TWO-STAGE SELECTION PROCEDURES BASED ON TESTS*

by

Klaus J. Miescke

University of Illinois at Chicago

Technical Report #83-39

Department of Statistics
Purdue University

September 1983

*This research was supported by the Office of Naval Research Contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Abstract

Suppose that k new treatments have been developed with the purpose of replacing the standard treatment with the best new one, provided that it is actually an improvement on the standard treatment. In a parametric approach, mainly under the assumption of MLR, procedures are considered which, at a first stage, screen out inferior treatments through statistical tests at a common level of significance α_0. If none (exactly one) is not eliminated, none (this one) will be used as a replacement. Otherwise, if more than one treatment overcomes this screening process, that one of the non-eliminated treatments will be chosen as the replacement which is judged to be the best, after additional data have been observed from the selected treatments. Topics of this paper are the questions of how to choose the terminal decision at the second stage and the tests at the first stage, respectively, and how to implement the appropriate procedures at certain pre-specified performance criteria.

Key Words: Multiple comparisons with a control; 2-stage procedures; screening procedures.

1. Introduction The following procedure, which is used in certain clinical studies, may serve as a motivation for the considerations in this paper. Suppose that k new treatments have been developed with the purpose of replacing the standard treatment with the best new one, provided that it is actually an improvement on the standard treatment. In a pilot study, each new treatment is applied several times and screened out if it is not considered to be significaltly better than the standard treatment. Hereby, judgement is gained through suitable statistical tests at a fixed level of α_0. If all k new treatments are eliminated the standard treatment will not be replaced. If exactly one new treatment is not eliminated this will be taken as a replacement. In all other cases, the remaining treatments are further examined in a follow up study through additional applications, and finally that one which appears to be the best will be used as a replacement of the standard treatment. The natural questions of how to choose the tests in the first stage and the terminal decision in the second stage are the topic of this paper.

Let $\pi_1,\ldots,\pi_k$ be k populations associated with unknown parameters $\theta_1,\ldots,\theta_k \in \Omega \subseteq \mathbb{R}$. Let $\theta_0 \in \Omega$ be a control value which may be known or unknown. In the latter case, assume that there is also a control population π_0. A population π_i is considered to be better than π_0 if $\theta_i > \theta_0$, $i=1,\ldots,k$. The goal is to determine, in two stages, whether there is any population better than the control and, in the affirmative, which one is associated with the largest parameter. Assume that samples $\underline{X}_i = \{X_{ij}\}_{j=1,\ldots,n_i}$, $i=0,1,\ldots,k$, and $\underline{Y}_j = \{Y_{ij}\}_{j=1,\ldots,m_i}$, $i = 1,\ldots,k$, can be drawn from $\pi_0,\pi_1,\ldots,\pi_k$ at

the first and at the second stage, respectively, which are mutually independent. Let $\{f_\theta\}_{\theta\in\Omega}$ be a given family of densities with respect to μ, the Lebesgue measure on $\mathbb{R}$ or the counting measure on any lattice in $\mathbb{R}$, and assume that for every $i\in\{0,1,\ldots,k\}$ all observations from π_i have a common distribution with density f_{θ_i}. Later on, after Theorem 1 has been proved, we will make the additional assumption that for every sample $\underline{Z}$ of size n from one population there exists a sufficient statistic $T_n(\underline{Z})$ such that the family of joint densities has nondecreasing likelihood ratios in T_n. For notational convenience let them be in the following denoted by $U_i = T_{n_i}(\underline{X}_i)$, $i = 0,1,\ldots,k$, $V_i = T_{m_i}(\underline{Y}_i)$, and $W_i = T_{n_i+m_i}(\underline{X}_i,\underline{Y}_i)$, $i=1,\ldots,k$.

To simplify the presentation the case of a known control value θ_0 will be considered first. Before we define a natural class of two-stage procedures in a concise way, let us briefly describe how these procedures will be typically applied. For every testing problem $H_i\colon \theta_i \le \theta_0$ versus $K_i\colon \theta_i > \theta_0$ the experimenter chooses a test based on $\underline{X}_i$ with a fixed level α_0 and another test based on $(\underline{X}_i,\underline{Y}_i)$ with a variable level α. At Stage 1 he discards all populations which are not significant at level α_0 under the first set of tests. If none (exactly one) is left, he decides that none (this one) is better than the control and is the best population. If more than one population survives he proceeds to Stage 2. At Stage 2, he draws additional samples $\underline{Y}_i$ from those populations which have been selected at Stage 1 and makes a final decision in favor of that population among the selected ones which has the smallest p-value (i.e. is most significant) under the associated second test.

If these tests are upper level tests, which for simplicity may be non-randomized for a moment to fix ideas, based on some real-valued statistics $\tilde{U}_i$ and $\tilde{W}_i$, say, $i = 1,\ldots,k$, then the procedure considered above can be equivalently described as follows: At Stage 1 all π_i's are selected with $\tilde{U}_i \geq c_i$ (where c_i is the α_0-fractile of $\tilde{U}_i$ under $\theta_i = \theta_0$), and a final decision is made in terms of the largest $\tilde{W}_i$ among the selected π_j's. The truncated versions of such procedures (i.e. which perform Stage 1 only) have been studied by several authors, see, for example, Gupta and Sobel (1958) and Lehmann (1961). For further references see Gupta and Panchapakesan (1979) Chapter 20. Some preliminary results concerning two-stage procedures of the type described above in the case of $n_1 = \ldots = n_k$ and $m_1 = \ldots = m_k$ can be found in Gupta and Miescke (1982), which include a comparison with the one-stage analog by Bechhofer and Turnbull (1978).

To begin with, let us point out that several definitions given in Miescke (1979) will be relevant in the sequel but for brevity are not repeated here. Especially, tests may be randomized ones taking values in $[0,1]$. This typically occurs in discrete cases or in continuous type cases where nonparametric (rank) tests are under concern. Thus significance statements as well as p-values are understood to be based on additional randomization schemes as are used in Miescke (1979). To be more specific, let $\underline{A} = (A_1,\ldots,A_k)$ and $\underline{B} = (B_1,\ldots,B_k)$ be the randomization schemes for the first and the second stage, respectively. Note that the $\underline{X}_i$'s, $\underline{Y}_j$'s, $\underline{A}$ and $\underline{B}$ altogether are assumed to be mutually independent.

The class $\mathfrak{D}$ of two-stage procedures. For $i = 1,\ldots,k$, let $\varphi_i = \{\varphi_{i,\alpha}\}_{\alpha\in[0,1]}$ be a right continuous and monotone (in α) unbiased test for H_i versus K_i which

is standardized at $\theta_i = \theta_0$. Assume that $\varphi_{i,1} \equiv 1$ outside of the support of the distribution of $\underline{X}_i$ at $\theta_i = \theta_0$. Let $\underline{\varphi}_{\alpha_0} = (\varphi_{1,\alpha_0},\ldots,\varphi_{k,\alpha_0})$ where $0 < \alpha_0 < 1$ is fixed. Analogously, let $\psi_i = \{\psi_{i,\alpha}\}_{\alpha\in[0,1]}$ be such a test for H_i versus K_i based on $(\underline{X}_i,\underline{Y}_i)$. Let $\underline{\psi} = (\psi_1,\ldots,\psi_k)$. Let $\mathfrak{S}$ be the class of all procedures of the following type $(\underline{\varphi}_{\alpha_0}, \underline{\psi})$:

Stage 1: Select π_i if $p_{\varphi_i}(\underline{X}_i,A_i)$, the p-value of $\underline{X}_i$ under φ_i, is smaller than α_0, $i = 1,\ldots,k$. If none (exactly one) of the populations is selected, stop and decide none (this one) is better than π_0 and is the best population. Otherwise proceed to Stage 2.

Stage 2: Among the selected populations decide finally in favor of that π_j which has the smallest p-value $p_{\psi_j}(\underline{X}_j,\underline{Y}_j,B_j)$ under ψ_j.

The following result will prove to be useful in various aspects, except for the important question of how to optimize the component $\underline{\psi}$ in $(\underline{\varphi}_{\alpha_0}, \underline{\psi})$.

Theorem 1. Let $(\underline{\varphi}_{\alpha_0}, \underline{\psi}) \in \mathfrak{S}$. For notational convenience, let $E_i = E_{\theta_i}(\varphi_{i,\alpha_0}(\underline{X}_i))$ and $F_i(\alpha) = E_{\theta_i}(\varphi_{i,\alpha_0}(\underline{X}_i)\psi_{i,\alpha}(\underline{X}_i,\underline{Y}_i))$, $\alpha\in[0,1]$, $i = 1,\ldots k$, $\underline{\theta}\in\Omega^k$. Then for every non-empty $D \subseteq \{1,\ldots,k\}$ and $\theta\in\Omega^k$,

(1) $P_{\underline{\theta}}$ {final decision falls into D}

$$= \int_0^1 \prod_{j\notin D} [1-F_j(\alpha)]\, d\Big(1 - \prod_{i\in D} [1-F_i(\alpha)]\Big),$$

(2) $P_{\underline{\theta}}$ {final decision is in favor of π_i}

$$= \int_0^1 \prod_{\substack{j=1\\ j\neq i}}^{k} [1-F_j(\alpha)]\, dF_i(\alpha), \quad i = 1,\ldots,k,$$

(3) $P_{\underline{\theta}}$ {final decision is made at Stage 1 in favor of π_i}

$$= \prod_{\substack{j=1 \\ j\neq i}}^{k} [1 - E_j]\, E_i, \qquad i = 1,\ldots,k,$$

(4) $P_{\underline{\theta}}$ {final decision is in favor of the control}

$$= \prod_{j=1}^{k} [1 - E_j].$$

Proof: It has been shown in Miescke (1979) [cf. (2.3) - (2.5) loc. cit.] that the distribution function of each p-value appearing in $(\underline{\varphi}_{\alpha_0}, \underline{\psi})$ equals to the power function of the corresponding test, which is a continuous function of $\alpha \in [0,1]$ at every fixed parameter point, and which at $\alpha = 1$ assumes the value one.

Let now D be a non-empty subset of $\{1,\ldots,k\}$. For $j = 1,\ldots,k$, let $p^*_{\psi_j}(\underline{X}_j, \underline{Y}_j, B_j)$ be equal to $p_{\psi_j}(\underline{X}_j, \underline{Y}_j, B_j)$ if $p_{\varphi_j}(\underline{X}_j, A_j) \leq \alpha_0$, and let it be equal to 1 otherwise. Then it is easy to see that the l.h.s. of (1) is equal to

$$P_{\underline{\theta}} \{ \min_{i\in D} p^*_{\psi_i}(\underline{X}_i, \underline{Y}_i, B_i) \} < \min_{j\notin D} \{p^*_{\psi_j}(\underline{X}_j, \underline{Y}_j, B_j)\} \}.$$

Since for $i \in \{1,\ldots,k\}$ and $\theta_i \in \Omega$, $P_{\theta_i} \{p^*_{\psi_i}(\underline{X}_i, \underline{Y}_i, B_i) \leq \alpha\}$ is equal to $F_i(\alpha)$ if $0 \leq \alpha < 1$, and is equal to 1 if $\alpha = 1$, (1) follows by standard arguments. (2) is a special case of (1) which was stated only because of its relevance in later applications. The verification of (3) and (4) is straightforward and therefore the proof is omitted.

Remark 1. It will be shown in Section 2 that under the assumption of monotone likelihood ratios (MLR), every $(\underline{\varphi}_{\alpha_0}, \underline{\psi}) \in \mathfrak{S}$ is dominated by $(\underline{\varphi}_{\alpha_0}, \underline{\psi}^*)$ if $n_1 + m_1 = \ldots = n_k + m_k$, where $\underline{\psi}^*$ consists of the uniformly most powerful (UMP) tests. Hereby the results of Theorem 1 will not be of great help. There is, however, a particular situation where (1) and (2) can be used for a similar purpose. Suppose that the data of Stage 1 are not available but the information which populations have been significant is a hand. Then one has to use tests at Stage 2 which depend only on the $\underline{Y}_i$'s from the selected populations. In this case every $F_i(\alpha)$ factorizes into the product of the two power functions of φ_{i,α_0} and $\psi_{i,\alpha}$, respectively, and therefore (1) and (2) are completely determined through these power functions. For example, if $D_+(\underline{\theta}) = \{i \mid \theta_i > \theta_0, i \in \{1,\ldots,k\}\}$ is not empty, then (1) for $D = D_+(\underline{\theta})$ is maximized by the procedure which uses the UMP-tests at both stages. This is true even if the $n_i + m_i$'s are not assumed to be all equal. Since these and related results in such a special case, however, are considered to be of less statistical importance they will not be discussed in further detail.

In the case of an unknown control parameter θ_0 some obvious changes have to be made. First of all the tests φ_{i,α_0} depend now on $(\underline{X}_i, \underline{X}_0)$, $i=1,\ldots,k$, whereas $\underline{\psi}$ remains the same as before. Let $\mathfrak{S}'$ denote the class of two-stage procedures of the type $(\underline{\varphi}_{\alpha_0}, \underline{\psi})$ in this case. The analog of Theorem 1 for $\mathfrak{S}'$ can be attained by replacing the right hand sides of (1) - (4) by their integrals with respect to the distribution of $\underline{X}_0$. If not explicitly stated otherwise, the results to be derived in the sequel for $\mathfrak{S}$ have analogous counterparts for $\mathfrak{S}'$ which will not be formulated or proved for brevity because of the close similarities.

2. Optimality Results For MLR-Families. In the following let $D_+(\underline{\theta}) = \{i \mid \theta_i > \theta_0,\ i \in \{1,\ldots,k\}\}$ be the "good" populations and $D_-(\underline{\theta}) = \{1,\ldots,k\} \setminus D_+(\underline{\theta})$ be the "bad" ones, $\underline{\theta} \in \Omega^k$. Also let us partition the parameter space Ω^k into $\Omega^k_- = \{\underline{\theta} \in \Omega^k \mid \theta_i \leq \theta_0,\ i = 1,\ldots,k\}$ and its complement Ω^k_+, say. A procedure is said to make a correct selection (CS) at $\underline{\theta} \in \Omega^k_-$ if all populations are eliminated at Stage 1, and it is said to make a correct selection at $\underline{\theta} \in \Omega^k_+$ if a final decision is made in favor of a population with the largest θ-value. Let the goal be now to find a procedure in $\mathcal{S}$ which has a large probability of a correct selection (PCS) on Ω^k. From now on we assume that the family $\{f_\theta\}_{\theta \in \Omega}$ has the MLR-property as specified in Section 1. Then the following partial solution to our problem can be given.

Theorem 2. Let $(\varphi_{\alpha_0}, \underline{\psi}) \in \mathcal{S}$. If $n_1 + m_1 = \ldots = n_k + m_k$ and $\psi_1 = \ldots = \psi_k$ then for all $\underline{\theta} \in \Omega^k$,

$$(5) \qquad \sum_{\sigma} P_{\sigma(\underline{\theta})}\{\text{CS under } (\varphi_{\alpha_0}, \underline{\psi})\} \leq \sum_{\sigma} P_{\sigma(\underline{\theta})}\{\text{CS under } (\varphi_{\alpha_0}, \underline{\psi}^*)\},$$

where $\underline{\psi}^*$ consists of the UMP-tests for H_i versus K_i, $i = 1,\ldots,k$, which in this case are all identical, and where the summation is with respect to all $k!$ permutations of $(1,\ldots,k)$. $\sigma(\underline{\theta}) = (\theta_{\sigma(1)},\ldots,\theta_{\sigma(k)})$.

Proof: Only an outline of the proof will be given since it follows by similar decision theoretic arguments as have been used previously in Gupta and Miescke (1983).

Under the assumptions stated above, the associated decision function of $(\varphi_{\alpha_0}, \underline{\psi})$ which determines final selections at Stage 2 is permutation invariant. The loss function which is implicitly employed is zero if a correct selection

is made, and is one, otherwise. Its component which is associated with final selections at Stage 2 is permutation invariant and favors selections of populations with large parameters.

Let now $\underline{\theta} \in \Omega_+^k$ be fixed and, in a Bayes approach, assume that the unknown parameter vector is random and has a prior distribution which gives equal mass 1/k! to all permutations $\sigma(\underline{\theta})$ of $\underline{\theta}$. Then the posterior distribution of the parameter vector, given $W_i = w_i$, $i = 1,\ldots,k$, has the decreasing in transposition (DT) property. From this fact and the properties of the loss function stated above it follows that the optimal final decision at Stage 2 is the natural one which is made with respect to the non-eliminated population with the largest W_i, where ties are broken at random. Clearly this is equivalent to selecting the non-eliminated population with the smallest p-value under test ψ^*. The proof is now completed by noting that (5) gives a comparison of the corresponding Bayes risks, where of course at all $\underline{\theta} \in \Omega_-^k$ the probabilities of a correct selection are the same for both procedures.

Corollary 1. If, under the assumptions of Theorem 2, additionally $n_1 = \ldots = n_k$ and $\varphi_{1,\alpha_0} = \ldots = \varphi_{k,\alpha_0}$ is given then for all $\underline{\theta} \in \Omega^k$

$$(6) \qquad P_{\underline{\theta}}\{\text{CS under } (\underline{\varphi}_{\alpha_0}, \underline{\psi})\} \leq P_{\underline{\theta}}\{\text{CS under } (\underline{\varphi}_{\alpha_0}, \underline{\psi}^*)\} .$$

Proof: Since both procedures considered here are completely permutation invariant, and since also the 0-1 loss function employed is permutation invariant, their risk functions are symmetric functions of $\underline{\theta} \in \Omega^k$. Therefore

all summands on the l.h.s. of (5) coincide and the same holds for the summands on the r.h.s. of (5).

Remark 2. The proof of Theorem 2 actually applies more generally to the following situation. No matter of how the populations are eliminated at Stage 1, if $n_1 + m_1 = \ldots = n_k + m_k$, and if only permutation invariant final decision functions are admitted at Stage 2, then every Bayes procedure w.r.t. any symmetric prior employs the natural rule at State 2. Of course, also from a non-Bayesian point of view, (5) is an intuitively appealing criterion. It simply reflects the lack of knowledge of how the sample sizes are associated with the k ordered populations parameters. Since there is not even an approximately similar result available in the case of unequal $n_i + m_i$'s, it is strongly recommended to "repair" the design of every experiment with unequal n_i's by choosing the m_i's appropriately to get equal overall sample sizes. Let us assume from now on that $n_1 + m_1 = \ldots = n_k + m_k = N$, say, holds.

Actually, in various selection problems authors have chosen their designs such that the statistics on which the natural final decision rule is based have joint distributions with the DT property. To mention a few relevant examples, Bechhofer (1954), Bechhofer, Dunnett and Sobel (1954), and Dudewicz and Dalal (1975) have done so to be able to implement their procedures at certain specified performance requirements. It may now be added that exactly in these designs the employed natural final decisions are optimal in terms of the risk or the PCS, respectively, uniformly on all parameter configurations.

Lack of the DT property in distributions of statistics used for final

decisions, in cases where this property could be attained in principle, should be considered as pathological designs. Even in the simplest case of a one-stage selection procedure serious difficulties arise, as may be illustrated by the following problem which was emphasized by Bechhofer (1982).

Example. Let $\bar{X}_1,\ldots,\bar{X}_k$ be independent sample means with unknown expectations $\theta_1,\ldots,\theta_k$ and with known but different variance $q_1,\ldots,q_k$ generated from k normal populations. For the problem of selecting a population with the largest θ-value no procedure exists which has a largest PCS, uniformly in $\underline{\theta} \in \mathbb{R}^k$. If $k = 2$ the natural rule is Bayes w.r.t. independent priors $N(0,q_1)$ and $N(0,q_2)$, respectively, and it can be seen to be admissible under the 0-1 loss function. However, if $k \geq 3$ the natural rule cannot be Bayes with respect to any (multivariate) normal prior. The question of whether it is admissible is still open. On the other hand, for every normal prior with expectation $(0,\ldots,0)$, for which the posterior distribution has the DT -property (which leads to a simple solution) the Bayes rule selects in terms of the largest $\bar{X}_i/q_i$, $i = 1,\ldots,k$.

In the remainder of this paper only procedures of the type $(\underline{\varphi}_{\alpha_0}, \underline{\psi}^*) \in \mathcal{D}$ will be considered. This is justified in view of the assumption $n_1 + m_1 = \ldots = n_k + m_k = N$ and of Theorem 2. The following result, which generalizes Theorem 3 of Gupta and Miescke (1982), can be used to find least favorable parameter configurations (LFC) of such procedures on suitable subspaces of Ω^k.

Theorem 3. Let $(\underline{\varphi}_{\alpha_0}, \underline{\psi}^*) \in \mathcal{D}$ where for every $i \in \{1,\ldots,k\}$ the power function of φ_{i,α_0} is nondecreasing in θ_i. Then the performance characteristics con-

sidered in Theorem 1 have the following monotonicity properties. (1) is nondecreasing in θ_i, $i \in D$ and nonincreasing in θ_j, $j \notin D$. (2) and (3) are nondecreasing in θ_i and nonincreasing in θ_j, $j \neq i$. (4) is nonincreasing in $\theta_1,\ldots,\theta_k$.

Proof: The assertions concerning (3) and (4) are obviously true. To prove those concerning (1) and (2) note that for $i \in \{1,\ldots,k\}$, $\underline{\theta} \in \Omega^k$, and $\alpha \in [0,1]$,

$$(7) \qquad F_i(\alpha) = E_{\theta_i}(\varphi_{i,\alpha_0}(\underline{X}_i))\, E_{\theta_i}\{\psi^*_{i,\alpha}(\underline{X}_i,\underline{Y}_i) \mid p_{\varphi_i}(\underline{X}_i,A_i) \leq \alpha_0\}.$$

The first factor on the r.h.s. of (7) is nondecreasing in θ_i according to the assumptions made above. The second factor can be seen to have the same property by applying Theorem 1 of Simons (1980) which guarantees that for every sample from a MLR-family, conditionally on any proportion of the information in the sample which one might choose to extract, likelihood ratios are still stochastically nondecreasing. Since $\psi^*_{i,\alpha}(\underline{X}_i,\underline{Y}_i)$ is a nondecreasing function of W_i, the proof is completed by noting that A_i could be ignored since the arguments apply to every situation $A_i = a_i$, where $a_i \in [0,1]$ is held fixed.

Corollary 2. Under the assumptions of Theorem 3, let $n_1 = \ldots = n_k$ and $\varphi_{1,\alpha_0} = \ldots = \varphi_{k,\alpha_0}$. Then for every $\underline{\theta} \in \Omega^k$ with $\theta_1 \leq \ldots \leq \theta_k$, the probability of a final decision in favor of population π_i is nondecreasing in $i \in \{1,\ldots,k\}$.

Proof: Given the assumptions above it can be seen from the proof of Theorem 3 that for $\theta_1 \leq \ldots \leq \theta_k$,

(8) $$F_1(\alpha) \leq F_2(\alpha) \leq \ldots \leq F_k(\alpha), \ \alpha \in [0,1].$$

Therefore the assertion follows from (2) by the same technique which was used in (IV) of Miescke (1979).

Theorem 4. Let $(\varphi_{\alpha_0}, \underline{\psi}^*) \in \mathfrak{S}$ where φ_{α_0} consists of consistent tests. Then for increasing sample sizes n_i and $m_i = N - n_i$, $i = 1,\ldots,k$, the probability of a correct selection tends to one at all $\underline{\theta} \in \Omega_+^k$ and at all $\underline{\theta} \in \Omega_-^k$ with $\theta_i < \theta_0$, $i = 1,\ldots,k$.

Proof: If $\underline{\theta} \in \Omega_+^k$ with exactly one coordinate greater than θ_0, or if $\underline{\theta} \in \Omega_-^k$ with $\theta_i < \theta_0$, $i = 1,\ldots,k$, the assertion follows immediately from (3) and (4), respectively. For all other $\underline{\theta} \in \Omega_+^k$, the probability that all populations π_i with $\theta_i > \theta_0$ will not be eliminated at Stage 1 tends to one. Moreover, $P_{\underline{\theta}}\{W_i > W_j \text{ for all } j \neq i\}$ also tends to one if θ_i is the unique maximum of $\theta_1,\ldots,\theta_k$. This can be seen as follows. Selecting in terms of the largest W_j is equivalent to selecting in terms of the smallest p-value under tests $\tilde{\psi}_j$ which are essentially the same tests as ψ_j^* but now standardized at θ_i. By Theorem 2 of Miescke (1979) it follows that

(9) $$P_{\underline{\theta}}\{W_i > W_j \text{ for all } j \neq i\} = \int_0^1 \prod_{j \neq i} [1 - E_{\theta_j}(\tilde{\psi}_{j,\alpha}(\underline{X}_j, \underline{Y}_j))]d\alpha,$$

which now can be seen to tend to one if N tends to infinity. If, however, all the good populations are not eliminated at Stage 1 and $W_i > W_j$, for all $j \neq i$, then a correct selection is made. This completes the proof in the given parameter configuration. The case of more than one best population can be treated similarly.

Focussing now on the first component $\underline{\varphi}_{\alpha_o}$ in the procedures $(\underline{\varphi}_{\alpha_o}, \underline{\psi}^*) \in \mathcal{D}$, the natural choice is of course $\underline{\varphi}^*_{\alpha_o}$ which consists of the corresponding UMP-tests for H_i versus K_i based on $\underline{X}_i$, or more precisely, based on $U_i, i = 1,\ldots,k$. Even though such a choice cannot be justified (not even in the case of $n_1 = \ldots = n_k$) by an overall improvement on the PCS, several strong reasons can be quoted in support of choosing $(\underline{\varphi}^*_{\alpha_o}, \underline{\psi}^*)$. First, of course, all results derived hitherto hold for this procedure. Second, the following can be stated.

Theorem 5. Among all procedures in $\mathcal{D}$, $(\underline{\varphi}^*_{\alpha_o}, \underline{\psi}^*)$ maximizes

(10) $P_{\underline{\theta}}\{CS\}$ at every $\underline{\theta} \in \Omega^k_-$,

(11) $P_{\underline{\theta}}\{CS \text{ at Stage } 1\}$ at every $\underline{\theta} \in \Omega^k_+$ with exactly one $\theta_i > \theta_0$,

(12) $E_{\underline{\theta}}$ (number of good populations selected at Stage 1) at every $\underline{\theta} \in \Omega^k_+$,

(13) $E_{\underline{\theta}}$ (number of bad populations eliminated at Stage 1) at every $\underline{\theta} \in \Omega^k$.

Proof: (10) follows from (4) and (11) follows from (3). All arguments are standard and are based on the well known properties of the power functions of the UMP-tests $\varphi^*_{i,\alpha_o}, i = 1,\ldots,k$. Therefore, no further details will be given.

Third, all permutation invariant procedures $(\underline{\varphi}_{\alpha_o}, \underline{\psi}^*)$, except $(\underline{\varphi}^*_{\alpha_o}, \underline{\psi}^*)$ ifself, can be modified, without changing the sizes of the selected subsets of populations at Stage 1, in a specific way which leads to an improvement on the PCS, provided that the family $\{f_\theta\}_{\theta \in \Omega}$ is a strongly unimodal exponential family. It should be noted that the modified procedure is also based on tests but is no longer a member of the class $\mathcal{D}$. More precisely, from Corollary 2 in Gupta and Miescke (1983) the following result can be derived.

Theorem 6. Let $\{f_\theta\}_{\theta \in \Omega}$ be a strongly unimodal (i.e. log-concave) exponential family. Then every $(\varphi_{\alpha_0}, \underline{\psi}^*) \in \mathcal{S}$ with $n_1 = \ldots = n_k$ and $\varphi_{1,\alpha_0} = \ldots = \varphi_{k,\alpha_0}$ can be improved by simply replacing the selected populations at Stage 1 by the same number of populations, but now by those which are associated with the largest U_i's, where ties are broken at random. Then for all $\underline{\theta} \in \Omega^k$,

$$(14) \qquad P_{\underline{\theta}}\{\text{CS under } (\varphi_{\alpha_0}, \underline{\psi}^*)\} \leq P_{\underline{\theta}}\{\text{CS under the modified procedure}\}.$$

3. Applications With Illustrations In The Normal Case. In applications the procedure $(\underline{\varphi}^*_{\alpha_0}, \underline{\psi}^*)$ usually will be implemented as to meet certain performance requirements. This will be described in this section and will be illustrated by the example of k normal populations $N(\theta_i, \sigma_i^2)$, $i = 1,\ldots,k$, where U_i and W_i are the corresponding sample means, $i = 1,\ldots,k$. Here the procedure can be considered to be the two-stage analog of the one-stage procedure by Bechhofer and Turnbull (1978). At first consider the basic requirement

$$(15) \qquad \inf\{ P_{\underline{\theta}}\{\text{CS under } (\underline{\varphi}^*_{\alpha_0}, \underline{\psi}^*) \mid \underline{\theta} \in \Omega^k_- \} = P_0^*,$$

where P_0^* is a predetermined constant. In view of (4) this can be accomplished by choosing α_0 to satisfy

$$(16) \qquad (1 - \alpha_0)^k = P_0^*.$$

Then in the normal case the procedure is of the following form. At Stage 1, population π_i is selected if $n_i^{1/2}(U_i - \theta_0) / \sigma_i \geq \Phi^{-1}(1 - \alpha_0)$, $i = 1,\ldots,k$, where Φ denotes the c.d.f. of $N(0,1)$. And at Stage 2, a final decision is made in terms of the largest W_j from the selected populations.

Since (15) actually involves only the properties of the procedure at the

first stage, the P_0^* - condition can be attained by employing techniques used for one-stage procedures. Thus (15) can be solved by taking recourse to relevant papers in this area. For further details see Gupta and Panchapakesan (1979).

A second requirement will typically employ the indifference zone approach which is due to Bechhofer (1954). Let $\Delta > 0$ be fixed and let $\Omega_\Delta^k = \{\underline{\theta} \in \Omega^k \mid \theta_i + \Delta \geq \theta_0, \theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots \theta_k$ for some $i\}$. Now consider the requirement

$$\text{(16)} \qquad \inf \{P_{\underline{\theta}} \{\text{CS under } (\underline{\varphi}_{\alpha_0}^*, \underline{\psi}^*)\} \mid \underline{\theta} \in \Omega_\Delta^k\} = P_1^* ,$$

where P_1^* is a second predetermined constant. Even though Theorem 3 can be used to find the LFC, it is technically too difficult to attain (16) exactly. Therefore, the following conservative approach, which over-protects the experimenter with respect to (16), is recommended and is easy to perform. Let $\underline{\theta} \in \Omega_\Delta^k$ with $\theta_i = \max \{\theta_1, \ldots, \theta_k\}$, say. If population π_i is selected at Stage 1 and W_i is the unique maximum of $W_1, \ldots, W_k$ then a correct selection is made. Therefore if the following two conditions are fulfilled, and β_1 and β_2 are chosen to meet $\beta_1 + \beta_2 - 1 = P_1^*$ then by Bonferroni's inequality it follows that the l.h.s. of (16) is not smaller than P_1^*. The conditions are

$$\text{(17)} \qquad E_{\theta_0 + \Delta}(\varphi_{j,\alpha_0}^*(\underline{X}_j)) \geq \beta_1, \; j = 1, \ldots, k, \text{ and}$$

$$\text{(18)} \qquad \inf \{P_{(\theta+\Delta, \theta, \ldots, \theta)} \{W_1 > W_2, \ldots, W_k\} \mid \theta \in \Omega\} \geq \beta_2 .$$

In the normal case it is well known (cf. Tamhane and Bechhofer (1979)) that Slepian's inequality leads to better results than Bonferroni's inequality.

Use of the former allows to choose β_1 and β_2 according to the condition $\beta_1\beta_2 = P_1^*$ which is preferable since $\beta_1 + \beta_2 - 1 < \beta_1\beta_2$ for $0 < \beta_1, \beta_2 < 1$.

To meet (17), standard techniques from the theory of testing hypotheses can be used. And (18) can be attained by using results of single-stage selection procedures in the indifference zone approach due to Bechhofer (1954).

REFERENCES

[1] Bechhofer, R. E. (1954): A single sample multiple decision procedure for ranking means of normal populations with known variances. Ann. Math. Statist. 25, 16-39.

[2] Bechhofer, R. E. (1982): Private communication.

[3] Bechhofer, R. E., Dunnett, C. W., and Sobel, M. (1954): A two-sample multiple-decision procedure for ranking means of normal populations with a common unknown variance. Biometrika 41, 170-176.

[4] Bechhofer, R. E. and Turnbull, B. W. (1978): Two (k+1) - decision selection procedures for comparing k normal means with a specified standard. J. Amer. Statist. Assoc. 73, 385-392.

[5] Dudewicz, E. J. and Dalal, S. R. (1975): Allocation of observations in ranking and selection with unequal variances. Sankhyā Ser. B 37, 28-78.

[6] Gupta, S. S. and Miescke, K. J. (1982): On the problem of finding a best population with respect to a control in two stages. In Statistical Decision Theory and Related Topics III (Ed. Gupta, S. S. and Berger, J. O.), Vol. 1, 473-496. Academic Press, New York.

[7] Gupta, S. S. and Miescke, K. J. (1983): Sequential selection procedures - A decision theoretic approach. Mimeo. Ser. #82-6, revised July 1983, Dept. of Statistics, Purdue Univ., W. Lafayette.

[8] Gupta, S. S. and Panchapakesan, S. (1979): Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations. J. Wiley, New York.

[9] Gupta, S. S. and Sobel, M (1958): On selecting a subset which contains all populations better than a standard. Ann. Math. Statist. 29, 235-244.

[10] Lehmann, E. L. (1961): Some model I problems of selection. Ann. Math. Statist. 32, 990-1012.

[11] Miescke, K. J. (1979): Identification and selection procedures based on tests. Ann. Statist. 7, 207-219.

[12] Simons, G. (1980): Extensions of the stochastic ordering property of likelihood ratios. Ann. Statist. 8, 833-839.

[13] Tamhane, A. C. and Bechhofer, R. E. (1979): A two-stage minimize procedure with screening for selecting the largest normal mean (II): An improved PCS lower bound and associated tables. Commun. Statist. - Theor. Meth. A8, 337-358.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report #83-39	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): TWO-STAGE SELECTION PROCEDURES BASED ON TESTS*	5. TYPE OF REPORT & PERIOD COVERED: Technical
6. PERFORMING ORG. REPORT NUMBER: Technical Report #83-39	
7. AUTHOR(s): Klaus J. Miescke	8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0455
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Purdue University, Department of Statistics, West Lafayette, IN 47907	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Washington, DC	12. REPORT DATE: September 1983
	13. NUMBER OF PAGES: 18
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Multiple comparisons with a control; 2-stage procedures; screening procedures.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Suppose that k new treatments have been developed with the purpose of replacing the standard treatment with the best new one, provided that it is actually an improvement on the standard treatment. In a parametric approach, mainly under the assumption of MLR, procedures are considered which, at a common level of significance α_0. If none (exactly one) is not eliminated, none (this one) will be used as a replacement. Otherwise, if more than one treatment overcomes this screening process, that one of the non-eliminated

DD FORM 1 JAN 73 1473

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

treatments will be chosen as the replacement which is judged to be the best, after additional data have been observed from the selected treatments. Topics of this paper are the questions of how to choose the terminal decision at the second stage and the tests at the first stage, respectively, and how to implement the appropriate procedures at certain pre-specified performance criteria.