The Annals of Statistics
1987, Vol. 15, No. 1, 449–455

EXTENDED-PAULSON SEQUENTIAL SELECTION

BY DON EDWARDS

University of South Carolina

Principles of multiple comparisons with the best treatment are used to construct a sequential procedure for selecting one of k normal populations with equal, known variance. The procedure makes a stronger confidence than traditional Indifference Zone procedures, without (asymptotically) inflating sample sizes. It is shown to be substantially superior in asymptotic sample size behavior to existing procedures making the stronger statement.

1. Introduction. Let θ_i be a location parameter for the probability distribution of independent X_{il}, $i = 1, \ldots, k$, $l = 1, 2, \ldots$. Suppose a large θ_i implies a good treatment, and the experimenter's goal is to select the best treatment, or one very nearly best. Let (k) denote the index of any treatment with largest θ_i, and let $\delta^* > 0$ be a specified constant.

DEFINITION 1.1. A treatment i is *δ^*-best* if $\theta_{(k)} - \theta_i \leq \delta^*$.

Let $\langle k \rangle$ denote the index of a treatment selected according to some rule using the data. Confidence statements at level P^*, $1/k < P^* < 1$, regarding treatment $\langle k \rangle$ have to date come in two varieties:

(1.1) ASSERTION 1.1 (Indifference Zone). "Treatment $\langle k \rangle$ is the δ^*-best treatment, if there is only one δ^*-best treatment."

(1.2) ASSERTION 1.2 (Confidence Bound). "Treatment $\langle k \rangle$ is a δ^*-best treatment."

Assertion 1.2 can also be written "$\theta_{(k)} - \theta_{\langle k \rangle} \leq \delta^*$," hence its name. Any procedure satisfying (1.1) but not (1.2) is not safe for use, since it cannot assert selection of a good treatment when there is more than one good treatment.

Sequential methods for choosing $\langle k \rangle$ have been extensively studied for $X_{il} \sim N(\theta_i, \sigma^2)$, σ^2 known. The problem is not adequately solved even in this simple setting. Procedures which eliminate inferior treatments early in the experiment have been proposed by Paulson (1964), Kao and Lai (1980), Hsu and Edwards (1983) and many others. Some of these [Paulson (1964) and Kao and Lai (1980, Section 3)], referred to here as Indifference Zone procedures, can make assertion (1.1) at P^* confidence, but have not been shown able to make assertion (1.2). Other procedures [Kao and Lai (1980, Section 5) and Hsu and Edwards (1983)] have been developed to satisfy (1.2), but do so at the expense of considerably slower elimination of inferior treatments. These latter procedures will be referred

Received January 1985; revised June 1986.

AMS 1980 *subject classifications*. Primary 62L99; secondary 62F07.

Key words and phrases. Sequential elimination, selection, indifference zone, multiple comparisons with the best.

to as (k)-retaining procedures and will be formally defined in Section 3. [It should be said that the Hsu and Edwards (1983) procedure makes even stronger inference than (1.2).] In Edwards (1985), a procedure which can make assertion (1.2) at P^* confidence was outlined. In Section 2 of this article a slightly different, more general procedure called an extended-Paulson sampling plan is proposed. In Section 3 it is shown to have asymptotic ($P^* \to 1$) sample size properties identical to the Indifference Zone procedures, and superior to the (k)-retaining procedures.

2. Extended-Paulson sequential selection. The principles of multiple comparisons with the best treatment were established by Hsu (1981), who gives Lemma 2.1 below for the fixed-sample-size normal case. The version here can be found with proof as Corollary 1 of Edwards and Hsu (1983). Let $x^+ = \max(x, 0)$.

LEMMA 2.1. *For any fixed* j, $1 \le j \le k$, *let* U_i^j, $i = 1, \ldots, k$, $i \ne j$, *be random variables satisfying*

$$P\left\{\bigcap_{j \ne i}\left[\theta_j - \theta_i \le U_i^j\right]\right\} \ge P^*, \tag{2.1}$$

for some probability measure P. For

$$U_i = \left(\max_{j \ne i} U_i^j\right)^+, \tag{2.2}$$

it will follow that

$$P\left\{\bigcap_{i=1}^{k}\left[\theta_{(k)} - \theta_i \le U_i\right]\right\} \ge P^*.$$

Let $T(\mathbf{X}_i; n) = T(X_{i1}, X_{i2}, \ldots, X_{in})$ be defined for each n, satisfying $T(\mathbf{X}_i + \delta\mathbf{1}; n) = \delta + T(\mathbf{X}_i; n)$ for any δ and $T(\mathbf{X}_i/\sigma; n) = T(\mathbf{X}_i; n)/\sigma$ for any $\sigma > 0$. $T(\mathbf{X}_i; n)$ will often be the arithmetic mean $\bar{X}_i$, but robustness considerations may dictate a different choice. Let $n_I \le n_T$ be experimenter-specified minimum and maximum sample sizes, allowing $n_T = \infty$, and $b(n)$ a boundary function defined for every $n_I \le n \le n_T$. Let $P_{\boldsymbol{\theta}, \sigma}$ denote probability computed under $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ and $\sigma > 0$. The boundaries used herein must satisfy the following probability requirement for each j:

$$P_{\mathbf{0},1}\left\{\bigcap_{i \ne j}\bigcap_{n=n_I}^{n_T}\left[T(\mathbf{X}_i; n) - T(\mathbf{X}_j; n) \ge -b(n)\right]\right\} \ge P^*. \tag{2.3}$$

For $T(\cdot) \equiv \bar{X}$, this is the requirement of Paulson (1964) with $\sigma b(n) = a/n - (\lambda - \delta^*)$, $0 < \lambda < \delta^*$ and $a > 0$. It is the requirement of Kao and Lai (1980, Section 3), who choose $\sigma b(n) = g_n^{-1}[(k-1)/(1-P^*)]$ for $g_n(x)$ defined by their equation (7), using δ for our δ^*. For our extended-Paulson procedure we will require (2.3) and $\sigma b(n_T) < \delta^*$. Typically, $\sigma b(n)$ will decrease in n, and n_T will be the first integer for which $\sigma b(n) < \delta^*$.

LEMMA 2.2. *Let N_i be any stopping times such that $n_I \le N_i \le n_T$ a.s. Let $N_{ij} = \min(N_i, N_j)$. If $b(\cdot)$ satisfies* (2.4) *for each* $1 \le j \le k$, *then*

$$U_i^j = U_i^j(\mathbf{N}) = \min_{n_I \le n \le N_{ij}} \left[T(\mathbf{X}_j; n) - T(\mathbf{X}_k; n) + \delta b(n)\right] \tag{2.4}$$

satisfy (2.11) *with* $P = P_{\theta,\sigma}$ *for each* $1 \le j \le k$, θ, *and* $\sigma > 0$.

The proof is very similar to the proof of Lemma 2.2 of Edwards (1985), and is omitted. Lemma 2.2 shows that the boundary requirement (2.3) is quite strong, allowing a great deal of flexibility in the choice of a sequential sampling plan. This flexibility has not been acknowledged in other papers on this subject.

Having taken n_i observations from treatment i, $i = 1, \dots, k$, we are at sampling point $\mathbf{n} = (n_1, \dots, n_k)$. Let $n_{ij} = \min(n_i, n_j)$, and define $U_i^j(\mathbf{n})$ as in (2.4) with n_{ij} for N_{ij}. Let $U_i(\mathbf{n}) = [\max_{j \ne i} U_i^j(\mathbf{n})]^+$, and stage$(\mathbf{n})$ = $\max(n_1, \dots, n_k)$. Any sampling plan with $n_I \le N_i \le n_T$ a.s. will give $U_i(\mathbf{N})$ to be simultaneous $P^* \times 100$ percent confidence bounds for $\theta_{(k)} - \theta_i$, $i = 1, \dots, k$. The strategy, then, will be to make sure the sampling plan provides *one* $U_i(\mathbf{N}) \le \delta^*$ [one index i such that $U_i^j(\mathbf{N}) \le \delta^*$, $\forall\, j \ne i$], with as little sampling of inferior treatments as possible.

Such a plan is defined below, in two phases, after sampling n_I observations from each treatment. The quantities $U_i^j(\mathbf{n})$ and $U_i(\mathbf{n})$ must be updated with every change in $\mathbf{n}$.

A. *Paulson phase*: At each stage$(\mathbf{n})$ $(= n_I, n_I + 1, \dots)$ let

$$G(\mathbf{n}) = \left\{j: U_i^j(\mathbf{n}) > \delta^*, \forall\, i \ne j\right\}. \tag{2.5}$$

(i) If $|G(\mathbf{n})| > 1$, sample once for each treatment in $G(\mathbf{n})$.

(ii) If $|G(\mathbf{n})| \le 1$, let $\langle k \rangle$ be the index such that

$$U_{\langle k \rangle}(\mathbf{n}) = \min\left\{U_i(\mathbf{n}): n_i = \text{stage}(\mathbf{n})\right\}.$$

If $U_{\langle k \rangle} \le \delta^*$, stop, and redefine $\langle k \rangle$ (if necessary) so that $U_{\langle k \rangle} = \min_i\{U_i(\mathbf{n})\}$. Else, go to B.

B. *Resampling phase*: Define the resampling set

$$R(\mathbf{n}) = \left\{j: U_{\langle k \rangle}^j(\mathbf{n}) > \delta^*\right\}.$$

Let j' be the index such that $U_{j'}(\mathbf{n}) = \min\{U_j(\mathbf{n}),\ j \in R(\mathbf{n})\}$. Sample repeatedly from treatment j' until one of the following occurs (these are hierarchical):

(i) $\min_i\{U_i(\mathbf{n})\} \le \delta^*$. If so, stop, and redefine $\langle k \rangle$ (if necessary) so that $U_{\langle k \rangle} = \min_i\{U_i(\mathbf{n})\}$.

(ii) j' drops out of $R(\mathbf{n})$. If so, revise $R(\mathbf{n})$ and j' and resume the resampling phase.

(iii) $n_{j'} = n_{\langle k \rangle} = \text{stage}(\mathbf{n})$. If so, sample one observation at a time from both the treatments j' and $\langle k \rangle$ until (i) or (ii) occurs, or $U_j^{\langle k \rangle}(\mathbf{n}) \le \delta^*$. In the last case, redefine $\langle k \rangle = j'$ and go to B.

REMARK. The Paulson phase (which will be executed only once) can be reexpressed as follows: at each n = stage($\mathbf{n}$), sample from treatments not previously eliminated, eliminating treatment j if $\max_{i \neq j}\{T(\mathbf{X}_i; n)\} - T(\mathbf{X}_j; n) \geq \sigma b(n) - \delta^*$ ($=$"$U_i^j \leq \delta^*$ for some $i \neq j$"). The maximum is taken over uneliminated treatments. If the *same* treatment has largest $T(\cdot)$ at each elimination point, this one will be treatment $\langle k \rangle$ after the Paulson phase, and

$$U_{\langle k \rangle}(\mathbf{N}) = \left[\max_{j \neq \langle k \rangle} U^j_{\langle k \rangle}(\mathbf{N})\right]^+ \leq \delta^*.$$

In this case, the resampling phase is not necessary. If, however, there are some treatments j such that $U^j_{\langle k \rangle}(\mathbf{n}) > \delta^*$ [i.e., $R(\mathbf{n}) \neq \phi$] resampling from these is prescribed. It is intuitive to sample from the better ones first [those with small $U_i(\mathbf{n})$, recalling $\theta_{(k)} - \theta_i \leq U_i(\mathbf{n})$]. The resampling phase above creates a kind of "duel" situation, with the winner to be the new $\langle k \rangle$ candidate.

With $T(\cdot) = \bar{X}$, the Paulson phase reduces to the Paulson (1964) or the Kao and Lai (1980, Section 3) procedure, depending on choice of $b(\cdot)$. The above discussion then shows that these procedures can make the confidence bound assertion (1.2) if the same treatment eliminates all other treatments.

THEOREM 2.1. *If $b(\cdot)$ satisfies* (2.3) *and $\sigma b(n_T) < \delta^*$, the extended-Paulson sequential procedure gives simultaneous $P^* \times 100$ percent confidence upper bounds $U_i(\mathbf{N})$ for $\theta_{(k)} - \theta_i$, $i = 1, \ldots, k$, and*

$$\min_i U_i(\mathbf{N}) = U_{\langle k \rangle}(\mathbf{N}) \leq \delta^*.$$

PROOF. It is implicit to the procedure that when stage($\mathbf{n}$) $= n_T$ is reached, it will be reached by more than one treatment. Let m denote the index of the treatment with largest $T(\mathbf{X}_i, n_T)$. Then for each other treatment i with $n_i = n_T$,

$$U^i_m(\mathbf{n}) \leq T(\mathbf{X}_i; n_T) - T(\mathbf{X}_m; n_T) + \sigma b(n_T) \leq \sigma b(n_T) \leq \delta^*. \tag{2.6}$$

It follows that there will be no further sampling on any treatment reaching $n = n_T$. Hence, the procedure satisfies $n_I \leq N_i \leq n_T$, $i = 1, \ldots, k$, and so the $U_i(\mathbf{N})$ are simultaneous confidence bounds for $\theta_{(k)} - \theta_i$, $i = 1, \ldots, k$. The procedure stops when $\min_i\{U_i(\mathbf{N})\} \leq \delta^*$, or when all treatments reach $n = n_T$. In the latter case, (2.6) implies $U_m(\mathbf{N}) \leq \delta^*$. □

3. Sample size comparisons. Let $N_i(P)$ and $N_i(EP)$ denote the treatment sample sizes after the Paulson phase and after the extended-Paulson procedure, respectively, $i = 1, \ldots k$. This section discusses the asymptotic behavior of these and other stopping times as $P^* \to 1$, with $\delta^* > 0$ fixed. Let $\delta_i = \theta_{(k)} - \theta_i$, $i = 1, \ldots, k$. Let n_I be fixed and for each P^*, let $b(n) \equiv b(n; P^*)$ be defined for all n, $n_I \leq n \leq n_T \equiv n_T(P^*)$. Assume $b(n; P^*)$ is decreasing in n, satisfying (2.3) and $\sigma b(n_T(P^*); P^*) \leq \delta^*$. Connect the points $b(n_I; P^*)$, $b(n_I + 1; P^*), \ldots,$ etc., to form a continuous decreasing function defined for all real $n \in [n_I, n_T(P^*)]$. Let $g(x; P^*)$ be the inverse of this function, defined and decreasing for all $x \in [b(n_T(P^*); P^*), b(n_I; P^*)]$. For $x < b(n_T(P^*); P^*)$,

define $g(x; P^*) \equiv n_T(P^*)$. Assume

$$\text{(3.1)}\qquad \begin{aligned} &\text{(i) } T(X_i; n) \to \theta_i \text{ a.s., } i = 1, \ldots, k, \ n \to \infty; \\ &\text{(ii) } n_T(P^*) \to \infty, \ P^* \to 1; \\ &\text{(iii) } b(n; P^*) \to \infty, \ P^* \to 1, \text{ for each } n \geq n_I. \end{aligned}$$

Note that (ii) and (iii) will usually follow from (2.3), and will imply $g(x; P^*) \to \infty$, $P^* \to 1$, for each x. Finally, assume $dg(x; P^*)/dx = g'(x; P^*)$ exists for each P^* and x, and

$$\text{(3.2)}\qquad \text{(iv)} \lim_{P^* \to 1} \left| \frac{g'(x; P^*)}{g(x; P^*)} \right| < B(x),$$

for some $B(x) < \infty$. Lemma 3.1 below generalizes Perng's (1969) result on $N_i(P)$ under Paulson's choice of $b(\cdot)$ and $T(\cdot)$. It also strengthens it, since it applies to all θ, not just those with $\delta_i > 0$ for all $i \neq (k)$. For every procedure discussed herein, $N_{(k)}/\max_{i \neq (k)}\{N_i\} \to 1$, so we need deal only with the properties of N_i for $i \neq (k)$.

LEMMA 3.1. *Under the assumptions stated immediately above,*

$$\frac{N_i(P)}{g\left[(\delta_i + \delta^*)/\sigma; P^*\right]} \to 1$$

in probability, as $P^* \to 1$, *for each* $i \neq (k)$.

PROOF. By Egoroff's theorem, for any $\varepsilon > 0$, $\exists\ B_1$, $P(B_1) < \varepsilon$, and n_ε such that off the event B_1,

$$\text{(3.3)}\qquad |T(X_i; n) - \theta_i| < \varepsilon,$$

for all $1 \leq i \leq k$ and $n \geq n_\varepsilon$. Note that, off B_1, for all $n \geq n_\varepsilon$,

$$\text{(3.4)}\qquad \left| \max_i T(\mathbf{X}_i; n) - T(\mathbf{X}_{(k)}; n) \right| < 2\varepsilon.$$

In the Paulson phase, a treatment j is sampled at stage($\mathbf{n}$) if

$$\text{(3.5)}\qquad U_i^j(\mathbf{n}) = \min_{n_I \leq n \leq n_{ij}} \left[T(\mathbf{X}_j; n) - T(\mathbf{X}_i; n) + \sigma b(n; P^*) \right] > \delta^*, \qquad i \neq j.$$

Let B_2 be the event that $[\min_i N_i(P) < n_\varepsilon]$. Choose P_1^* such that for all $P^* > P_1^*$, $b(n; P^*)$ is sufficiently large for each $n < n_\varepsilon$ to insure $P(B_2) < \varepsilon$. Off the event $[B_1 \cup B_2]$ the following inequalities hold at stage($\mathbf{n}$) $= N_i(P)$ and $N_i(P) - 1$ [using (3.4)]:

$$\text{(3.6)}\qquad \begin{aligned} &T(\mathbf{X}_i; N_i(P)) - T(\mathbf{X}_{(k)}; N_i(P)) + \sigma b(N_i(P); P^*) \leq \delta^* + 2\varepsilon, \\ &T(\mathbf{X}_i; N_i(P) - 1) - T(\mathbf{X}_{(k)}; N_i(P) - 1) + \sigma b(N_i(P) - 1; P^*) > \delta^* - 2\varepsilon. \end{aligned}$$

These, with (3.3), give

$$b(N_i(P); P^*) \le (\delta^* + \delta_i + 4\varepsilon)/\sigma,$$
$$b(N_i(P) - 1; P^*) > (\delta^* + \delta_i - 4\varepsilon)/\sigma. \tag{3.7}$$

These imply

$$\frac{g[(\delta^* + \delta_i - 4\varepsilon)/\sigma; P^*] + 1}{g[(\delta^* + \delta_i)/\sigma; P^*]} \ge \frac{N_i(P)}{g[(\delta^* + \delta_i)/\sigma; P^*]} \ge \frac{g[(\delta^* + \delta_i + 4\varepsilon)/\sigma; P^*]}{g[(\delta^* + \delta_i)/\sigma; P^*]}. \tag{3.8}$$

First-order Taylor series expansions of $g(\cdot)$ about $(\delta^* + \delta_i)/\sigma$ on both the left- and right-hand sides of (3.8) give, for P^* sufficiently large,

$$\left| \frac{N_i(P)}{g[(\delta^* + \delta_i)/\sigma; P^*]} - 1 \right| \le \frac{5\varepsilon}{\sigma} B[(\delta^* + \delta_i)/\sigma]$$

off $[B_1 \cup B_2]$. $P(B_1 \cup B_2) < 2\varepsilon$, ε arbitrary, completes the proof. □

REMARK. If every $\delta_i > 0$, $i \ne (k)$, and $\varepsilon < (\min_{i \ne (k)} \delta_i)/2$, then (3.3) yields $\max_i T(\mathbf{X}_i; n) \equiv T(\mathbf{X}_{(k)}; n)$ for $n \ge n_e$, off B_1. In this case, off $B_1 \cup B_2$,

$$U^i_{(k)}(\mathbf{N}) \le T(\mathbf{X}_i; N_i(P)) - T(\mathbf{X}_{(k)}; N_i(P)) + \sigma b(N_i(P), P^*) \le \delta^*,$$

for each $i \ne (k)$, after the Paulson phase. Hence, the extended-Paulson procedure reduces to the Paulson, and $N_i(EP) = N_i(P)$, $i = 1, \ldots, k$.

COROLLARY 3.1. *Under the assumptions of Lemma* 3.1, *and assuming* $\delta_i > 0$ *for* $i \ne (k)$,

$$\frac{N_i(EP)}{N_i(P)} \to 1$$

in probability, as $P^* \to 1$, $i = 1, \ldots, k$.

The (k)-retaining rules follow the format of the Paulson phase, with one major change. With appropriate choice of $T(\cdot)$ and $b(\cdot)$ satisfying (2.3) and $\sigma b(n_T; P^*) \le \delta^*$, they specify $G(\mathbf{n})$ of (2.5) replacing δ^* by 0, and then impose the restriction that sampling stop at n_T. Under these rules, treatment (k) is retained, with probability P^*, in the set of treatments sampled at each stage($\mathbf{n}$).

If $N_i(KR)$ is the sample size for treatment i under a (k)-retaining rule, clearly $N_i(KR) \ge N_i(P)$, $i = 1, \ldots, k$. The difference is quantified in Lemma 3.2.

LEMMA 3.2. *Under the assumptions of Lemma* 3.1,

$$\frac{N_i(KR)}{\min[n_T(P^*), g(\delta_i/\sigma; P^*)]} \to 1$$

in probability, as $P^* \to 1$, $i \ne (k)$.

PROOF. Follows the proof of Lemma 3.1, replacing δ^* by 0 in (3.5). Now, either the modified inequalities (3.6) hold, or $N_i(KR) = n_T(P^*)$. The expressions (3.7) then become

$$b(N_i(KR); P^*) \leq \max[b(n_T(P^*); P^*), (\delta_i + 4\varepsilon)/\sigma]$$

and

$$b(N_i(KR) - 1; P^*) > \max[b(n_T(P^*) - 1; P^*), (\delta_i - 4\varepsilon)/\sigma].$$

The remainder of the proof follows as before. □

Combining Lemmas 3.1 and 3.2, we can gauge the approximate savings in using the extended-Paulson procedure over a (k)-retaining rule, with the assumptions of Lemma 3.1 and $\delta_i > 0$ for all $i \neq (k)$:

$$\frac{N_i(EP)}{N_i(KR)} \sim \frac{g[(\delta_i + \delta^*)/\sigma; P^*]}{g[(\delta_i/\sigma); P^*]} \leq 1,$$

for $i \neq (k)$, noticing that $n_T(P^*) = g[(\delta_i/\sigma); P^*]$ for $\delta_i < \delta^*$.

It would appear that the (k)-retaining rules take no sample-size advantage of the fact that the experimenter considers any treatment with $\delta_i < \delta^*$ to be essentially a "best" treatment.

Acknowledgments. The author would like to thank an Associate Editor and a referee for helpful comments.

REFERENCES

EDWARDS, D. (1985). Modifications to sequential selection rules to strengthen associated confidence statements. In *The Frontiers of Modern Statistical Inference Procedures* (E. J. Dudewicz, ed.) 237–249. American Sciences Press, Columbus, Ohio.

EDWARDS, D. G. and HSU, J. C. (1983). Multiple comparisons with the best treatment. *J. Amer. Statist. Assoc.* **78** 965–971. Corrigendum **79** 965.

HSU, J. C. (1981). Simultaneous confidence intervals for all distances from the "best." *Ann. Statist.* **9** 1026–1034.

HSU, J. C. and EDWARDS, D. G. (1983). Sequential multiple comparisons with the best. *J. Amer. Statist. Assoc.* **78** 958–964.

KAO, S. C. and LAI, T. L. (1980). Sequential selection procedures based on confidence sequences for normal populations. *Comm. Statist. A—Theory Methods* **9** 1657–1676.

PAULSON, E. (1964). A sequential procedure for selecting the population with the largest mean from k normal populations. *Ann. Math. Statist.* **35** 174–180.

PERNG, S. K. (1969). A comparison of the asymptotic expected sample sizes of two sequential procedures for ranking problem. *Ann. Math. Statist.* **40** 2198–2202.

DEPARTMENT OF STATISTICS
UNIVERSITY OF SOUTH CAROLINA
COLUMBIA, SOUTH CAROLINA 29208