Some Sequential Procedures for Selecting the Better Bernoulli Treatment by Using a Matched Samples Design

AJIT C. TAMHANE*

The problem of selecting the better Bernoulli treatment by using a matched samples design is considered in the framework of the indifference-zone approach. In Tamhane (1980) a fixed sample procedure (FSP) for this problem was proposed. Here three sequential procedures are considered: (a) a curtailed sampling procedure (CSP), (b) a procedure based on the Wald sequential probability ratio test (SPRT), and (c) a procedure based on the 2-SPRT proposed by Lorden (1976). Comparisons are made between these procedures based on their expected *total* (tied and untied) sample sizes. It is pointed out that the CSP is the only closed procedure (in terms of the total number of observations) among the three. In addition, for any parameter configuration, the CSP is at least as efficient as the FSP, whereas the SPRT and 2-SPRT can be less efficient than the FSP when success probabilities of the treatments are close. Thus the CSP is a useful practical procedure.

KEY WORDS: Ranking; Selection; Indifference-zone approach; Curtailed sampling; Sequential probability ratio test; 2-Sequential probability ratio test.

1. INTRODUCTION

Suppose that we have two competing medical treatments and we wish to design a clinical trial to select the more effective treatment. Further suppose that the response of a patient to either treatment can be simply classified as a "success" or a "failure" (Bernoulli outcomes). The probabilities of success associated with the two treatments are assumed to be fixed throughout the trial. The treatment having the higher success probability is referred to as the "better" (more effective) treatment.

In a previous article (Tamhane 1980), a fixed sample procedure (FSP) for selecting the better treatment by using a matched samples design was proposed. (For a comparison of a matched samples design with an independent samples design, see McKinlay 1977.) In the present article, it is assumed that the patients enter the trial sequentially. Three selection procedures using a matched samples design in the sequential setting are considered: (a) a curtailed sampling procedure (CSP), (b) a sequential procedure based on the Wald (1947) sequential probability ratio test (SPRT), and (c) a sequential procedure based on Lorden's (1976) 2-SPRT, which approximately solves the Kiefer–Weiss (1957) problem of minimizing the maximum expected sample size. These sequential procedures are shown to guarantee a specified requirement on the probability of a correct selection. The performances of these three procedures are compared in relation to the FSP based on their expected total sample sizes.

A number of closed sequential procedures have been proposed in the literature to remedy some of the shortcomings of the Wald SPRT resulting from its open-ended nature (e.g., see Armitage 1975). The 2-SPRT is included as a representative of such closed procedures. It should be noted, however, that all of these procedures are really open in the present context because of the possibility of tied observations, which they ignore in making stopping and terminal decisions; they are closed only in terms of untied observations, and the SPRT is always open. Thus the CSP is the only truly closed sequential procedure among the ones considered here.

There are some difficulties in implementing a matched samples design in the sequential setting, particularly in applications involving human or animal subjects. So that sampling can be done sequentially, it is clearly necessary that the responses to the treatments be relatively quick; but equally important, it is also necessary that both of the responses in a matched pair be obtained almost simultaneously. This latter requirement precludes, in many cases, using two different subjects who are matched on relevant attributes or using the same subject on two different occasions as a matched pair. Still, in cases in which experimental units come naturally in pairs (e.g., eyes or limbs of the same person) and the treatments do not interact, the use of the sequential matched samples design seems appropriate. An example of such a design is provided by Fertig et al. (1964), who reported a clinical trial in which two topical anesthetic drugs were compared by applying them simultaneously to the oral mucous membrane on the two sides of each patient's mouth, the drugs being allocated randomly to the two sides. After 4 minutes each area was tested with a probe for the presence of topical mucosal anesthesia, and the outcome was recorded as a success or failure. In Section 6, their data will be used to demonstrate the three sequential procedures. (For additional references to sequential matched samples designs with Bernoulli outcomes, see Armitage 1975, p. 89.) Although a hypothesis-testing approach is employed in these applications, the underlying objective in many cases is *selection* of the better treatment, to which goal the present article is devoted.

2. FORMULATION OF THE PROBABILITY REQUIREMENT

To formulate a precise probability requirement, let us first introduce some notation that is similar to that in Tamhane (1980).

* Ajit C. Tamhane is Associate Professor, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60201-9990. Part of this research was done while the author was visiting Cornell University on sabbatical leave during the 1982–1983 academic year on partial support from the U.S. Army Research Office, Durham Contract DAAG-29-81-K-0168. The author wishes to thank Robert Bechhofer and Lionel Weiss for helpful discussions and the editor and the referees for suggestions for improvement.

Journal of the American Statistical Association
June 1985, Vol. 80, No. 390, Theory and Methods

Let T_1 and T_2 denote the two treatments, and let π_{ij} denote the probability that a matched observation results in outcome i on T_1 and outcome j on T_2 $(i, j = 0, 1)$, where 1 denotes a success and 0 denotes a failure; $\sum\sum \pi_{ij} = 1$. Let $p_1 = \pi_{11} + \pi_{10}$ and $p_2 = \pi_{11} + \pi_{01}$ be the success probabilities of T_1 and T_2, respectively, and let $p_{[1]} \le p_{[2]}$ denote the ordered p_i. The π_{ij} are assumed to be unknown, but it is assumed that the experimenter is able to specify an upper bound π^* $(0 < \pi^* \le 1)$ on the probability $\pi = \pi_{10} + \pi_{01}$ of an untied observation. (We can regard π as a measure of how poorly the pairs are matched. If the experimenter thinks that matching is good, then he or she will specify a small π^* and vice versa.)

It is desired to select the treatment associated with $p_{[2]}$, the better treatment. Selection of the better treatment is referred to as a *correct selection*. Attention is restricted to procedures that guarantee the following probability requirement:

$$\Pr(\text{correct selection}) \ge P^* \quad \text{whenever} \quad p_{[2]} - p_{[1]} = \delta \ge \delta^* \quad \text{and} \quad \pi_{10} + \pi_{01} = \pi \le \pi^*, \tag{2.1}$$

where $\{\pi^*, \delta^*, P^*\}$ are preassigned constants, $0 < \pi^* \le 1$, $0 < \delta^* \le \pi^*$, and $\frac{1}{2} < P^* < 1$.

3. CURTAILED SAMPLING PROCEDURE

3.1 Description of the Procedure

Let $X_{ij}^{(m)}$ denote the number of trials that result in outcome i on T_1 and outcome j on T_2 $(i, j = 0, 1)$ when $m \ge 1$ matched observations have been taken on the two treatments. In the following, let n denote the fixed sample size required by the FSP to guarantee (2.1) [a table of n for selected (π^*, δ^*) and $P^* = .90$ and .95 is given in Tamhane 1980]. The CSP operates as follows: Continue taking one matched observation at a time on T_1 and T_2 until the first $m \le n - 1$ for which

$$|X_{10}^{(m)} - X_{01}^{(m)}| \ge n - m \tag{3.1}$$

holds. Stop sampling and select the treatment associated with $\max[X_{10}^{(m)}, X_{01}^{(m)}]$. If (3.1) does not hold for any $m \le n - 1$, then stop sampling with the nth observation and make the selection as in the FSP; that is, select T_1 (T_2) if $X_{10}^{(n)} > X_{01}^{(n)}$ $[X_{10}^{(n)} < X_{01}^{(n)}]$ and break the tie at random if $X_{10}^{(n)} = X_{01}^{(n)}$.

3.2 Probability of a Correct Selection

The main result of this section is summarized in the following theorem.

Theorem 1. For any parameter values π_{ij} $(i, j = 0, 1)$, we have

$$\Pr(\text{correct selection} \mid \text{CSP}) = \Pr(\text{correct selection} \mid \text{FSP}) \tag{3.2}$$

when both the CSP and FSP are based on the same n. In particular, the CSP guarantees the probability requirement (2.1) if and only if the FSP does.

Proof. The proof follows from theorem 5.1 of Bechhofer and Kulkarni (1982).

Remark 1. The first draft of this article contained a proof of Theorem 1 based on random walk arguments. Later, Bechhofer and Kulkarni (1982) and Jennison (1983) generalized the result of Theorem 1 to more general settings (more than two treatments and general sampling strategies with the only restriction that no more than n observations be taken on any treatment).

Remark 2. This theorem shows that the CSP attains the *same* probability of a correct selection as does the FSP at *every* parameter configuration when both are based on the same sample size n. The sample size n may be selected without any reference to the indifference-zone probability requirement (2.1) (e.g., n may be selected based on economic considerations). The theorem enables us to calculate the *exact* probability of a correct selection attained by the CSP at any parameter configuration by using the corresponding formula [see eqs. (3.1) and (3.2) of Tamhane 1980] for the FSP.

3.3 Expected Sample Size

Let N denote the number of steps required until termination, and $E(N \mid \text{CSP})$, the expected sample size required by the CSP based on sample size $= n$. We see that the upper limit on N is n and the lower limit on N is n_0, where $n_0 = n/2$ if n is even and $n_0 = (n + 1)/2$ if n is odd. Thus

$$1 \le n/[E(N \mid \text{CSP})] \le 2. \tag{3.3}$$

The quantity $E(N \mid \text{CSP})$ can be evaluated by the usual method of solving recursively a system of difference equations obtained by considering the stopping times of the random walk $Y_m = X_{10}^{(m)} - X_{01}^{(m)}$; here $Y_0 \equiv 0$ and at each step Y_m goes up one unit with probability π_{10}, goes down one unit with probability π_{01}, and stays unchanged with probability $1 - \pi$. A direct approach is employed here, which exploits the fact that because of the unit slopes of the stopping boundaries (3.1), once the random walk hits a boundary, it must either stay on it or exit outside. Thus the calculation of the first passage-time probabilities [given by (3.6) below] is greatly simplified.

Let u_m denote the probability that the random walk Y_m stops at the mth step with the stopping boundaries given by (3.1). Then

$$E(N \mid \text{CSP}) = \sum_{m=n_0}^{n} m u_m. \tag{3.4}$$

To obtain an expression for u_m, define

$$v_{m,y} = \Pr(Y_m = y) = \sum (m!/i!j!k!)\pi_{10}^i \pi_{01}^j (1 - \pi)^k, \tag{3.5}$$

where the sum extends over all $i, j, k \ge 0$, $i + j + k = m$, and $i - j = y$; that is, $i = (m + y - k)/2$, $j = (m - y - k)/2$, and the sum extends over admissible values of k. Then for $n_0 \le m \le n - 1$, one has

$$u_m = (1 - \pi_{01})v_{m-1,n-m} + \pi_{10}v_{m-1,n-m-1} + (1 - \pi_{10})v_{m-1,-n+m} + \pi_{01}v_{m-1,-n+m+1}, \tag{3.6}$$

and

$$u_n = 1 - \sum_{m=n_0}^{n-1} u_m. \tag{3.7}$$

Substituting (3.5) in (3.6) and making some simplification, one obtains for $n_0 \le m \le n - 1$,

$$u_m = \sum_{k=0}^{2m-n} \left\{ (1 - I_{k,n}) \cdot \frac{(m-1)!}{\left(\frac{n-k-1}{2}\right)!\left(m - \frac{n+k+1}{2}\right)!k!} \cdot (1-\pi)^k \cdot [(1-\pi_{01})\pi_{10}^{(n-k-1)/2}\pi_{01}^{m-[(n+k+1)/2]} + (1-\pi_{10})\pi_{10}^{m-[(n+k+1)/2]}\pi_{01}^{(n-k-1)/2}] + I_{k,n} \cdot \frac{(m-1)!}{\left(\frac{n-k-2}{2}\right)!\left(m - \frac{n+k}{2}\right)!k!} \cdot (1-\pi)^k \cdot [\pi_{10}^{(n-k)/2}\pi_{01}^{m-[(n+k)/2]} + \pi_{10}^{m-[(n+k)/2]}\pi_{01}^{(n-k)/2}] \right\}, \quad (3.8)$$

where $I_{k,n} = 1$ if k and n are both odd or both even and 0 otherwise. Thus (3.7) and (3.8) give the exact distribution of N, and these together with (3.4) give an exact formula for $E(N \mid \text{CSP})$. It should be noted that (3.8) simplifies when one or more of the π_{ij} are zero. Thus when $\pi_{10} = 1$ or $\pi_{01} = 1$, one has $u_{n_0} = 1$ and $u_i = 0$ for $i \neq n_0$; when $\pi_{10} = \pi_{01} = 0$, one has $u_n = 1$ and $u_i = 0$ for $i \neq n$; when $\pi = 1$, one has for $n_0 \le m \le n - 1$

$$u_m = \frac{(m-1)!}{\left(\frac{n-1}{2}\right)!\left(m - \frac{n+1}{2}\right)!} \times [\pi_{10}^{(n+1)/2}\pi_{01}^{m-[(n+1)/2]} + \pi_{10}^{m-[(n+1)/2]}\pi_{01}^{(n+1)/2}] \quad \text{if } n \text{ is odd,}$$

$$= \frac{(m-1)!}{\left(\frac{n}{2}-1\right)!\left(m - \frac{n}{2}\right)!} [\pi_{10}^{n/2}\pi_{01}^{m-(n/2)} + \pi_{10}^{m-(n/2)}\pi_{01}^{n/2}] \quad \text{if } n \text{ is even;} \quad (3.9)$$

and finally, when $\pi_{01} = 0$, $\pi_{10} > 0$, and $1 - \pi > 0$, one has for $n_0 \le m \le n - 1$

$$u_m = \frac{(m-1)!}{(n-m)!(2m-n-1)!}(1-\pi)^{2m-n-1}\pi_{10}^{n-m} + \frac{(m-1)!}{(n-m-1)!(2m-n)!}(1-\pi)^{2m-n}\pi_{10}^{n-m} \quad (3.10)$$

3.4 Wiener Process Approximation to the Expected Sample Size

For large n, the random walk Y_i can be approximated by a Wiener process, and expressions derived by Anderson (1960) can be used as approximations to $E(N \mid \text{CSP})$. From equation (5.13) of Anderson (1960) one obtains for large n

$$E(N \mid \text{CSP}) \cong \left(\frac{c}{d-\mu}\right)\phi(\mu\sqrt{n}) \sum_{i=0}^{\infty} (-1)^i(2i+1) \times \left[R\left(\frac{2(i+1)c + \mu n}{\sqrt{n}}\right) - R\left(\frac{2ic - \mu n}{\sqrt{n}}\right)\right] + \text{an analogous expression with } \mu \text{ replaced by } -\mu, \quad (3.11)$$

where $\mu = \delta/\sigma$, $\delta = \pi_{10} - \pi_{01}$, $\sigma^2 = \pi_{10} + \pi_{01} - (\pi_{10} - \pi_{01})^2 = \pi - \delta^2$, $c = n/\sigma$, $d = -1/\sigma$, $R(x) = [1 - \Phi(x)]/\phi(x)$ is the Mill's ratio, and $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal distribution and density functions, respectively.

The series (3.11) converges very slowly, but for large n, even with up to 200 terms in the series, it is much cheaper to compute than the exact formula and gives accurate results (but a bit on the low side). For example, for $\delta = \delta^* = .1$, $\pi = \pi^* = .1$, $P^* = .90$, and $n = 16$, the exact value of $E(N \mid \text{CSP})$ is 14.628 and the Wiener approximation is 14.443, giving a 1.26% error. For $\delta = \delta^* = .1$, $\pi = \pi^* = .9$, $P^* = .90$, and $n = 147$, the exact value of $E(N \mid \text{CSP})$ is 132.962 and the Wiener approximation is 132.557, giving a .305% error.

4. PROCEDURE BASED ON THE SEQUENTIAL PROBABILITY RATIO TEST

4.1 Description of the Procedure

This procedure operates as follows: Continue taking one matched observation at a time on T_1 and T_2 until the first m for which

$$|X_{10}^{(m)} - X_{01}^{(m)}| \ge d^* \quad (4.1)$$

is satisfied, where

$$d^* = \text{the smallest integer} \ge \frac{\log_e\left(\frac{P^*}{1-P^*}\right)}{\log_e\left(\frac{\pi^* + \delta^*}{\pi^* - \delta^*}\right)} \quad (4.2)$$

(For $P^* < \frac{1}{2} + \delta^*/2\pi^*$, in particular for $\delta^* = \pi^*$, define $d^* = 1$.) Stop sampling when (4.1) is satisfied and select T_1 (T_2) if $X_{10}^{(m)} > X_{01}^{(m)}$ [$X_{10}^{(m)} < X_{01}^{(m)}$].

This procedure is related to the Wald SPRT for an underlying hypothesis-testing problem. Consider a trinomial distribution having cells (1, 0), (0, 1), and (1, 1) ∪ (0, 0), with respective cell probabilities of π_{10}, π_{01}, and $1 - \pi$. Then the foregoing procedure is exactly the Wald SPRT for the symmetric test of H_1: $\pi_{10} = (\pi^* + \delta^*)/2$ and $\pi_{01} = (\pi^* - \delta^*)/2$ versus H_2: $\pi_{10} = (\pi^* - \delta^*)/2$ and $\pi_{01} = (\pi^* + \delta^*)/2$, with the two error probabilities equal to $1 - P^*$ and a direct correspondence between accepting H_i and selecting T_i ($i = 1, 2$). This procedure will be referred to as the SPRT.

4.2 Probability of a Correct Selection

Theorem 2. With the choice of d^* given in (4.2), the SPRT guarantees the probability requirement (2.1).

Proof. Using elementary random walk methods for finding the probability of ruin in the classical ruin problem (e.g., see Feller 1968, chap. 14, sec. 2), it can be shown that for any values of the π_{ij},

$$\Pr(\text{correct selection} \mid \text{SPRT}) = 1 \Big/ \left[1 + \left(\frac{\pi - \delta}{\pi + \delta}\right)^{d^*}\right] \quad (4.3)$$

Now for $\delta > 0$, (4.3) is strictly decreasing in π; and for $\pi > \delta$, (4.3) is strictly increasing in δ. Therefore the infimum of (4.3) over $\pi \le \pi^*$ and $\delta \ge \delta^*$ is achieved at $\pi = \pi^*$ and $\delta = \delta^*$ (the so-called least favorable configuration; LFC). It is straightforward to check that this infimum $\ge P^*$, and in fact any $d \ge d^*$, will also guarantee (2.1).

4.3 Expected Sample Size

Again using elementary random walk methods for finding the expected duration of the game in the classical ruin problem, which involves solving a system of difference equations (Feller 1968, chap. 14, sec. 4), we can derive the formula

$$E(N \mid \text{SPRT}) = \frac{d^*}{\delta}\left\{\frac{1 - \left(\frac{\pi - \delta}{\pi + \delta}\right)^{d^*}}{1 + \left(\frac{\pi - \delta}{\pi + \delta}\right)^{d^*}}\right\} \quad \text{for } \delta > 0$$

$$= \frac{(d^*)^2}{\pi} \quad \text{for } \delta = 0. \qquad (4.4)$$

5. PROCEDURE BASED ON THE 2-SPRT

5.1 Description of the Procedure

Kiefer and Weiss (1957) considered the problem of finding an optimal test that, for a simple versus a simple hypothesis-testing problem with given error probabilities, minimizes the expected total sample size at a third parameter point. Under certain conditions, they characterized such a test as a Bayes sequential test having convergent nonlinear boundaries; this test is a member of the class of generalized SPRT's (GSPRT's) introduced by Weiss (1953). In the symmetric normal and Bernoulli testing problems, this also solves the problem of minimizing the maximum expected total sample size.

For the problem of testing hypotheses on a Bernoulli parameter, Weiss (1962) and Freeman and Weiss (1964) showed how the boundaries of the GSPRT can be computed by using the backward induction method. This algorithm, however, involves trial and error as well as considerable computations. Lorden (1976) proposed a procedure (referred to as the 2-SPRT) that is much simpler to implement and approximately solves the Kiefer–Weiss problem. Lorden's 2-SPRT will now be described in the context of this article's problem.

Take one matched observation at a time on T_1 and T_2 but consider only the untied observations, that is, the observations that result in outcome (1, 0) or (0, 1). Let $Z_i = 1$ (0) if the ith untied outcome is (1, 0) [(0, 1)], and let $S_m = \sum_{i=1}^{m} Z_i$ be the number of successes on T_1 among the first m untied observations. The Z_i are iid Bernoulli random variables with parameter $\theta = \pi_{10}/\pi = \frac{1}{2} + \delta/2\pi$. Consider the symmetric test of H_1: $\theta = \frac{1}{2} + \Delta^*$ versus H_2: $\theta = \frac{1}{2} - \Delta^*$ (where $\Delta^* = \delta^*/2\pi^*$) with the two error probabilities equal to $1 - P^*$. For convenience of notation, let $\theta_1 = \frac{1}{2} + \Delta^*$, $\theta_2 = \frac{1}{2} - \Delta^*$, and $\theta_0 = (\theta_1 + \theta_2)/2 = \frac{1}{2}$. The 2-SPRT for this problem operates as follows: Stop after $m < M$ [given by (5.3)] untied observations and accept H_2 (select T_2) if

$$\left(\frac{\theta_0}{\theta_1}\right)^{S_m}\left(\frac{1 - \theta_0}{1 - \theta_1}\right)^{m - S_m} \geq \frac{1}{2(1 - P^*)},$$

that is, if

$$S_m \leq \frac{m \log_e\left(\frac{1}{1 - 2\Delta^*}\right) + \log_e 2(1 - P^*)}{\log_e\left(\frac{1 + 2\Delta^*}{1 - 2\Delta^*}\right)}; \qquad (5.1)$$

accept H_1 (select T_1) if

$$\left(\frac{\theta_0}{\theta_2}\right)^{S_m}\left(\frac{1 - \theta_0}{1 - \theta_2}\right)^{m - S_m} \geq \frac{1}{2(1 - P^*)},$$

that is, if

$$S_m \geq \frac{m \log_e(1 + 2\Delta^*) - \log_e 2(1 - P^*)}{\log_e\left(\frac{1 + 2\Delta^*}{1 - 2\Delta^*}\right)}. \qquad (5.2)$$

If both (5.1) and (5.2) are not satisfied, then continue sampling until the number of untied observations is M, where

M = the smallest integer

$$\geqq (2 \log_e[2(1 - P^*)])/[\log_e(1 - 4\Delta^{*2})]; \qquad (5.3)$$

at $m = M$ accept H_1 (accept H_2) if $S_M > M/2$ ($S_M < M/2$), and randomize with equal probability between the two decisions if $S_M = M/2$.

5.2 Probability of a Correct Selection

From Lorden (1976) it is known that this 2-SPRT controls the error probabilities at or below $1 - P^*$ at H_1 and H_2. The following theorem can now be stated.

Theorem 3. The 2-SPRT guarantees the probability requirement (2.1).

Proof. The Z_i ($1 \leq i \leq m$) admit a scalar sufficient statistic S_m for every $m \geq 1$ such that the distribution of S_m has the monotone likelihood ratio property in θ. Thus the result of J. Ghosh (1960) applies, and we can conclude that the operating characteristic (OC) function,

$$\Pr_\theta\{\text{accept } H_1 \mid \text{2-SPRT}\} = \Pr_\theta\{\text{select } T_1 \mid \text{2-SPRT}\},$$

is nondecreasing in θ. But $\theta = \frac{1}{2} + \delta/2\pi$ is increasing in δ and decreasing in π. Therefore, since $\Pr_\theta\{\text{select } T_1 \mid \text{2-SPRT}\} \geq P^*$ at H_1: $\theta = \theta_1$, where $\delta = \delta^*$ and $\pi = \pi^*$, the same holds for $\delta \geq \delta^*$ and $\pi \leq \pi^*$.

5.3 Expected Sample Size

The expected number of untied observations required by the 2-SPRT, $E(N_0 \mid \text{2-SPRT})$, can be evaluated by solving recursively a system of difference equations (the details are omitted). The expected total (tied and untied) sample size required by the 2-SPRT can then be obtained by using the formula

$$E(N \mid \text{2-SPRT}) = [E(N_0 \mid \text{2-SPRT})]/\pi. \qquad (5.4)$$

This method was used in the computer programs for making numerical comparisons between the 2-SPRT and its two sequential competitors (this numerical study is discussed in Section 7).

6. AN EXAMPLE

Fertig et al. (1964) reported sequentially collected data on the presence (1) or absence (0) of topical mucosal anesthesia by using two topical anesthetic drugs, A and B, on the oral mucous membrane of the two sides of each patient's mouth.

Table 1. Presence (1) or Absence (0) of Topical Mucosal Anesthesia

Patient (m)	Drug A	Drug B	$X_{10}^{(m)} - X_{01}^{(m)}$	Patient (m)	Drug A	Drug B	$X_{10}^{(m)} - X_{01}^{(m)}$
1	0	0	0	24	1	1	5
2	0	0	0	25	1	1	5
3	0	1	−1	26	0	1	4
4	1	0	0	27	1	1	4
5	0	0	0	28	1	1	4
6	0	0	0	29	1	0	5
7	0	0	0	30	1	1	5
8	0	0	0	31	0	0	5
9	1	0	1	32	0	1	4
10	1	0	2	33	1	1	4
11	1	0	3	34	1	0	5
12	0	0	3	35	0	0	5
13	0	0	3	36	0	0	5
14	1	1	3	37	1	0	6
15	1	1	3	38	1	0	7
16	1	1	3	39	1	1	7
17	1	1	3	40	1	1	7
18	1	1	3	41	1	0	8
19	1	1	3	42	1	0	9
20	1	0	4	43	1	1	9
21	0	0	4	44	0	0	9
22	0	0	4	45	1	0	10
23	1	0	5				

Source: Fertig et al. (1964).

The data were collected by Kutscher of Columbia University. Out of the 47 patients observed, 2 patients had missing data on drug B. Only the data on the remaining 45 patients is considered here, and it is presented in Table 1.

For illustration purposes only, suppose that these data were collected for the FSP with $n = 45$. From table 1 in Tamhane (1980), we see that $n = 45$ corresponds to $\delta^* = .2$, $\pi^* = .7$, and $P^* = .95$. [There are other triples (δ^*, π^*, P^*) that will also correspond to $n = 45$.] These values shall be used to calculate the design constants of the SPRT and the 2-SPRT.

First, let us apply the CSP with $n = 45$ to the data in Table 1. It is seen that for $m = 38$, (3.1) is satisfied for the first time. Thus sampling could have been terminated after the 38th patient, and drug A could have been selected as the more effective anesthetic. This would guarantee the probability requirement (2.1) with (δ^*, π^*, P^*) given above.

Next let us apply the SPRT. Substituting $\delta^* = .2$, $\pi^* = .7$, and $P^* = .95$ in (4.2) we get $d^* = 6$. Thus the SPRT terminates with the 37th patient with the same decision as above.

Finally, to apply the 2-SPRT, calculate the boundaries (5.1) and (5.2) for S_m—the number of times drug A *beats* drug B—where m is now the index of *untied* responses. The lower boundary is $.5725m - 3.9174$, and the upper boundary is $.4276m + 3.9174$. The cutoff limit for the number of untied responses is $M = 55$ from (5.3). It can be readily checked that S_m first exits the upper boundary for the 38th patient ($m = 13$th untied response) when the upper boundary equals 9.4762 and $S_m = 10$.

Thus in this particular example, all three procedures terminate at about the same point and reach the same decision. The average performances of these procedures, however, can be markedly different depending on the underlying true δ and π values. The following section carries out these performance comparisons.

7. PERFORMANCE COMPARISONS

In this section, the performances of the three sequential procedures are compared in terms of their $E(N)$ values when they all guarantee the same probability requirement (2.1). To facilitate the comparison, take the sample size n needed by the FSP to guarantee (2.1) as the benchmark and define the relative efficiencies (RE's) of the sequential procedures with respect to the FSP as follows:

$$\mathrm{RE}_{\mathrm{CSP}} = \frac{n}{E(N \mid \mathrm{CSP})}, \qquad \mathrm{RE}_{\mathrm{SPRT}} = \frac{n}{E(N \mid \mathrm{SPRT})},$$

$$\mathrm{RE}_{\text{2-SPRT}} = \frac{n}{E(N \mid \text{2-SPRT})}. \tag{7.1}$$

Note that RE values greater than unity favor the corresponding sequential procedures over the FSP. The n values needed in the RE computations were taken from table 1 of Tamhane (1980).

Computations were carried out for $P^* = .90$; $\delta^* = .1, .3$; and $\pi^* = .5, .7, .9$. Analogous computations were made for $P^* = .95$, but the results were similar and hence are not reported here. For each value of δ^*, three values of δ were considered: $\delta = 0$, δ^*, and π^*. For each value of π^*, three values of π were considered: $\pi = \max(\delta, \pi^*/2)$ (denoted by θ^* in Table 2), π^*, and 1. The case $\pi = 1$ deals with the possibility that the experimenter's specification of the upper bound π^* on π is actually violated by the true π. The preceding nine combinations cover the possible range of (δ, π) values. The results of these computations are presented in Table 2.

Inspection of the results shows that $\mathrm{RE}_{\mathrm{SPRT}}$ and $\mathrm{RE}_{\text{2-SPRT}}$ are higher than $\mathrm{RE}_{\mathrm{CSP}}$ in almost all cases except when $\delta = 0$ and $\pi = \theta^* < \pi^*$. For favorable configurations (large values of δ and π), the $\mathrm{RE}_{\mathrm{SPRT}}$ values are the highest (higher than those of $\mathrm{RE}_{\text{2-SPRT}}$), whereas for unfavorable configurations (small values of δ and π), the $\mathrm{RE}_{\mathrm{SPRT}}$ values are the lowest and the $\mathrm{RE}_{\text{2-SPRT}}$ values are slightly higher. Thus in terms of the expected sample sizes, the SPRT and 2-SPRT are preferred over the CSP for favorable parameter configurations, with the SPRT gaining preference for more favorable configurations and the 2-SPRT gaining preference for somewhat less favorable configurations. But when δ is close to zero, $\mathrm{RE}_{\mathrm{SPRT}}$ and $\mathrm{RE}_{\text{2-SPRT}}$ are both less than unity, and from (3.3) it is known that $1 \leq \mathrm{RE}_{\mathrm{CSP}} \leq 2$ always. Thus for δ close to zero, the SPRT and 2-SPRT are both less efficient than the FSP, with the SPRT being more so. This fact of course is well known for the SPRT (see B. Ghosh 1970, secs. 3.6.1–3.6.2, or Wetherill 1975, p. 23); but for the 2-SPRT (or more generally for GSPRT's), it is not as widely known.

Perhaps more important is the fact that the CSP is a *closed* procedure requiring at most n observations whereas the SPRT and 2-SPRT are both open procedures in the present problem, the latter because of the possibility of ties. Thus although the $E(N)$'s required by the SPRT and 2-SPRT are smaller in favorable configurations, the distributions of the N's are skewed to right (that for the SPRT being more so), resulting in occasionally large values of N, which could lead to bottlenecks in practice. The security afforded by the closed nature of the CSP is very reassuring in such cases, and therefore the CSP is recommended as a useful practical procedure.

Table 2. Relative Efficiency of CSP, SPRT, and 2-SPRT With Respect to FSP for $P^* = .90$

		$\pi^* = .5$			$\pi^* = .7$			$\pi^* = .9$		
δ^*	δ	1	π^*	θ^{*a}	1	π^*	θ^*	1	π^*	θ^*
					CSP					
.1	π^*	1.482	1.494	1.494	1.700	1.694	1.694	1.887	1.888	1.888
	δ^*	1.113	1.105	1.100	1.117	1.106	1.100	1.103	1.106	1.100
	0	1.083	1.063	1.044	1.081	1.063	1.044	1.063	1.063	1.044
.3	π^*	1.376	1.314	1.314	1.701	1.644	1.644	1.900	1.855	1.855
	δ^*	1.261	1.283	1.248	1.377	1.295	1.276	1.347	1.302	1.282
	0	1.194	1.183	1.124	1.291	1.193	1.133	1.244	1.194	1.133
					SPRT					
.1	π^*	6.769	6.750	6.750	9.975	9.975	9.975	3.230	3.230	3.230
	δ^*	2.507	1.610	1.367	2.141	1.742	1.451	1.927	1.824	1.502
	0	2.250	1.125	.563	1.781	1.247	.623	1.470	1.323	.662
.3	π^*	2.813	2.250	2.250	2.831	2.800	2.800	3.600	3.600	3.600
	δ^*	2.453	1.530	1.350	1.644	1.364	1.201	1.420	1.360	1.204
	0	2.250	1.125	.563	1.333	.933	.467	1.000	.900	.450
					2-SPRT					
.1	π^*	5.669	4.500	4.500	7.319	6.650	6.650	9.147	8.820	8.820
	δ^*	2.796	1.623	1.196	2.155	1.665	1.229	1.807	1.685	1.243
	0	2.642	1.321	.661	1.935	1.354	.677	1.524	1.371	.686
.3	π^*	2.385	1.500	1.500	2.446	2.100	2.100	3.017	2.880	2.880
	δ^*	2.254	1.238	.900	1.792	1.370	.964	1.642	1.528	1.109
	0	2.182	1.091	.545	1.631	1.142	.571	1.400	1.260	.630

[a] $\theta^* = \max(\delta, \pi^*/2)$.

Theoretically the SPRT requires the smallest $E(N)$ (and hence has the highest RE) when $\pi = \pi^*$ and $\delta = \delta^*$ among all procedures guaranteeing (2.1) by the optimality result of Wald and Wolfowitz (1948). In the numerical computations, however, the 2-SPRT achieves higher RE at this configuration in some cases—a discrepancy that can be explained by the fact that an excessive overshoot is caused by the upward rounding of d^* [cf. (4.2)]. Some other comments on the numerical results are as follows: For any (δ^*, π^*), if π is fixed then all RE's decrease as δ decreases. This is of course no surprise, since any procedure that samples sequentially would require larger $E(N)$ as the two treatments get closer in their success probabilities. Moreover, in most cases all three RE's increase as π increases for fixed δ. This happens because as π increases, each matched observation is more likely to yield different responses on the two treatments. Since large values of the statistic $|X_{10}^{(m)} - X_{01}^{(m)}|$ are favorable to an early termination of all three sequential procedures, they require smaller $E(N)$'s if π is large.

[*Received September 1981. Revised October 1984.*]

REFERENCES

Anderson, T. W. (1960), "A Modification of the Sequential Probability Ratio Test to Reduce the Sample Size," *Annals of Mathematical Statistics,* 31, 165–197.

Armitage, P. (1975), *Sequential Medical Trials* (2nd ed.), New York: Halsted Press.

Bechhofer, R. E., and Kulkarni, R. V. (1982), "Closed Adaptive Sequential Procedures for Selecting the Best of $k \geq 2$ Bernoulli Populations," in *Statistical Decision Theory and Related Topics* (Vol. 3), eds. S. S. Gupta and J. O. Berger, New York: Academic Press, pp. 61–108.

Feller, W. (1968), *An Introduction to Probability Theory and Its Applications* (Vol. 1; 3rd ed.), New York: John Wiley.

Fertig, J. W., Chilton, N. W., and Varma, A. O. (1964), "Studies in the Design and Analysis of Dental Experiments. 9: Sequential Analysis (Sign Test)," *Journal of Oral and Therapeutical Pharmacology,* 1, 45–56.

Freeman, D., and Weiss, L. (1964), "Sampling Plans Which Approximately Minimize the Maximum Expected Sample Size," *Journal of the American Statistical Association,* 59, 67–88.

Ghosh, B. K. (1970), *Sequential Tests of Statistical Hypotheses,* New York: Addison Wesley.

Ghosh, J. K. (1960), "On the Monotonicity of the OC of a Class of Sequential Probability Ratio Tests," *Calcutta Statistical Association Bulletin,* 9, 139–144.

Jennison, C. (1983), "Equal Probability of Correct Selection for Certain Bernoulli and Multinomial Selection Procedures," *Communications in Statistics,* Ser. A, 12, 2887–2896.

Kiefer, J., and Weiss, L. (1957), "Some Properties of Generalized Probability Ratio Tests," *Annals of Mathematical Statistics,* 28, 57–74.

Lorden, G. (1976), "2-SPRT's and the Modified Kiefer–Weiss Problem of Minimizing an Expected Sample Size," *Annals of Statistics,* 4, 281–291.

McKinlay, S. M. (1977), "Pair-Matching—A Reappraisal of a Popular Technique," *Biometrics,* 33, 725–735.

Tamhane, A. C. (1980), "Selecting the Better Bernoulli Treatment Using a Matched Samples Design," *Journal of the Royal Statistical Society,* Ser. B, 42, 26–30.

Wald, A. (1947), *Sequential Analysis,* New York: John Wiley.

Wald, A., and Wolfowitz, J. (1948), "Optimum Character of the Sequential Probability Ratio Test," *Annals of Mathematical Statistics,* 19, 326–339.

Weiss, L. (1953), "Testing One Simple Hypothesis Against Another," *Annals of Mathematical Statistics,* 24, 273–281.

——— (1962), "On Sequential Tests Which Minimize the Maximum Expected Sample Size," *Journal of the American Statistical Association,* 57, 551–566.

Wetherill, G. B. (1975), *Sequential Methods in Statistics* (2nd ed.), New York: Halsted Press.