Proceedings of the 2003 Winter Simulation Conference
S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds.

INFERENCES FROM INDIFFERENCE-ZONE SELECTION PROCEDURES

E. Jack Chen

BASF Corporation
3000 Continental Drive - North
Mount Olive, NJ 07828, U.S.A.

W. David Kelton

Department of Quantitative Analysis and
Operations Management
University of Cincinnati
Cincinnati, OH 45221, U.S.A.

ABSTRACT

Two-stage indifference-zone selection procedures have been widely studied and applied. It is known that most indifference-zone selection procedures also guarantee multiple comparisons with the best confidence intervals with half-width corresponding to the indifference amount. We provide the statistical analysis of multiple comparisons with a control confidence interval that bounds the difference between each design and the unknown best and multiple comparisons with the best confidence intervals. The efficiency of selection procedures can be improved by taking into consideration the differences of sample means, using the variance reduction technique of common random numbers, and using sequentialized selection procedures. An experimental performance evaluation demonstrates the validity of the confidence intervals and efficiency of sequentialized selection procedures.

1 INTRODUCTION

When evaluating k alternative system designs, we would like to select one as the best and to control the probability that the selected design really is the best. Let μ_i denote the expected response of design i. Our goal is to find the design with the smallest expected response $\mu^* = \min_{1 \le i \le k} \mu_i$. If the goal is to select a design with the biggest expected response, just replace min with max in the above. We achieve this goal by using a class of ranking and selection (R&S) procedures. Let CS denote the event of "correct selection." In a stochastic simulation, CS can never be guaranteed with certainty. The probability of CS, denoted by P(CS), becomes higher as sample sizes become larger.

Most indifference-zone selection procedures are directly or indirectly developed based on Dudewicz and Dalal's (1975) or Rinott's (1978) indifference-zone selection procedures and efficiency is still a key concern for application of simulation to R&S problems. There are several new approaches aiming to improve the efficency of R&S procedures; Chen and Kelton (2000) take into account the difference between sample means, Kim and Nelson (2001) propose using sequentialized selection procedures. For an overview of earlier methods of R&S see Law and Kelton (2000). Furthermore, it is known that under the same assumptions of the indifference-zone selection procedures, most indifference-zone selection procedures also guarantee multiple comparisons with the best (MCB) confidence intervals (CIs) with half-width corresponding to the indifference amount; see Section 2.1. We are also interested in how much better the unknown best is relative to each alternative. In this paper, we provide the statistical analysis of multiple comparisons with a control (MCC) CIs that bound the difference between each design and the unknown best design as well as MCB CIs. Furthermore, we provide the rationale of taking into account the difference of sample means when computing the sample sizes and the effect of using common random numbers (CRN) with indifference-zone selection procedures to reduce variances. We then propose sequentializing Chen's (2002) adjusted ETSS (enhanced two-stage selection) procedure to improve P(CS) and efficiency of R&S procedures.

The rest of this paper is organized as follows. In Section 2, we provide the background of indifference zone and Rinott's selection procedure. In Section 3, we present our analysis of MCC and MCB CIs as well as techniques to improve the efficiency of two-stage selection procedures. In Section 4, we show our empirical-experimental results. In Section 5, we give concluding remarks.

2 BACKGROUND

First, some notation:

- X_{ij}: the observation from the j^{th} replication or batch of the i^{th} design,
- N_i: the number of replications or batches for design i,
- μ_i: the expected performance measure for design i, i.e., $\mu_i = E(X_{ij})$,

$\bar{X}_i(r)$**:** the sample mean performance measure for design i with r samples, i.e., $\sum_{j=1}^{r} X_{ij}/r$,

$\bar{X}_i$**:** the final sample mean performance measure for design i, i.e., $\sum_{j=1}^{N_i} X_{ij}/N_i$,

σ_i^2**:** the variance of the observed performance measure of design i from one replication or batch, i.e., $\sigma_i^2 = \mathrm{Var}(X_{ij})$,

$S_i^2(N_i)$**:** the sample variance of design i with N_i replications or batches, i.e., $S_i^2(N_i) = \sum_{j=1}^{N_i}(X_{ij} - \bar{X}_i)^2/(N_i - 1)$.

2.1 Indifference-Zone Selection Procedures

Let μ_{i_l} be the l^{th} smallest of the μ_i's, so that $\mu_{i_1} \leq \mu_{i_2} \leq \ldots \leq \mu_{i_k}$. Our goal is to select a design with the smallest expected response μ_{i_1}. However, in practice if the difference between μ_{i_1} and μ_{i_2} is very small, we might not care if we mistakenly choose design i_2, whose expected response is μ_{i_2}. The "practically significant" difference d^* (a positive real number) between the best and a satisfactory design is called the indifference zone in the statistical literature, and it represents the smallest difference that we care about. Therefore, we want a procedure that avoids making a large number of replications or batches to resolve differences less than d^*. That is, we want P(CS) $\geq P^*$ provided that $\mu_{i_2} - \mu_{i_1} \geq d^*$, where the minimal CS probability P^* and the "indifference" amount d^* are both specified by the user. That means we want to select a system i such that $\mu_i - \mu_{i_1} < d^*$. Some literature refers to this event as the probability of good selection (P(GS)) and use P(CS) to indicate the event in which we select system i_1. In this paper, we do not distinguish between the two and use P(CS) to indicate the event that we select a good design.

One way to look at the indifference amount is that indifference-zone procedures need to rank the designs with a desired confidence when their performance measures differ by more than d^*. More specifically, if design j has $\mu_j < \mu_{i_1} + d^*$, then indifference-zone procedures do not guarantee the order of these designs with a desired confidence. On the other hand, if $\mu_j \geq \mu_{i_1} + d^*$, indifference-zone procedures do guarantee the order of these designs with a desired confidence.

2.2 The Two-Stage Rinott Procedure

The two-stage procedure of Rinott (1978) has been widely studied and applied. Let n_0 be the number of initial replications or batches. The first-stage sample means $\bar{X}_i(n_0)$, and marginal sample variances

$$S_i^2(n_0) = \sum_{j=1}^{n_0} \frac{(X_{ij} - \bar{X}_i(n_0))^2}{n_0 - 1},$$

for $i = 1, 2, \ldots, k$ are computed. Based on the number of initial replications or batches n_0 and the sample variance $S_i^2(n_0)$ obtained from the first stage, the number of additional simulation replications or batches for each design in the second stage is $N_i - n_0$, where

$$N_i = \max(n_0, \lceil (hS_i(n_0)/d^*)^2 \rceil), \text{ for } i = 1, \ldots, k, \quad (1)$$

where $\lceil z \rceil$ is the smallest integer that is greater than or equal to the real number z, and h (which depends on k, P^*, and n_0) is a constant that solves Rinott's (1978) integral (h can be found from the tables in Wilcox (1984), or can be calculated by the FORTRAN program *rinott* in Bechhofer et al. (1995)). We then compute the overall sample means $\bar{X}_i = \sum_{j=1}^{N_i} X_{ij}/N_i$, and select the design with the smallest $\bar{X}_i$ as the best one. Basically, the computing budget is allocated proportionally to the estimated sample variances. Moreover, the derivation of this procedure is based on the *least favorable configuration* (LFC, i.e., assuming $\mu_{i_l} = \mu_{i_1} + d^*$, for all $l = 2, 3, \ldots, k$). However, in reality, we rarely encounter the LFC, so this procedure is conservative. That is, it generally allocates more samples than needed in order to reach the desired correct-selection guarantees.

Nelson and Matejcik (1995) show that most indifference-zone procedures not only guarantee a probability of correct selection, but they also guarantee MCB CI coverage probability of at least P^* with the half-width of the CI corresponding to the indifference amount d^* under the same assumptions for indifference-zone procedures. That is,

$$\mathrm{P}[\mu_i - \min_{j \neq i} \mu_j \in [(\bar{X}_i - \min_{j \neq i} \bar{X}_j - d^*)^-, (\bar{X}_i - \min_{j \neq i} \bar{X}_j + d^*)^+], \text{ for } i = 1, 2, \ldots, k] \geq P^*,$$

where $(x)^- = \min(0, x)$ and $(x)^+ = \max(0, x)$. These confidence intervals bound the difference between each design and the best of the others.

3 STATISTICAL ANALYSIS

In this section we examine the relationship between the sample-size allocation strategy of Rinott's indifference-zone selection procedure and the CI half-width and provide techniques to improve the efficiency of R&S procedures.

3.1 Multiple Comparisons with a Control

Multiple comparisons with a control (with design i_1 as the control) provides simultaneous CIs for the parameters $\mu_{i_l} - \mu_{i_1}$, for $l = 2, 3, \ldots, k$. These CIs bound the difference between the performance of each design and design i_1 with

a prespecified confidence level. The probability of correct selection is estimated by

$$\begin{aligned} \text{P(CS)} &= \text{P}[\bar{X}_{i_1} < \bar{X}_{i_l}, \text{ for } l = 2, 3, \ldots, k] \\ &= \text{P}[\frac{\bar{X}_{i_1} - (\bar{X}_{i_l} - d^*)}{\sqrt{\sigma_{i_l}^2/N_{i_l} + \sigma_{i_1}^2/N_{i_1}}} < \\ &\quad \frac{d^*}{\sqrt{\sigma_{i_l}^2/N_{i_l} + \sigma_{i_1}^2/N_{i_1}}}, \text{ for } l = 2, 3, \ldots, k]. \end{aligned}$$

Let

$$Z_{i_l} = \frac{\bar{X}_{i_1} - (\bar{X}_{i_l} - d^*)}{\sqrt{\sigma_{i_l}^2/N_{i_l} + \sigma_{i_1}^2/N_{i_1}}}$$

and

$$Q_{i_l} = \frac{h}{\sqrt{\sigma_{i_l}^2/S_{i_l}^2(r) + \sigma_{i_1}^2/S_{i_1}^2(r)}}, l = 2, 3, \ldots, k.$$

Since $N_i \geq (h/d^*)^2 S_i^2(r)$ for all i,

$$\frac{d^*}{\sqrt{\sigma_{i_l}^2/N_{i_l} + \sigma_{i_1}^2/N_{i_1}}} \geq Q_{i_l} \text{ for } l = 2, 3, \ldots, k. \quad (2)$$

Moreover, the variables $Y_i = (n_0 - 1)S_i^2(n_0)/\sigma_i^2$, for $i = 1, 2, \ldots, k$ are independent χ^2 variables with $n_0 - 1$ degrees of freedom, so the variables Q_{i_l}, for $l = 2, 3, \ldots, k$, have the same distribution. Let $\Phi(x)$ be the cdf (cumulative distribution function) of the standard normal distribution and $Q_{i_l} = Q$, for $l = 2, 3, \ldots, k$. Then, under the LFC,

$$\text{P(CS)} \geq \text{P}[Z_{i_l} < Q_{i_l}, l = 2, 3, \ldots, k] \geq \text{E}(\Phi^{k-1}(Q)).$$

The first inequality follows from (2) and the second inequality follows from Slepian's inequality (Tong 1980) since Z_{i_l}'s are positively correlated. In Rinott's procedure, the critical value h is computed such that $\text{E}(\Phi^{k-1}(Q)) = P^*$. Furthermore, (2) can be rewritten as

$$d^* \geq Q\sqrt{\sigma_{i_l}^2/N_{i_l} + \sigma_{i_1}^2/N_{i_1}} \text{ for } l = 2, 3, \ldots, k. \quad (3)$$

When testing the null hypothesis $\text{H}_0 : \mu_{i_l} \leq \mu_{i_1}$, the test statistic that will be used to make a decision whether or not to reject the null hypothesis is

$$z = \frac{\bar{X}_{i_l} - \bar{X}_{i_1}}{\sigma_{\bar{X}_{i_l} - \bar{X}_{i_1}}},$$

where $\sigma^2_{\bar{X}_{i_l} - \bar{X}_{i_1}}$ is the variance of the values $\bar{X}_{i_l} - \bar{X}_{i_1}$. Let $z_{1-\alpha}$ denote the $1 - \alpha$ quantile of the standard normal distribution. We reject the null hypothesis only if $z > z_{1-\alpha}$, or similarly

$$\bar{X}_{i_l} - \bar{X}_{i_1} > z_{1-\alpha}\sigma_{\bar{X}_{i_l} - \bar{X}_{i_1}} = w_{i_l},$$

where w_{i_l} is the one-tailed $1-\alpha$ CI half width. By definition, w_{i_l} ensures $\text{P}[\mu_{i_l} - \mu_{i_1} \geq \bar{X}_{i_l} - \bar{X}_{i_1} - w_{i_l}] \geq 1 - \alpha$. Moreover, for us to conclude with confidence $1 - \alpha$ that $\mu_{i_l} > \mu_{i_1}$ the lower endpoint of the one-tailed $1 - \alpha$ CI must be positive, i.e., $\bar{X}_{i_l} - \bar{X}_{i_1} - w_{i_l} > 0$. For details on the duality of confidence intervals and hypothesis tests see Rice (1995).

By symmetry of the normal distribution,

$$\text{P}[(\bar{X}_{i_l} - \bar{X}_{i_1}) + w_{i_l} \geq \mu_{i_l} - \mu_{i_1}] \geq 1 - \alpha. \quad (4)$$

The simultaneous one-tailed P^* CIs half-width, with design i_1 as a control, is

$$w_{i_l} = z_h\sqrt{\sigma_{i_l}^2/N_{i_l} + \sigma_{i_1}^2/N_{i_1}},$$

where z_h is a critical value such that $\text{E}(\Phi^{k-1}(z_h)) = P^*$. To achieve

$$\text{P}[\bar{X}_{i_l} - \bar{X}_{i_1} > 0] \geq 1 - \alpha,$$

the sample sizes N_i should be large enough so that $\mu_{i_l} - \mu_{i_1} > w_{i_l}$. Note that the half-width w_{i_l} depends on the sample sizes.

Since $Q = z_h$, it follows from (3) that under the LFC the sample sizes determined by Rinott's procedure guarantee

$$d^* \geq w_{i_l} \text{ for } l = 2, 3, \ldots, k.$$

Consequently,

$$\text{P}[\mu_i - \mu_{i_1} \in [\bar{X}_i - \bar{X}_{i_1} - d^*, \infty] \text{ for } i \neq i_1] \geq P^*,$$

$$\text{P}[\mu_i - \mu_{i_1} \in [-\infty, \bar{X}_i - \bar{X}_{i_1} + d^*] \text{ for } i \neq i_1] \geq P^*,$$

and

$$\text{P}[\mu_i - \mu_{i_1} \in [\bar{X}_i - \bar{X}_{i_1} - d^*, \bar{X}_i - \bar{X}_{i_1} + d^*] \text{ for } i \neq i_1] \geq 2P^* - 1.$$

Let the set $I = \{1, 2, \ldots, k\}$ includes all k designs under consideration. Since $\mu_{i_1} - \mu_{i_1} (= 0)$ is within the above three CIs with probability 1 and $\mu_i - \mu_{i_1} \geq 0$ for all i, we have the following result.

Theorem 1 *Under the same assumptions for indifference-zone procedures, the sample sizes determined*

by Rinott's indifference-zone selection procedure also guarantee with high confidence that

1. *The simultaneous lower one-tailed confidence intervals,*

$$P[\mu_i - \mu_{i_1} \in [(\bar{X}_i - \bar{X}_{i_1} - d^*)^+, \infty], \forall i] \geq P^*.$$

2. *The simultaneous upper one-tailed confidence intervals,*

$$P[\mu_i - \mu_{i_1} \in [0, \bar{X}_i - \bar{X}_{i_1} + d^*], \forall i] \geq P^*.$$

3. *The simultaneous two-tailed confidence intervals,*

$$P[\mu_i - \mu_{i_1} \in [(\bar{X}_i - \bar{X}_{i_1} - d^*)^+, \bar{X}_i - \bar{X}_{i_1} + d^*], \forall i] \geq 2P^* - 1.$$

Furthermore, $\bar{X}_b \leq \bar{X}_{i_1}$, where $\bar{X}_b = \min_{i \in I} \bar{X}_i$. Therefore,

$$\begin{aligned} &\mathrm{P}[\mu_i - \mu_{i_1} \in [(\bar{X}_i - \bar{X}_{i_1} - d^*)^+, \infty]] \geq \\ &\quad \mathrm{P}[\mu_i - \mu_{i_1} \in [(\bar{X}_i - \bar{X}_b - d^*)^+, \infty]] \end{aligned}$$

and

$$\begin{aligned} &\mathrm{P}[\mu_i - \mu_{i_1} \in [0, \bar{X}_i - \bar{X}_{i_1} + d^*]] \leq \\ &\quad \mathrm{P}[\mu_i - \mu_{i_1} \in [0, \bar{X}_i - \bar{X}_b + d^*]]. \end{aligned}$$

In practice we do not know which design is the true best and can substitute the sample mean of the unknown best with the best sample mean to construct upper one-tailed CIs.

We, therefore, have the following result.

Theorem 2 *Under the same assumptions for indifference-zone procedures, the sample sizes determined by Rinott's indifference-zone selection procedure also guarantee with high confidence that the difference between each design and the unknown best is bounded by the difference between the sample mean of each design and the best sample mean plus the indifference amount. That is,*

$$P[\mu_i - \mu_{i_1} \in [0, \bar{X}_i - \bar{X}_b + d^*], \forall i] \geq P^*,$$

where $\bar{X}_b = \min_{i \in I} \bar{X}_i$.

Since $b \in I$, Theorem 2 implies

$$\begin{aligned} &\mathrm{P}[\mu_b - \mu_{i_1} \in [0, \bar{X}_b - \bar{X}_b + d^*]] \geq P^*, \textit{i.e.}, \\ &\mathrm{P}[\mu_b - \mu_{i_1} \leq d^*] \geq P^*. \end{aligned}$$

Once the procedure has selected the design b, the event $\mu_b - \mu_{i_1} \leq d^*$ is either true or false. However, if we try this selection procedure many times (approaching infinite), the frequency of the event is true will be greater than or equal to P^*. In words, the selected design b will be within d^* of the best design with high confidence, as the indifference-zone R&S procedures advertised.

3.2 Multiple Comparisons with the Best

Multiple comparisons with the best provides simultaneous CIs for the parameter $\mu_i - \min_{j \neq i} \mu_j$, $i = 1, 2, \ldots, k$. These CIs bound the difference between the performance of each design and the best of the others with a prespecified confidence level. We follow the discussion of Nakayama (1997) to construct MCB intervals. Define the events

$$E = \{\mu_i - \mu_{i_1} \leq \bar{X}_i - \bar{X}_{i_1} + d^*, \forall i \neq i_1\},$$

$$E_L = \{\mu_i - \min_{j \neq i} \mu_j \geq (\bar{X}_i - \min_{j \neq i} \bar{X}_j - d^*)^-, \forall i\},$$

$$E_U = \{\mu_i - \min_{j \neq i} \mu_j \leq (\bar{X}_i - \min_{j \neq i} \bar{X}_j + d^*)^+, \forall i\},$$

$E_T = \{\mu_i - \min_{j \neq i} \mu_j \in [(\bar{X}_i - \min_{j \neq i} \bar{X}_j - d^*)^-, (\bar{X}_i - \min_{j \neq i} \bar{X}_j + d^*)^+], \forall i\}$.

Note that E is the event that the upper one-tailed confidence intervals for multiple comparison with a control, with the control being design i_1, contain all of the true difference $\mu_i - \mu_{i_1}$. From the second item of Theorem 1 in Section 3.1, we know that $\mathrm{P}[E] \geq P^*$. Now following an argument developed by Edwards and Hsu (1983), we have that $E \subset E_L \cap E_U$, which will establish the result $\mathrm{P}[E_T] \geq P^*$.

First we prove that $E \subset E_L$:

$$\begin{aligned} E &\subset \{\mu_{i_1} - \mu_j \geq \bar{X}_{i_1} - \bar{X}_j - d^*, \forall j \neq i_1\} \\ &\subset \{\mu_{i_1} - \mu_{i_2} \geq \bar{X}_{i_1} - \bar{X}_j - d^*, \forall j \neq i_1\} \\ &\subset \{\mu_i - \mu_{i_2} \geq (\bar{X}_i - \min_{j \neq i} \bar{X}_j - d^*)^-, \forall i\} \\ &\subset \{\mu_i - \min_{j \neq i} \mu_j \geq (\bar{X}_i - \min_{j \neq i} \bar{X}_j - d^*)^-, \forall i\}, \end{aligned}$$

where the second step follows since $\mu_{i_1} - \mu_{i_2} \geq \mu_{i_1} - \mu_j$ for all $j \neq i_1$ and the third step follows since $\mu_i - \mu_{i_2} \geq 0$ for all $i \neq i_1$ and $(x)^- \leq 0$.

Now we show $E \subset E_U$.

$$\begin{aligned} E &\subset \{\mu_i - \mu_{i_1} \leq \bar{X}_i - \min_{j \neq i} \bar{X}_j + d^*, \forall i \neq i_1\} \\ &\subset \{\mu_i - \min_{j \neq i} \mu_j \leq (\bar{X}_i - \min_{j \neq i} \bar{X}_j + d^*)^+, \forall i\}, \end{aligned}$$

where the first step follows since $\min_{j \neq i} \bar{X}_j \leq \bar{X}_{i_1}$ for all $i \neq i_1$ and the last step follows since $\mu_{i_1} - \min_{j \neq i_1} \mu_j \leq 0$ and $(x)^+ \geq 0$. Hence, $E \subset E_L \cap E_U$, and the proof is complete. These MCB CIs are the same as those established in Nelson and Matejcik (1995).

3.3 The Adjusted ETSS Procedure

It is informative that Rinott's selection procedure also guarantees MCB CIs with half-width corresponding to the indifference amount, but these CIs come at a cost. If the objective is to select the best design instead of estimating the difference of sample means, then the sample sizes only need to be large enough so that the half-widths are smaller than the differences between means; see (4). Chen and Kelton (2000) propose the ETSS procedure that takes into account not only the sample variances, but also the difference between the sample means across designs. The sample-size allocation strategy of the ETSS procedure is similar to that of Chen et al. (2000), who show that with a fixed computational budget the probability of selecting the best design can be asymptotically maximized when the allocated sample sizes satisfy certain ratios. Chen (2002) refines the ETSS procedure by adding a conservative adjustment to increase P(CS).

Let $\bar{X}_b(n_0) = \min_{i=1}^{k} \bar{X}_i(n_0)$, $U(\bar{X}_b(n_0))$ be the upper one-tailed P^* confidence limit of μ_b, and

$$d_i = \max(d^*, \bar{X}_i - U(\bar{X}_b(n_0))).$$

The adjusted ETSS procedure computes the number of required simulation replications or batches for each design based on the following formula

$$N_i = \max(n_0, \lceil (hS_i(n_0)/d_i)^2 \rceil), \text{ for } i = 1, \ldots, k. \quad (5)$$

The Adjusted ETSS Algorithm:

1. Simulate n_0 replications.
2. For each design i, compute the needed number of additional replications $N_i - n_0$. Here N_i will be computed according to (5).
3. Simulate $N_i - n_0$ additional replications for each design i.
4. Return the values b and $\bar{X}_b$, where $\bar{X}_b = \min_{i=1}^{k} \bar{X}_i$.

The difference between (5) and (1) is that d_i is being used instead of d^*. The differences between the sample means are embedded in d_i; consequently, this procedure will allocate fewer replications or batches to the less promising design i, whose sample mean $\bar{X}_i$ are far in excess of $\bar{X}_b$. Let $w_{i,b}$ denote the one-tailed $(P^*)^{\frac{1}{k-1}}$ CI half-with of $\mu_i - \mu_b$. Following the discussion of Section 3.1, the sample sizes determined by the adjusted ETSS procedure should guarantee

$$d_i \geq w_{i,b} \ \forall i.$$

Subset pre-selection is a screening device that attempts to improve the efficiency of the selection procedures by selecting a (random-size) subset of the k alternative designs that contains the best design. Inferior designs are excluded from further simulation to reduce the overall computational efforts. Chen (2001) shows that the ETSS procedure has an intrinsic subset pre-selection built-in, so a separate subset pre-selection process is not needed.

3.4 Multiple Comparison Confidence Intervals of the Adjusted ETSS

Using the same reasoning in Sections 3.1 and 3.2, multiple comparison CIs can also be constructed with the outcomes of the adjusted enhanced two-stage selection procedure. The simultaneous upper one-tailed confidence intervals,

$$\mathrm{P}[\mu_i - \mu_{i_1} \in [0, \bar{X}_i - \bar{X}_{i_1} + w_{i,i_1}], \forall i] \geq P^*.$$

Theorem 3 *Under the same assumptions for indifference-zone procedures, the sample sizes determined by the adjusted enhanced two-stage selection procedure also guarantee with high confidence that*

$$P[\mu_i - \mu_{i_1} \in [0, \max_{j \neq i}(\bar{X}_i - \bar{X}_j + w_{i,j})^+], \forall i] \geq P^*.$$

Theorem 4 *Under the same assumptions for indifference-zone procedures, the sample sizes determined by the adjusted enhanced two-stage selection procedure also guarantee with high confidence that multiple comparisons with the best confidence interval coverage probability will be at least P^* with the half-width of the confidence interval corresponding to the difference between the sample means. That is,* $P[\mu_i - \min_{j \neq i} \mu_j \in [\max_{j \neq i}(\bar{X}_i - \bar{X}_j - w_{i,j})^-, \max_{j \neq i}(\bar{X}_i - \bar{X}_j + w_{i,j})^+], \forall i] \geq P^*$.

Since the sample sizes should be large enough to ensure $w_{i,b} \leq d_i$, the sample sizes should also be large enough to ensure $w_{i,j} \leq \max(d_i, d_j)$. Hence, the CI half-width $w_{i,j}$ in Theorems 3 and 4 can be approximated by $\max(d_i, d_j)$. Furthermore, if $d_i = d^*$ for all i, i.e., under the LFC, then the MCC (with i_1 as the control) equation in Theorem 3 can be simplified to $\mathrm{P}[\mu_i - \mu_{i_1} \in [0, \bar{X}_i - \bar{X}_b + d^*], \forall i] \geq P^*$ and the MCB equation in Theorem 4 can be simplified to $\mathrm{P}[\mu_i - \min_{j \neq i} \mu_j \in [(\bar{X}_i - \min_{j \neq i} \bar{X}_j - d^*)^-, (\bar{X}_i - \min_{j \neq i} \bar{X}_j + d^*)^+], \forall i] \geq P^*$.

3.5 Using Common Random Numbers

We can use common random numbers to improve P(CS) without any further assumptions. Let $\mathrm{P}_I(\mathrm{CS})$ denote the probability of correct selection with independent sampling, $\mathrm{P}_C(\mathrm{CS})$ denote P(CS) with CRN, event E_l, for $l = 2, 3, \ldots, k$, denote $\bar{X}_{i_l} - \bar{X}_{i_1} > 0$, and let $\mathrm{P}_I(E_l)$ and $\mathrm{P}_C(E_l)$ denote the probability of event E_l with independent sampling and with CRN, respectively. With independent

sampling across alternatives, the E_l's are positively correlated and by Slepian's inequality (Tong 1980)

$$\mathrm{P}_I(\mathrm{CS}) = \mathrm{P}_I[E_l \text{ for } l = 2, 3, \ldots, k] \geq \prod_{l=2}^{k} \mathrm{P}_I[E_l].$$

Equality holds for $k = 2$, and for $k > 2$ strict inequality holds. Using the notions defined in Section 3.1, under the LFC, $\mathrm{P}[E_l] = \Phi(Q_{i_l})$, and $Q_{i_l} = Q$. Since the critical constant h ensures $\mathrm{E}(\Phi^{k-1}(Q)) = P^*$, after the constant h is assigned a numeric value, $\Phi(Q) = (P^*)^{\frac{1}{k-1}}$ and P(CS) $\geq P^*$.

It is known that in some cases

$$\mathrm{P}_C[E_l \text{ for } l = 2, 3, \ldots, k] < \prod_{l=2}^{k} \mathrm{P}_C[E_l].$$

However,

$$\prod_{l=2}^{k} \mathrm{P}_I[E_l] \leq \prod_{l=2}^{k} \mathrm{P}_C[E_l],$$

so the following may still hold

$$\mathrm{P}_C[E_l \text{ for } l = 2, 3, \ldots, k] \geq \mathrm{P}_I[E_l \text{ for } l = 2, 3, \ldots, k].$$

That is, $\mathrm{P}_C(\mathrm{CS})$ is still greater than or equal to $\mathrm{P}_I(\mathrm{CS})$. However, it has not been proved that $\mathrm{P}_C(\mathrm{CS})$ is always larger than or equal to $\mathrm{P}_I(\mathrm{CS})$.

By the *Bonferroni* inequality (Law and Kelton 2000)

$$\mathrm{P(CS)} = \mathrm{P}[E_l, \text{ for } l = 2, 3, \ldots, k] \geq 1 - \sum_{l=2}^{k} (1 - \mathrm{P}[E_l]).$$

If we find the constant h with

$$P = (1 - \frac{1 - P^*}{k - 1})^{k-1},$$

then the sample sizes will guarantee $\mathrm{P}[E_l] \geq 1 - (1 - P^*)/(k-1)$ for $l = 2, 3, \ldots, k$. Thus, $\sum_{l=2}^{k}(1 - \mathrm{P}[E_l]) \leq 1 - P^*$ and P(CS) $\geq P^*$. Hence, we can use CRN to increase $\mathrm{P}[E_l]$ for $l = 2, 3, \ldots, k$ and P(CS) without any further assumptions. For example, if $k = 10$ and we want to have P(CS) ≥ 0.95, we use

$$P = (1 - \frac{1 - P^*}{k - 1})^{k-1} = 0.951097$$

to find the constant h. If h is obtained with $P^* = 0.95$ and $k = 10$, we state that P(CS) ≥ 0.948852, i.e., $1 - (k - 1)(1 - (P^*)^{1/(k-1)})$. Similarly, we can use CRN with the ETSS and its variants to improve P(CS).

3.6 Sequentializing the Adjusted ETSS Procedure

Chen and Kelton (2003) propose sequentializing the ETSS procedure to eliminate the drawback of two-stage procedures and to improve its efficiency. Rinott's procedure and its variants are based on $\mathrm{P(CS)}=\mathrm{P}[\bar{X}_{i_l} > \bar{X}_{i_1}, \text{ for } l = 2, 3, \ldots, k] \geq P^*$ and $\mathrm{P}[\bar{X}_{i_l} > \bar{X}_{i_1}] \geq (P^*)^{\frac{1}{k-1}}$, for $l = 2, 3, \ldots, k$. To further improve the efficiency of sequentialized ETSS procedure they perform all pairwise comparisons at each iteration. Inferior designs i such that $\mathrm{P}[\bar{X}_i > \bar{X}_j] \geq P^{1/(k-1)}$ for some design j will be excluded from further simulation at each iteration.

Since it is computationally intensive to perform all pairwise comparisons when the number of designs under consideration is large, users can modify the procedure to perform pairwise comparisons between design i and the best b designs or when the number of designs under contention has been reduced to a pre-determined number.

The Sequentialized Adjusted ETSS Procedure:

1. Initialize the set I to include all k designs. Simulate $r = n_0$ replications or batches for each design $i \in I$. Set the iteration number $j = 0$, and $N_{1,j} = N_{2,j} = \ldots = N_{k,j} = n_0$, where $N_{i,j}$ is the sample size allocated for design i at the j^{th} iteration. Let $\bar{X}_{i,j}$ denote the sample mean of design i at the j^{th} iteration.
2. Let $\bar{X}_{b,j} = \min_{i \in I} \bar{X}_{i,j}$. For all $i \in I$, compute $\hat{d}_{i,j} = \max(d^*, \bar{X}_{i,j} - U(\bar{X}_{b,j}))$, where $U(\bar{X}_{b,j})$ is the upper one-tailed P^* confidence limit of μ_b at the j^{th} iteration, and compute
$$\delta_{i,j+1} = \lceil ((hS_i(r))/\hat{d}_{i,j})^2 - r)^+ \rceil.$$
3. Set $j = j + 1$ and the incremental sample size at the j^{th} iteration $\delta_j = \min_{i \in I}\{\delta_{i,j} | \delta_{i,j} > 0\}$.
4. If $i \neq b$ and $\delta_{i,j} = 0$, delete design i from I.
5. Perform all pairwise comparisons and delete inferior design i from I.
6. For all $i \in I$, simulate additional δ_j samples and set $r = r + \delta_j$. If there is more than one element (or the pre-determined number of best designs) in I, go to step 2.
7. Return the values b and $\bar{X}_b$, where $\bar{X}_b = \min \bar{X}_i$, $1 \leq i \leq k$ and i was not eliminated by all pairwise comparisons.

In the sequential procedure, all the alternatives $1 \leq i \leq k$ are included in the set I initially for R&S. If all $k - 1$ designs were eliminated from I through the two-sample-t test, then $\mathrm{P}[i_1 \in I] \geq P^*$. On the other hand, if some designs were eliminated from I because its required additional sample size at the j^{th} iteration $\delta_{i,j} = 0$, then the procedure can only guarantee P(CS) $\geq P^*$. We use the equation $S_i^2(r) = (\sum_j^r X_{ij}^2/r - \bar{X}_i^2(r))r/(r-1)$ to compute the sample variance so that we are only required to store the

Table 1: $\hat{P}$(CS) for Experiment 1

	$P^* = 0.90$		$P^* = 0.95$	
n_0	20	30	20	30
$\hat{P}$(CS)	0.9866	0.9868	0.9956	0.9939
PSC	0.9390	0.9355	0.9674	0.9677
PC2	1.0000	1.0000	1.0000	1.0000
PC3	0.9886	0.9896	0.9941	0.9954
PC4	0.9912	0.9901	0.9953	0.9951
PC5	0.9886	0.9880	0.9942	0.9950
PC6	0.9892	0.9892	0.9957	0.9949
PC7	0.9912	0.9883	0.9931	0.9958
PC8	0.9876	0.9904	0.9948	0.9947
PC9	0.9886	0.9882	0.9945	0.9935
PC10	0.9899	0.9872	0.9948	0.9948

triple $(r, \sum_{j=1}^{r} X_{ij}, \sum_{j}^{r} X_{ij}^2)$, instead of the entire sequence $(X_{i1}, X_{i2}, \ldots, X_{ir})$.

4 EMPIRICAL EXPERIMENTS

In this section we present some empirical results obtained from simulations. Instead of using systems simulation examples, which offer less control over the factors that affect the performance of a procedure, we use various normally distributed random variables to represent the systems. Since the MCB CIs and the effects of using CRN with selection procedures are well known, we focus on the MCC CIs.

4.1 Experiment 1: Rinott Procedure

There are ten alternative designs under consideration. Suppose $X_{ij} \sim \mathcal{N}(i, 6^2)$, $i = 1, 2, \ldots, 10$, where $\mathcal{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. We want to select a design with the minimum mean: design 1. The indifference amount d^* is set to 1.00 for all cases. We use two different initial replications $n_0 = 20$ and 30. Furthermore, 10,000 independent experiments are performed to obtain the proportion of all $k-1$ CIs contain the true value simultaneously; we denote this event as PSC.

The results of experiment 1 are in Table 1. We list the observed $\hat{P}$(CS) (the proportion of the design 1 is selected), PSC, and the proportion of individual CIs that contain the true value, i.e., PCl = P$[\mu_{i_l} - \mu_{i_1} \in [0, \bar{X}_{i_l} - \bar{X}_b + d^*]]$, for $l = 2, 3, \ldots, k$, which should be greater than or equal to the nominal value of $(P^*)^{\frac{1}{k-1}}$, i.e., 0.988362 $(0.90^{\frac{1}{9}})$ and 0.994317 $(0.95^{\frac{1}{9}})$ for $P^* = 0.90$ and 0.95, respectively. Note that PC1 = P$[\mu_{i_1} - \mu_{i_1} \in [0, \bar{X}_{i_1} - \bar{X}_b + d^*]] = 1$.

Since Rinott's procedure is based on the LFC, it is conservative. The observed $\hat{P}$(CS)'s are all greater than the nominal values. The proportion of these $k-1$ CIs that contain the true value is greater than P^* even though some of the proportion of the individual CIs that contain the true value is less than $(P^*)^{\frac{1}{k-1}}$. Note that even when the CI for $\mu_i - \mu_{i_i}$ does not contain the true value, it is because $\mu_i - \mu_{i_1} > \bar{X}_i - \bar{X}_b + d^*$ and we may still have $\bar{X}_i > \bar{X}_{i_1}$.

Table 2: $\hat{P}$(CS) for Experiment 2

	$P^* = 0.90$		$P^* = 0.95$	
n_0	20	30	20	30
$\hat{P}$(CS)	0.9866	0.9868	0.9956	0.9939
PP2	0.9866	0.9868	0.9956	0.9939
PP3	1.0000	1.0000	1.0000	1.0000
PP4	1.0000	1.0000	1.0000	1.0000
PP5	1.0000	1.0000	1.0000	1.0000
PP6	1.0000	1.0000	1.0000	1.0000
PP7	1.0000	1.0000	1.0000	1.0000
PP8	1.0000	1.0000	1.0000	1.0000
PP9	1.0000	1.0000	1.0000	1.0000
PP10	1.0000	1.0000	1.0000	1.0000

4.2 Experiment 2: Adjusted ETSS Procedure

In this experiment, we use the adjusted ETSS procedure to perform R&S. All the settings are kept the same as in experiment 1. Since the upper confidence limit $\max_{j \neq i}(\bar{X}_{i_l} - \bar{X}_j + \max(d_{i_l}, d_j))$ is likely to be large so that PCl = P$[\mu_{i_l} - \mu_{i_1} \in [0, \max_{j \neq i}(\bar{X}_{i_l} - \bar{X}_j + \max(d_{i_l}, d_j))]]$, for $l = 2, 3, \ldots, k$ will be high, we list the proportion of individual correct pairwise comparisons with the control, i.e., PPl = P$[\bar{X}_{i_l} > \bar{X}_{i_1}]$, for $l = 2, 3 \ldots, k$. The results of experiment 2 are summarized in Table 2. We list the observed $\hat{P}$(CS) and PPl for $l = 2, 3 \ldots, k$, which should be greater than or equal to $(P^*)^{\frac{1}{k-1}}$, i.e., 0.988362 and 0.994317 for $P^* = 0.90$ and 0.95, respectively. Note that PP1 = P$[\bar{X}_{i_1} > \bar{X}_{i_1}] = 0$.

PPl for $l = 3, 4, \ldots, 10$ are 1.0. Since $\mu_{i_2} = \mu_{i_1} + d^*$ and $d_{i_2} \geq d^*$, the PP2's are just below the nominal value $(P^*)^{\frac{1}{k-1}}$. Furthermore, PP2 and $\hat{P}$(CS) have the same values, indicating when the procedure makes a wrong selection, it selects design 2. Even though the sample sizes allocated for inferior designs whose sample means $\bar{X}_i$ are far in excess of $\bar{X}_b$ are smaller relative to Rinott's procedure, the adjusted ETSS procedure obtains higher than the specified nominal value for MCC. The adjusted ETSS procedure can significantly improve the efficiency of Rinott's procedure when the objective is to select the best design.

4.3 Experiment 3: Sequentialized Adjusted ETSS

In this section, we present some empirical results obtained from simulations using the Rinott, ETSS, ETSS$_a$ (adjusted ETSS with $U(\bar{X}_b)$ being the upper $100P^*$% confidence limit), SARS (sequentialized adjusted ETSS without multiple comparisons, i.e., step 5 of the algorithm in Section 3.6 is not performed), and SAMC (SARS with multiple comparisons). We considered only the LFC,

Table 3: $\hat{P}$(CS) and Sample Sizes for Experiment 3

	$P^* = 0.90$		$P^* = 0.95$	
Procedure	$\hat{P}$(CS)	$\overline{T}$	$\hat{P}$(CS)	$\overline{T}$
Rinott(20)	0.9326	4259	0.9650	5412
ETSS(20)	0.6834	1820	0.7318	2317
ETSS$_a$(20)	0.8735	3347	0.9346	4640
SARS(20)	0.9529	3840	0.9800	5165
SAMC(20)	0.9363	2731	0.9705	3820
Rinott(30)	0.9320	4057	0.9655	5120
ETSS(30)	0.7662	2013	0.8029	2516
ETSS$_a$(30)	0.8976	3326	0.9530	4520
SARS(30)	0.9475	3655	0.9773	4872
SAMC(30)	0.9338	2730	0.9742	3705

$\mu_1 + d^* = \mu_2 = \ldots = \mu_{10}$, because the minimum P(CS) should occur at this configuration, where $\mu_1 = 0$ and $d^* = 1$. The variances are $\sigma_i^2 = 6^2$ for $i = 1, 2 \ldots, k$. We compute $\hat{P}$(CS) of those procedures using two different initial replications $n_0 = 20$ and 30. As discussed in Section 3.5, we suggest using P instead of P^* when finding the critical constant h. However, since the constant h is listed only for several commonly used probabilities, ETSS and its variants use P^* to approximate $P = (1-\frac{1-P^*}{k-1})^{k-1}$ when finding the critical constant h in this experiment. The approximations are generally close to their true values. For example, when $k = 10$ and $P^* = 0.90$, we have $P \approx 0.904333$.

Table 3 lists the results of experiment 3. The $\hat{P}$(CS) columns list the proportion of correct selection. The $\overline{T}$ column lists the average of the number of total simulation replications ($\overline{T} = \sum_{R=1}^{10000} \sum_{i=1}^{k} N_{R,i}/10000$, and $N_{R,i}$ is the total number of replications or batches for design i in the R^{th} independent run) used in each procedure. The Rinott(20), ETSS(20), ETSS$_a$(20), SARS(20), and SAMC(20) rows list the results of the respective procedure with the initial sample size $n_0 = 20$ (and similarly for $n_0 = 30$). Rinott's procedure is conservative, but when we encounter the LFC, it is very effective. In this setting, i.e., under the LFC and with large variances, the observed $\hat{P}$(CS)'s of the ETSS procedure are less than the specified P^*. We don't think this is a major drawback of the ETSS procedure since we rarely encounter the LFC and it may not be too costly to select design j whose μ_j is equal to or just greater than $\mu_{i_1} + d^*$. The conservative adjustment effectively increases $\hat{P}$(CS). Since ETSS and adjusted ETSS take into account sample means, they have better performance with a larger initial sample sizes. The sequentialized procedures eliminate the drawback of relying too heavily on information obtained in just one stage and achieve higher $\hat{P}$(CS) than do the two-stage procedures. Among these five procedures tested, SAMC is the most effective in achieving high $\hat{P}$(CS) with small sample sizes.

5 CONCLUSIONS

We have provided the statistical analysis of MCC CIs of each design and the unknown best design and MCB CIs of indifference-zone selection procedures. Rinott's indifference-zone selection procedure also guarantees simultaneous MCC and MCB CIs with half-width corresponding to the indifference amount. However, if the objective is to select a good design and not to estimate the difference of sample means, then Rinott's selection procedure is less desirable since these tight CIs come at a cost. The procedures of Chen and Kelton (2000) and Chen (2002) provide an effective enhancement to selecting the best design by taking into account the differences of sample means and can be used when the number of designs under consideration is large. However, the half-width of the CIs for the differences between μ_i and μ_{i_1} is around $\max(d^*, \mu_i - \mu_{i_1})$ instead of d^*. This approach is consistent with the philosophy of ordinal comparison (Ho et al. 1992). That is, in our solution technique we are more interested in whether a given design is better than the others rather than the accuracy of the performance measures. Furthermore, CRN can be used with selection procedures to increase P(CS) without any further assumptions.

To reduce the drawback that two-stage selection procedures rely heavily on the first-stage information, we sequentialized the adjusted ETSS procedure and incorporated all pairwise comparisons to eliminate inferior designs at earlier iterations to improve the overall computational effort as well as the probability of correct selection. Since the sequentialized version of the procedures have much better performance, we strongly recommend using the sequentialized version instead of two-stage procedures.

REFERENCES

Bechhofer, R. E., T. J. Santner, and D. M. Goldsman. 1995. *Design and Analysis of Experiments for Statistical Selection, Screening and Multiple Comparisons*. New York: John Wiley & Sons, Inc.

Chen, E. J. 2001. Using common random numbers with two-stage selection procedures. In *Proceedings of the 2001 Winter Simulation Conference*, ed. B.A. Peters, J.S. Smith, D.J. Medeiros, and M.W. Rohrer, 408–416. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Chen, E. J. 2002. A conservative adjustment to the ETSS procedure. In *Proceedings of the 2002 Winter Simulation Conference*, eds. E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 384–391. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Chen, E. J. and W. D. Kelton. 2000. An enhanced two-stage selection procedure. In *Proceedings of the 2000 Winter*

Simulation Conference, ed. J.A. Joines, R. Barton, P. Fishwick, and K. Kang, 727–735. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Chen, E. J. and W. D. Kelton. 2003. Sequential selection procedures: Using sample means to improve efficiency. Submitted.

Chen, C. H., J. Lin, E. Yücesan, and S. E. Chick. 2000. Simulation budget allocation for further enhancing the efficiency of ordinal optimization. *Journal of Discrete Event Dynamic Systems*, **10**(3), 251–270.

Dudewicz, E. J. and S. R. Dalal. 1975. Allocation of observations in ranking and selection with unequal variances. *Sankhya* B37: 28–78.

Edwards, D. G. and J. C. Hsu. 1983. Multiple comparisons with the best treatment. *Journal of the American Statistical Association*, **78**, 965–971.

Ho, Y. C., R. S. Sreenivas, and P. Vakili. 1992. Ordinal optimization of DEDS. *Journal of Discrete Event Dynamic Systems*, **2**, 61–68.

Kim, S.-H. and B. L. Nelson. 2001. A fully sequential procedure for indifference-zone selection in simulation. *ACM Transactions on Modeling and Computer Simulation*, **11**, 251–273.

Nakayama, M. K. 1997. Multiple-comparison procedures for steady-state simulations. *Annals of Statistics*, **25**, 2433–2450.

Nelson, B. L. and F. J. Matejcik. 1995. Using common random numbers for indifference-zone selection and multiple comparisons in simulation. *Management Science*, **41**, 1935–1945.

Law, A. M. and W. D. Kelton. 2000. *Simulation Modeling and Analysis*. 3rd ed. New York: McGraw-Hill.

Rice, J. A. 1995. *Mathematical Statistics and Data Analysis*, 2nd ed., Duxbury Press, Belmont, California.

Rinott, Y. 1978. On Two-stage selection procedures and related probability inequalities. *Communications in Statistics* A7: 799-811.

Tong, Y. L. 1980. *Probability Inequalities in Multivariate Distributions.* New York: Academic Press.

Wilcox, R. R. 1984. A table for Rinott's selection procedure. *Journal of Quality Technology* Vol. 16, No. 2: 97-100.

AUTHOR BIOGRAPHIES

E. JACK CHEN is a Senior Staff Specialist with BASF Corporation. He received a Ph.D. degree from University of Cincinnati. His research interests are in the area of computer simulation. His email address is <chenej@basf.com>.

W. DAVID KELTON is a Professor in the Department of Quantitative Analysis and Operations Management at the University of Cincinnati. He received a B.A. in mathematics from the University of Wisconsin-Madison, an M.S. in mathematics from Ohio University, and M.S. and Ph.D. degrees in industrial engineering from Wisconsin. His research interests and publications are in the probabilistic and statistical aspects of simulation, applications of simulation, and stochastic models. He is co-author of *Simulation Modeling and Analysis* (3rd ed., 2000, with Averill M. Law), and *Simulation With Arena* (2nd ed., 2002, with Randall P. Sadowski and Deborah A. Sadowski), both published by McGraw-Hill. Currently, he serves as Editor-in-Chief of the *INFORMS Journal on Computing*, and has been Simulation Area Editor for *Operations Research*, the *INFORMS Journal on Computing*, and *IIE Transactions*, as well as Associate Editor for *Operations Research*, the *Journal of Manufacturing Systems*, and *Simulation*. From 1991 to 1999 he was the INFORMS co-representative to the Winter Simulation Conference Board of Directors and was Board Chair for 1998. In 1987 he was Program Chair for the WSC, and in 1991 was General Chair. His email and web addresses are <david.kelton@uc.edu> and <www.cba.uc.edu/faculty/keltonwd>.