (±26.84)	69.940 (±27.97)	72.348 (±25.83)		
.6	17.386 (±6.26)	24.384 (±11.95)	37.908 (±19.51)	56.096 (±24.96)	64.419 (±24.99)	61.903 (±22.09)			
.5	23.090 (±10.24)	37.411 (±19.64)	57.606 (±26.84)	64.419 (±24.99)	58.140 (±21.29)				
.4	33.121 (±16.51)	58.878 (±28.87)	69.940 (±27.97)	61.903 (±22.09)					
.3	55.007 (±29.71)	84.351 (±34.25)	72.348 (±25.83)						
.2	112.627 (±52.18)	93.228 (±33.61)							
.1	167.162 (±60.05)								

P_1

* Entries are symmetric; we constrain $P_1 \leq P_2$.

TABLES 2A and 2B about here

Looking closely at Tables 2A and 2B, one decides that the most tro ublesome situation is when $P_1 = P_2$. When the two cure rates are equal or nearly equal, and especially when they are near 0 or 1, not only does the average sample size increase, but the variance does also. It would be helpful to have an empirical frequency distribution of sequential sample size for these situations. Figures 2A and 2B are a compromise. Rather than plot the entire empirical frequency distribution, several percentiles of the distribution are plotted. These values have been smoothed somewhat since the curve is known to be symmetric around the point $P_1 = P_2 = 0.50$. With these data, one can now make statements about the probability of very long runs, for example when $P_1 = P_2 = .20$, 95% of the time the trial should take less than 141 pairs, using Plan A.

Figures 2A and 2B about here

Comparison of simulation data with Bross' performance specifications

In this section, some of the previous types of simulation results are compared with the specifications given by Bross; in producing these comparisons, however, no smoothing has been performed. Because our simulation results are tabulated on the basis of P_1 and P_2 for ease of use, they are not all commensurate with Bross' benchmarks. In particular, the properties of the two Bross sequential plans are more easily expressed in terms of different parameters, p* and K

FIGURE 2A

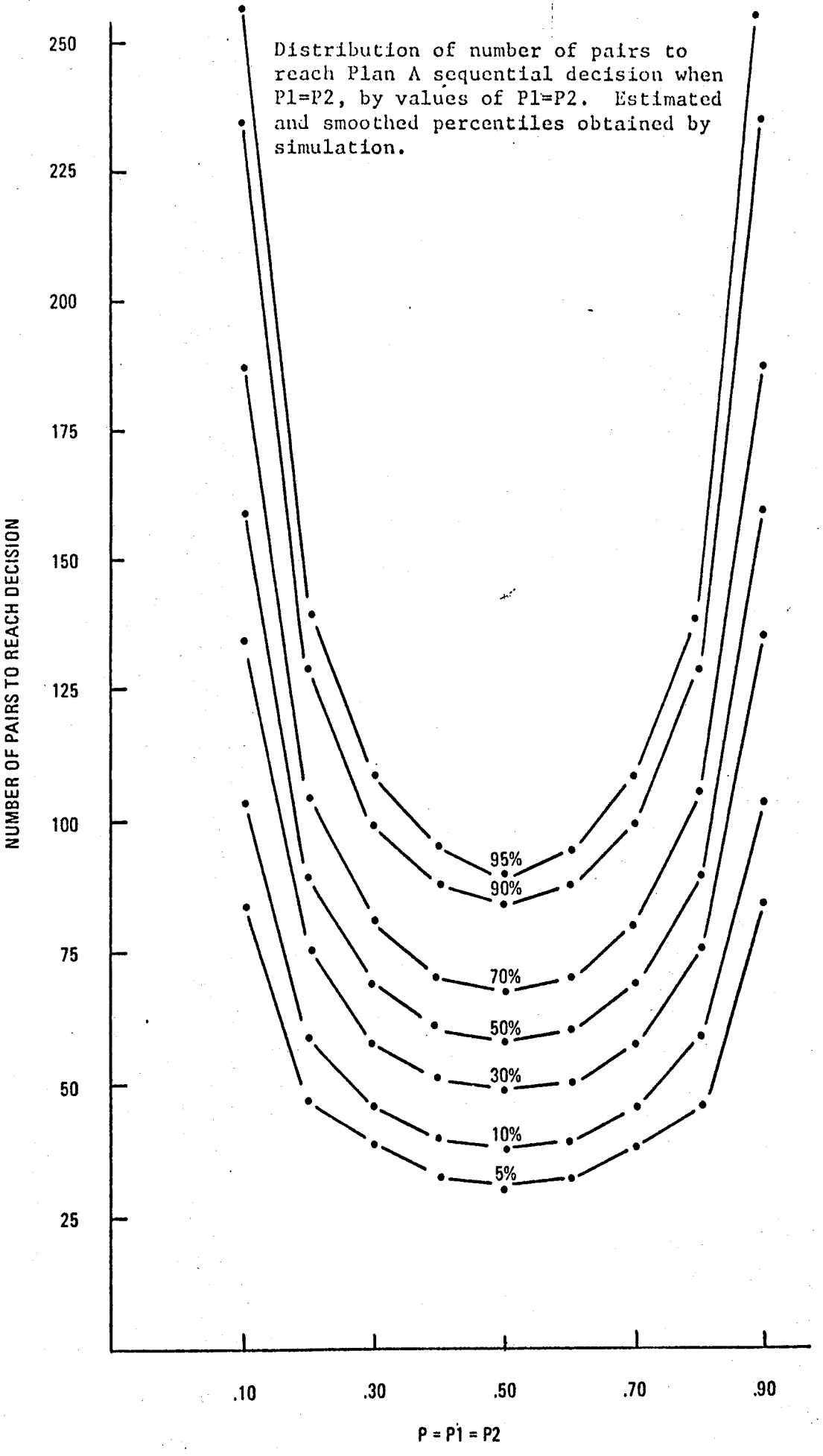
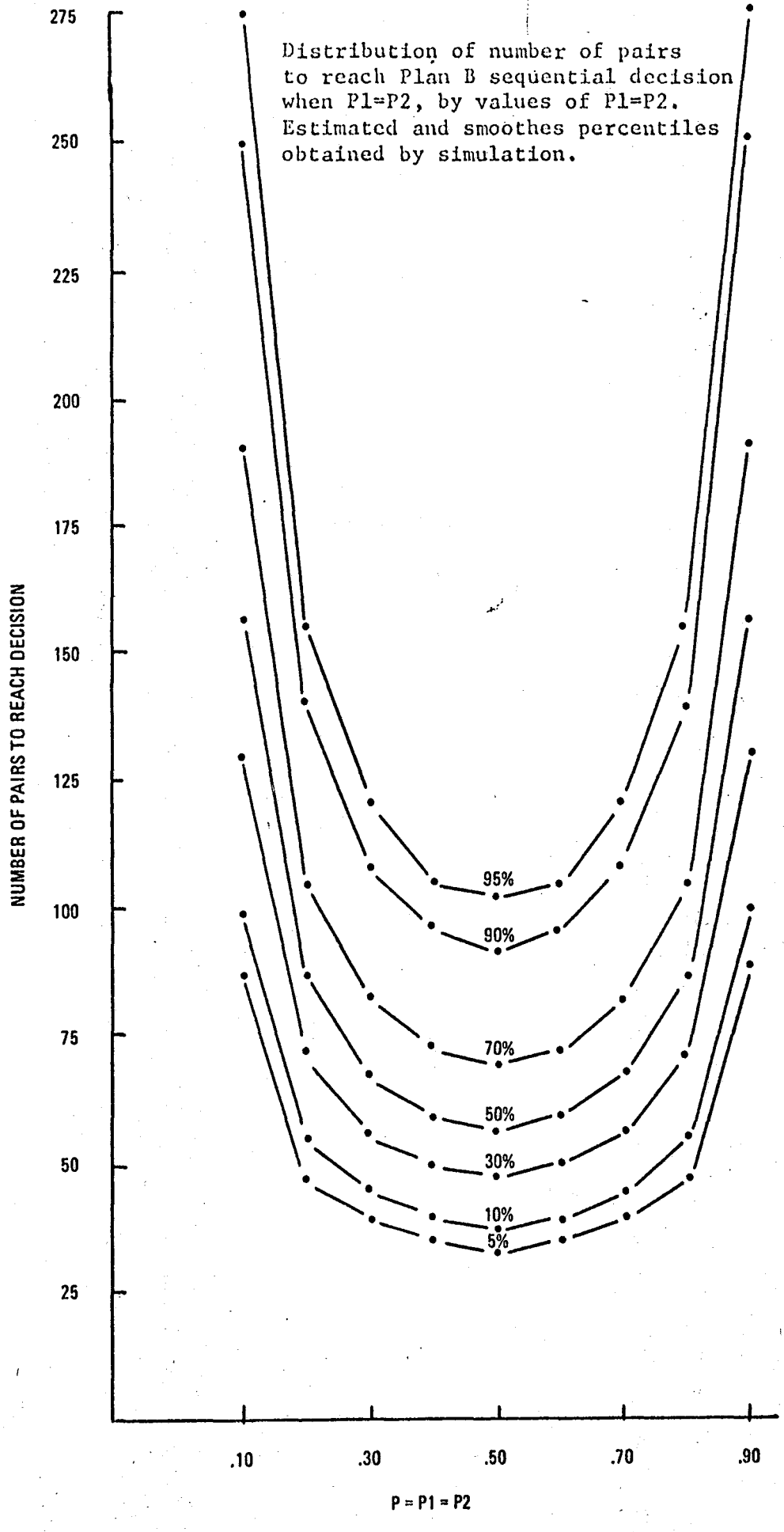

Distribution of number of pairs to reach Plan A sequential decision when P1=P2, by values of P1=P2. Estimated and smoothed percentiles obtained by simulation.
NUMBER OF PAIRS TO REACH DECISION
250
225
200
175
150
125
100
75
50
25
95%
90%
70%
50%
30%
10%
5%
.10
.30
.50
.70
.90
P = P1 = P2

FIGURE 2B


Distribution of number of pairs to reach Plan B sequential decision when P1=P2, by values of P1=P2. Estimated and smoothes percentiles obtained by simulation.
NUMBER OF PAIRS TO REACH DECISION
275
250
225
200
175
150
125
100
75
50
25
95%
90%
70%
50%
30%
10%
5%
.10
.30
.50
.70
.90
P = P1 = P2

defined as $p^* = P_1(1-P_2)/(P_1(1-P_2) + (1-P_1)P_2)$ and $K = 1/(P_1(1-P_2) + P_2(1-P_1))$ (Armitage [1] uses, respectively, $\Theta = p^*$ and $\phi = 1/K$).

The value of $1/K$ is the proportion of "chartable", i.e., untied, pairs to be expected for given values of P_1 and P_2. Thus p^* is the proportion of "successes" for treatment 1, having adjusted for the number of chartable pairs. When $P_1 = P_2$ then one can easily see that $p^* = 0.50$ and when $P_1 \neq P_2$, the greater the difference, the further p^* is from 0.50. In deriving the power for the sequential test of H_o, it is the value of p^* which is important, and in estimating the average number of pairs to decision, it is the value of K which is important.

TABLE 3 about here

Table 3 compares the computed and theoretical values for power, and Figures 3A and 3B make that comparison for the median sample size. We have used the maximum number of cases for these comparisons; for example, when $p^* = 0.50$, the data for all situations where $P_1 = P_2$ is used, whereas when $p^* = .7$ we have only the case of $P_1 = .30$, $P_2 = .50$. It should be noted here that since the labeling of "cure" or "fail" is arbitrary, the values of P_1 and P_2, and hence p^*, are symmetric around 0.50, i.e., $p^* = .70$ is the same as $p^* = .30$. This fact is implicit in Tables IA and IB, where the constraint $P_1 \leq P_2$ is imposed without loss of generality, and in Table 3 we table $p^* = .5, .4, .3$ rather than Bross' .5, .6, .7.

TABLE 3

Percentage of trials which reach, respectively, the three sequential decisions $P_1 < P_2$, $P_1 = P_2$ and $P_1 > P_2$: comparison of theoretical and simulated values for various choices of p*.

Plan A

	Simulation Results			Bross Theoretical Values		
	$P_1 < P_2$	$P_1 = P_2$	$P_1 > P_2$	$P_1 < P_2$	$P_1 = P_2$	$P_1 > P_2$
p*						
.5	10.63	78.72	10.65	10.75	78.50	10.75
.4	49.00	49.90	1.10	49.95	49.21	0.84
.3	90.05	9.90	.05	90.54	9.46	0.01

Plan B

	Simulation Results			Bross Theoretical Values		
	$P_1 < P_2$	$P_1 = P_2$	$P_1 > P_2$	$P_1 < P_2$	$P_1 = P_2$	$P_1 > P_2$
p*						
.5	5.01	90.36	4.63	4.90	90.20	4.90
.4	36.80	62.95	0.25	37.12	62.88	0.20
.3	85.55	14.45	0.00	85.95	14.04	0.01

Figure 3 about here

For the average sample size data, Bross gives a formula for the median path length depending upon a design parameter and the previously defined parameter K. His formula is:

average median length of path = $2 \cdot K \cdot$ (median length),

where the factor of 2 is present because Bross counts individuals (and not pairs) and the median length is dependent on the p* value and the design chosen (plan A or B). It should be pointed out that these calculations are based on a two-tailed test. Sample sizes would be somewhat reduced if one-tailed tests were performed. Bross does not derive estimates of sample size for the one-tailed tests and neither do we. The agreement between observed and predicted values is quite good for both power and average sample size.

Conclusion

In providing more detailed tables of power and average sample size for the two Bross designs, we hope that we will have encouraged the use of these sequential tests among statisticians. The close agreement between the previously published performance specifications and our simulation results helps justify the choice of that technique. Finally, we hope that our data on the distribution of average sample size (and particularly the upper "tail" of this distribution) will be especially

FIGURE 3A

Average number of pairs to sequential decision by values of P=P1=P2: Comparison of simulation results with expected values for Plan A.

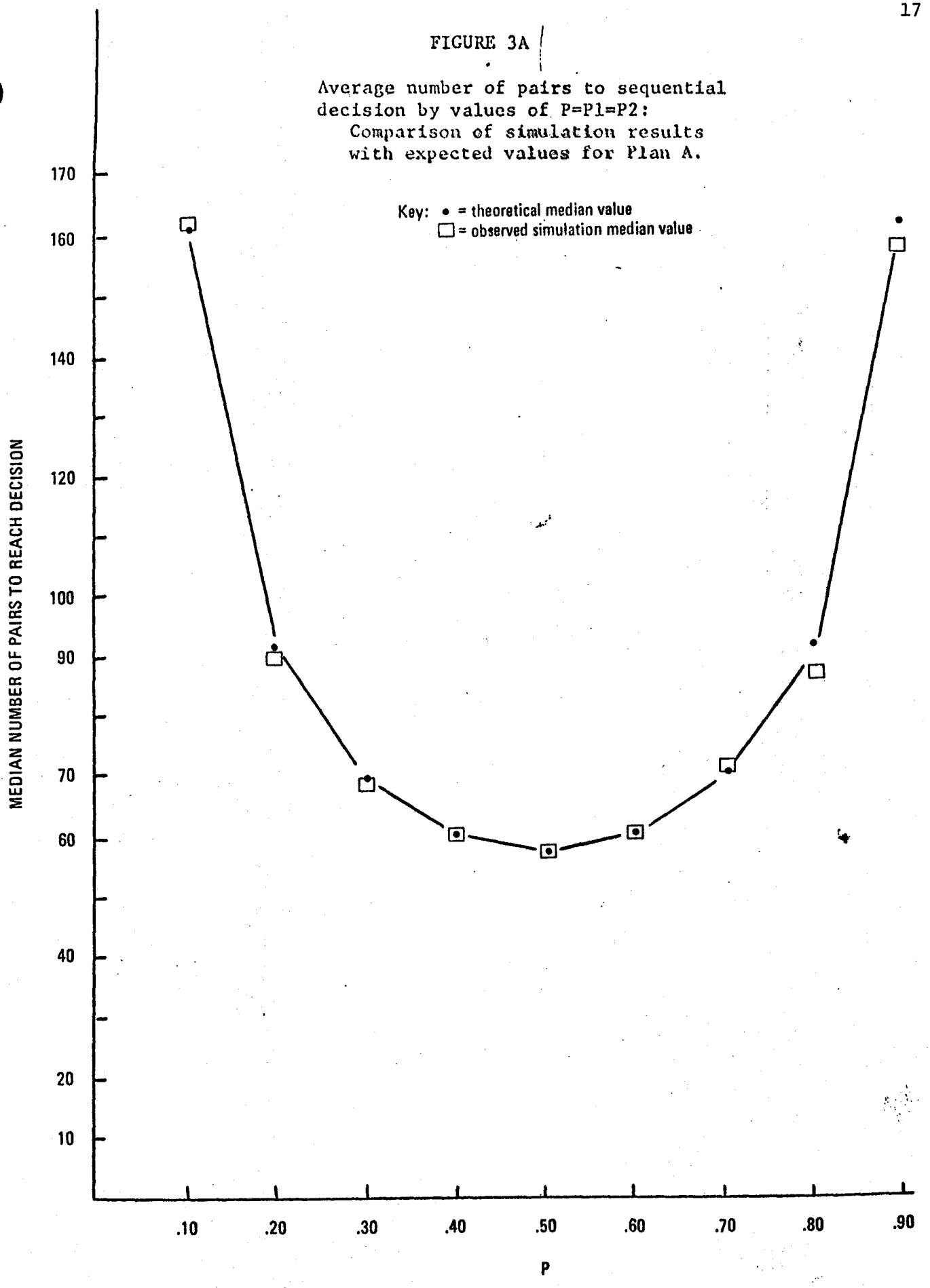
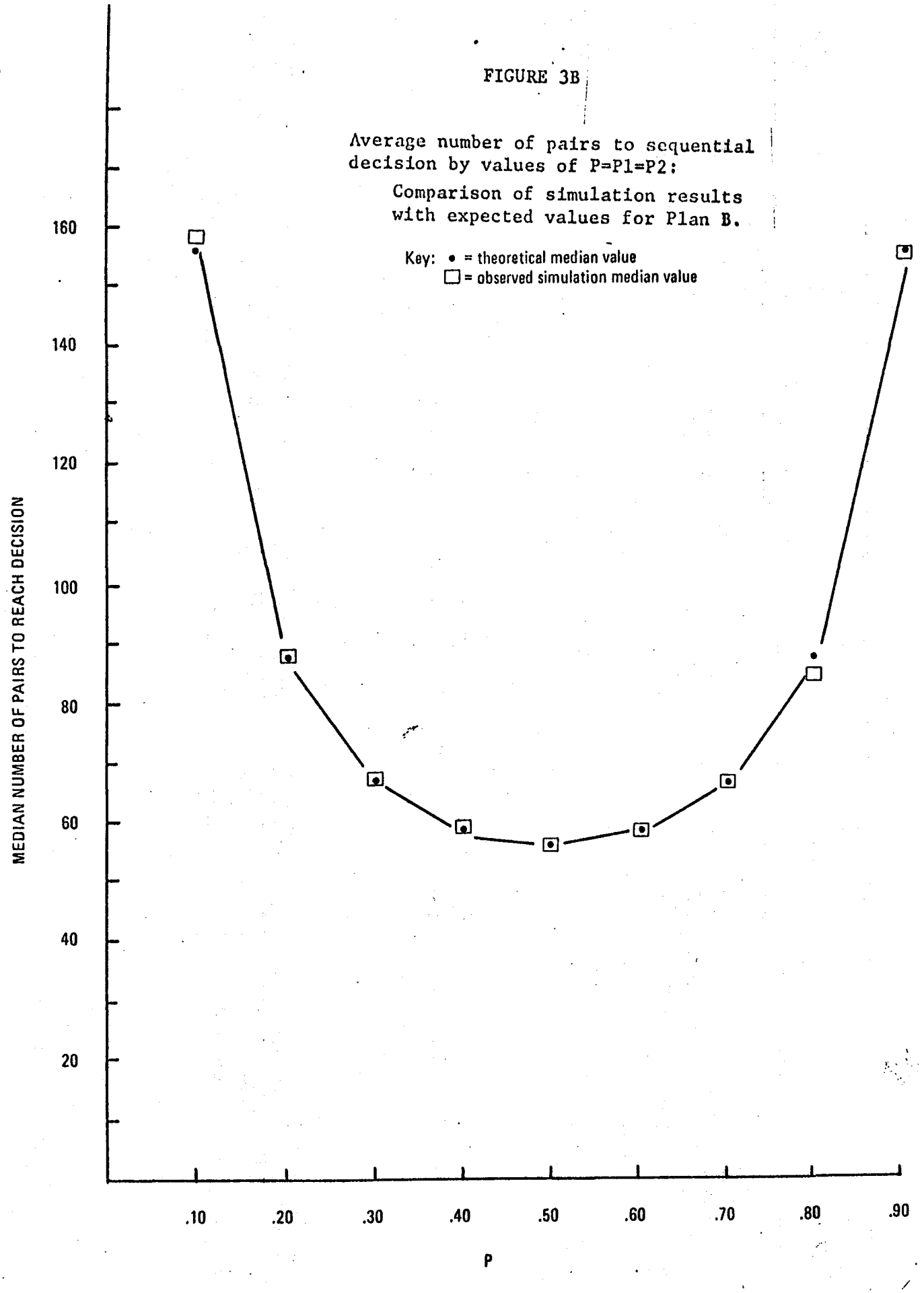


FIGURE 3B

Average number of pairs to sequential decision by values of P=P1=P2:

Comparison of simulation results with expected values for Plan B.



useful to those statisticians involved in planning experiments and estimating their costs.

Acknowledgements

The author would like to give special thanks to the members of the Statistical Services Division, Burroughs Wellcome Co. for their help, and especially to Hale Sweeny for his original proposal of this topic. In addition, the author would like to thank the Burroughs Wellcome Co. for its kind donation of the computer resources to accomplish this project.

Bibliography

[1] Armitage, P., Sequential Medical Trials, Blackwell Scientific Publications, Oxford, 1960.

[2] Bross, I., Sequential Medical Plans, Biometrics 8 (1952), pp 188-205.