

## A SURVEY OF RANKING, SELECTION, AND MULTIPLE COMPARISON PROCEDURES FOR DISCRETE-EVENT SIMULATION

James R. Swisher

Lotus Biochemical Corporation  
7335 Lee Highway  
Radford, VA 24141, U.S.A.

Sheldon H. Jacobson

Department of Mechanical and Industrial Engineering  
University of Illinois at Urbana-Champaign  
Urbana, IL 61801, U.S.A.

### ABSTRACT

Discrete-event simulation models are often constructed so that an analyst may compare two or more competing design alternatives. This paper presents a survey of the literature for two widely-used statistical methods for selecting the best design from among a finite set of  $k$  alternatives: ranking and selection (R&S) and multiple comparison procedures (MCPs). A comprehensive survey of each topic is presented along with a summary of recent unified R&S-MCP approaches. In addition, an example of the application of Nelson and Matejck's (1995) combined R&S-MCP procedure is given.

### 1 INTRODUCTION

The most common goal of discrete-event simulation models is to choose the best system design from among a set of competing alternatives. *Simulation optimization* provides a structured approach to determine optimal input parameters (i.e., the best system design), where optimality is measured by a function of (steady state or transient) output performance measures associated with a simulation model (Jacobson et al. 1999). In the 1970s and 1980s most simulation optimization techniques were applicable when the input parameters to be optimized were continuous. These techniques were path search based, involving gradient estimation techniques (e.g., response surface methodology, perturbation analysis) imbedded in a stochastic approximation algorithm (Jacobson and Schruben 1989). Over the last decade, however, two statistical techniques, ranking and selection (R&S) and multiple comparison procedures (MCPs), have gained popularity in simulation optimization. These techniques are applicable when the input parameters are discrete and the number of designs to be compared is both discrete and small (i.e., 2 to 20). The particular method that is applicable is dependant upon the type of comparison desired by the analyst and the properties of the simulation output data. Jacobson and Schruben (1989), Fu (1994a,

1994b), and Jacobson et al. (1999) provide extensive reviews of simulation optimization techniques including, but not limited to, R&S and MCPs.

R&S procedures are statistical methods specifically developed to select the best system or a subset that contains the best system design from a set of  $k$  competing alternatives (Goldsman and Nelson 1994). In general, these methods ensure the probability of a correct selection at or above some user-specified level. MCPs specify the use of certain pairwise comparisons to make inferences in the form of confidence intervals (Fu 1994a) about relationships among all designs. In short, R&S provides the experimenter with the best system design while MCPs provide information about the relationships among the designs (e.g., how much better the best design is in comparison to the alternatives).

For a general overview of R&S and MCPs, several sources exist. Law and Kelton (1991) provide an excellent introduction to R&S with corresponding references to more mathematically intense treatments. Likewise, Sanchez (1997) gives an overview of R&S with several sample scenarios and an extensive list of references. Goldsman (1983) provides a good perspective of R&S as it stood in relation to simulation in the early 1980s. Wen and Chen (1994) present single-stage sampling procedures for different MCPs. Goldsman and Nelson (1994, 1998) provide comprehensive state-of-the-art reviews of R&S and MCPs in simulation. Where possible, they attempt to unify R&S and MCP perspectives. Goldsman et al. (1991) compares R&S to MCPs and interactive analysis. In addition, two excellent texts are available: Bechhofer et al. (1995) details R&S and provides practical hints for practitioners while Hsu (1996) details the theory and application of MCPs.

The following notation will be used throughout this discussion: Let  $Y_{ij}$  represent the  $j$ th independent and identically distributed (i.i.d.) simulation output (replication or batch mean) of a performance measure from the  $i$ th design alternative, for  $i = 1, 2, \dots, k$  and  $j = 1, 2, \dots, n$ . Let  $\mu_i = E[Y_{ij}]$  denote the expected value of the output from the

$i$ th design alternative and let  $\sigma_i^2 = \text{Var}[Y_{ij}]$  denote its variance. Let  $\mu_{[1]} \leq \mu_{[2]} \leq \dots \leq \mu_{[k]}$  denote the ordered but unknown expected values for the outputs of the  $k$  alternatives. Let  $Y_j = (Y_{1j}, Y_{2j}, \dots, Y_{kj})'$  be the  $k \times 1$  vector of outputs across all design alternatives for output  $j$  and assume that  $Y_1, Y_2, \dots$  are i.i.d. with multivariate normal distribution  $Y_j \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  where  $\boldsymbol{\mu}$  is the unknown mean vector,  $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_k)'$ , and  $\boldsymbol{\Sigma}$  is the unknown variance-covariance matrix. In addition, the use of the subscript “.” indicates averaging with respect to that subscript. For example, the average design alternative output performance measure value across all replications (or batch means) is denoted by  $\bar{Y}_i = \sum_{j=1}^n Y_{ij} / n$ .

The remainder of this paper is organized as follows: Section 2 addresses R&S procedures (indifference zone, subset selection, and other approaches). Section 3 covers MCPs (paired-t, Bonferroni, all-pairwise comparisons; all-pairwise multiple comparisons; multiple comparisons with a control; and multiple comparisons with the best). Section 4 provides insight into R&S-MCP unification, including combined R&S-MCP techniques. Section 5 presents an application of Nelson and Matejcek's (1995) combined procedure. Conclusions and future directions for R&S and MCPs are provided in Section 6.

## 2 RANKING AND SELECTION

Ranking and selection is a commonly prescribed method for selecting the best system from among a set of competing alternatives. The fundamentals for R&S were first proposed by Bechhofer (1954). A majority of the work in R&S can be classified into two general approaches: indifference zone selection and subset selection. Indifference zone selection is treated in Section 2.1, while Section 2.2 treats subset selection approaches. Section 2.3 treats those R&S procedures which fall outside of the two general approaches.

### 2.1 Indifference Zone Procedures

In 1954, Bechhofer introduced the concept of R&S. He describes a problem in which the goal is to select the population with the largest mean for some population statistic from a set of  $k$  normal populations. This population is referred to as the “best.” Typically, an experimenter takes a certain number of observations from each population ( $Y_{ij}$ ) and selects the best population using statistics from these observations. Since the observations are realizations of random variables, it is possible that the experimenter will not select the best population. However, if the best population is selected, the experimenter is then said to have made the correct selection (CS).

In addition, an experimenter may be indifferent (at some level) in the selection of a population when two populations are nearly the same. That is, if  $\mu_{[k]} - \mu_{[k-1]}$  is

very small, then the experimenter may view the populations as essentially the same and not have a preference between the two. To quantify this, define  $\delta$ , the indifference zone. If  $\mu_{[k]} - \mu_{[k-1]} < \delta$ , the experimenter is said to be indifferent to choosing  $\mu_{[k]}$  or  $\mu_{[k-1]}$ . Define the probability of correct selection as  $P\{\text{CS}\} = P\{\mu_{[k]} > \mu_{[i]}, \forall i \neq k \mid \mu_{[k]} - \mu_{[i]} \geq \delta\} \geq P^*$  where  $\{\delta, P^*\}$  are pre-specified by the experimenter. Since  $P\{\text{CS}\} = 1/k$  could be achieved by simply choosing a population at random,  $1/k < P^* < 1$  is required.

The original indifference zone R&S procedure proposed by Bechhofer (1954) assumes unknown means,  $\mu_1, \dots, \mu_k$ , and *known, common* variance,  $\sigma^2$ , for all  $i$  populations ( $i = 1, \dots, k$ ). Bechhofer's procedure is a *single-stage* procedure. That is, the total number of observations required,  $N$ , is determined *a priori* by the experimenter's choice of  $\{\delta, P^*\}$ . Specifically,  $N = \lceil (c_{k,P^*} \sigma / \delta)^2 \rceil$ , where  $c_{k,P^*}$  is a constant defined in a table (see, for example, Bechhofer 1954) and  $\lceil x \rceil$  is the ceiling function of  $x$ . After taking  $N$  independent observations from each of the  $k$  populations, the population with the largest sample mean is selected as the best.

Indifference zone R&S procedures need not be single-stage. Indeed, following Bechhofer (1954), several articles approach the problem as a *multi-stage* problem. That is, the experimenter makes a user-specified number of observations, checks certain stopping criteria and then either continues sampling or stops and selects the best system. Paulson (1964) and Bechhofer et al. (1968) present such methodologies. The major disadvantage of these approaches is the continued requirement for common, known variance among populations. When a simulation analyst is modeling a system that does not physically exist, it is often impossible to know the output performance measure's variance. In addition, modeling an existing system still may not allow the analyst to know the output performance measure's variance because of the potentially high cost or practical infeasibility of data collection. Moreover, even when the variance is known, ensuring common variance across system designs may be difficult. For these reasons, modern indifference zone R&S procedures typically require *neither equal nor known* variances.

Although a small number of articles (e.g., Zinger and St. Pierre 1958 and Zinger 1961) present procedures for *unequal, but known* variances, no articles presented an indifference zone methodology which requires *neither equal nor known* variances until Dudewicz and Dalal in 1975. Their research represented a major step forward in R&S methodology, making the application of indifference zone techniques particularly suitable to discrete-event simulation. They present a two-stage procedure in which the experimenter chooses  $\delta, P^*$ , and  $n_0$  where  $n_0$  is the number of observations to be made during the first stage of the procedure. The first stage variances are then used to

determine the number of second stage observations required. A weighted average of the first and second stage sample means is then used to select the best system (i.e., the system with the largest weighted average). Dudewicz (1976) presents the same procedure with applications to simulation. Rinott (1978) presents a modification to the second stage of Dudewicz and Dalal (1975) that in some cases yields a greater  $P\{CS\}$ , but may prescribe more total observations.

Koenig and Law (1985) extend the indifference zone approach for use as a screening procedure. They present a method for selecting a subset of size  $m$  (user-specified) of the  $k$  systems so that with probability at least  $P^*$ , the selected subset will contain the best system. The procedure for this method requires only the selection of a different table constant when computing the second stage sample size as compared to Dudewicz and Dalal (1975). They also present a slightly different screening procedure with the goal of selecting the  $m$  (user-specified) best of  $k$  systems without regard to rank. This procedure requires minor modifications to the definitions of  $\delta$  and  $P^*$  for implementation.

It is interesting to note that the indifference zone procedures discussed thus far select the best system as a univariate problem (i.e., with a single output measure of system performance). Obviously, an experimenter could weight several output performance measures to form a single measure of effectiveness (see, for example, Morrice et al. 1998, Swisher 1999). However, Dudewicz and Taneja (1978) present a multivariate procedure which does not require reduction to a univariate model. Goldsman (1987) presents extensions of Bechhofer's original work to multivariate (two-factor) models. Further extension of the multivariate procedure for R&S remains an open research area.

Gray and Goldsman (1988) provide an application of indifference zone R&S for choosing the best airspace configuration for a major European airport. Goldsman (1986) also provides a tutorial for those interested in the basics of the indifference zone approach.

## 2.2 Subset Selection Procedures

In contrast to indifference zone procedures, Gupta (1956) presents a procedure for producing a subset of *random size* that contains the best system, with user-specified probability  $P^*$  without the specification of an indifference zone (i.e.,  $\delta = 0$ ). This procedure and others like it are known as subset selection R&S procedures. An application of this procedure to motor vehicle fatality data can be found in Gupta and Hsu (1977). Gupta and Santner (1973) and Santner (1975) extend the original methodology to allow for user specification of a maximum size for the subset. Like the original indifference zone R&S procedures, the original subset selection procedures

required *equal and known* variances among design alternatives. For this reason, subset selection R&S procedures have rarely been applied to discrete-event simulation. However, Sullivan and Wilson (1989) present a procedure that allows *unknown and unequal* variance, as well as the specification of an indifference zone. Although this procedure makes subset selection more attractive for simulation, indifference zone procedures are still more popular. In most cases, an analyst wishes to determine the best system, not identify a subset containing the best (Ho et al. 1992). In addition, if one wishes to identify a subset containing the best, specialized indifference zone procedures allow the *a priori* specification of the subset's size (Koenig and Law 1985).

## 2.3 Other R&S Approaches

As R&S procedures have become more popular, several researchers have attempted to model the problem in a different manner. Chen (1988) and Goldsman (1984a, 1984b) both model the problem as that of selecting the multinomial cell which has the largest underlying probability. Chen focuses on subset selection problems using this model while Goldsman focuses on indifference zone problems. Goldsman (1985) explores the use of standardized time series theory to determine variance estimators for R&S methodologies.

More recently, Chen (1995) and Chen et al. (1996) have formulated the R&S problem as an multi-stage optimization problem in which clearly inferior designs are discarded in the earlier stages. They use Chernoff bounds to estimate the gradient information and then apply the steepest descent method to solve the problem. Chen et al. (1997) extend this work by presenting a different method for estimating the gradient information. Chen et al. (1998) present a further extension of this work that accounts for simulation experiments with different system structures. Through numerical experimentation they find this approach to be more efficient than Chen et al. (1997).

Chick (1997) presents a Bayesian analysis of selecting the best simulated system. Inoue and Chick (1998) compare Bayesian and frequentist approaches for selecting the best system. Chick and Inoue (1998) extend Chick's (1997) work to the study of sampling costs and value of information arguments to improve the computational efficiency of identifying the best system.

## 3 MULTIPLE COMPARISON PROCEDURES

In contrast to R&S procedures in which the goal is to make a decision, the goal of MCPs is to identify the differences between systems' performance (not guarantee a decision). Four general classes of MCPs have been developed: paired- $t$ , Bonferroni, all-pairwise comparisons is discussed in Section 3.1, all-pairwise multiple comparisons (MCA) is

covered in Section 3.2, multiple comparisons with a control (MCC) is discussed in Section 3.3, and Section 3.4 treats multiple comparisons with the best (MCB). Note that MCB is by far the most widely used of these methodologies.

### 3.1 Paired-*t*, Bonferroni, All-Pairwise Comparisons

Fu (1994a) refers to paired-*t*, Bonferroni, all-pairwise comparisons as the brute force approach to MCPs. In this approach, one simply examines all possible pairwise confidence intervals for system designs. That is,  $(\mu_i - \mu_j) \in (\bar{Y}_i - \bar{Y}_j) \pm t_{n-1, \alpha/2} / \sqrt{n}$  for all *i-j* pairs. Here, there will be  $k(k-1)/2$  confidence intervals constructed. Due to the Bonferroni inequality, each confidence interval must be made at level  $(1-\alpha)/[k(k-1)/2]$  in order to have a confidence interval of at least  $(1-\alpha)$  for all intervals simultaneously. Clearly, for any more than 10 alternatives, the width of the individual confidence intervals becomes quite large. Unfortunately, unless there is a clear winner among the systems (i.e., a system with the confidence interval for the difference with all other pairs that is strictly positive), one gains little inference from this procedure.

### 3.2 MCA

MCA has its origins in Tukey (1953) and is similar to the brute-force method, except that instead of constructing separate confidence intervals and using Bonferroni to determine an overall confidence bound, a simultaneous set of confidence intervals at an overall  $(1-\alpha)$  level is formed. Tukey's simultaneous confidence intervals are  $(\mu_i - \mu_j) \in (\bar{Y}_i - \bar{Y}_j) \pm r_{k, k(n-1)}^\alpha s / \sqrt{n}$  for all  $i \neq j$  where  $s$  is the *pooled* sample standard deviation and  $r_{k, k(n-1)}^\alpha$  is the upper  $\alpha$  quantile of the studentized range distribution with  $k$  systems and  $k(n-1)$  degrees of freedom (see tables in Hochberg and Tamhane 1987).

Like the previous method, MCA requires  $k(k-1)/2$  confidence intervals be constructed. In contrast to the brute force method, MCA obtains an overall simultaneous confidence level with the same confidence half-widths for each pairwise comparison, while the brute-force method obtains a different confidence half-width for each pairwise comparison and uses Bonferroni to establish a bound on the overall confidence. Yang and Nelson (1991) provide a revision for MCA which allows the use of common random numbers (CRN) and control variates.

### 3.3 MCC

There are circumstances under which an experimenter wishes to compare a set of alternatives to a pre-defined control. MCC (Dunnnett 1955) is the construction of  $(k-1)$

simultaneous confidence intervals in comparison to a fixed control. This method is particularly useful when one wishes to compare design alternatives to the current design (Bratley et al. 1987). Yang and Nelson (1991) provide a revision for MCC which allows the use of CRN and control variates. Bofinger and Lewis (1992) expand the traditional MCC procedure by describing two-stage MCC procedures. Damerджи and Nakayama (1996) describe a two-stage MCC procedure that allows different probability distributions among alternatives.

### 3.4 MCB

MCB procedures have their origin in Hsu (1984) and Hsu and Nelson (1988). MCB's intent is similar to that of R&S procedures: determine the best system from a set of alternatives. MCB attacks this problem by forming simultaneous confidence intervals on the parameters  $\mu_i - \max_{j \neq i} \mu_j$  for  $i = 1, 2, \dots, k$ . These  $(k-1)$  confidence intervals bound the difference between the expected performance of each system and the best of the other systems. To apply MCB in discrete-event simulation, the simulation runs must be independently seeded and the simulation output must be normally distributed, or averaged so that the estimators used are (approximately) normally distributed.

Yang and Nelson (1989, 1991) and Nelson and Hsu (1993) present modifications to the MCB procedure that incorporate two variance reduction techniques (control variates and CRN). Their results suggest that using variance reduction can lead to correct selections with higher probabilities. Nelson's (1993) robust MCB procedure allows CRN under the assumption the variance-covariance matrix exhibits sphericity (see Section 4). Nakayama (1997b, 1999) presents a two-stage MCB procedure that also uses CRN for steady-state simulations, and shows that it is asymptotically valid. Goldsman and Nelson (1990) present a MCB procedure for steady state simulation experiments. They also present results on how the batch size can impact the probability of a correct selection when using the simulation technique of batch means. Nakayama (1995) presents general results on applying MCB using batch means in steady-state simulations. Yuan and Nelson (1993) discuss MCB procedures for steady-state simulations, where the simulation outputs are assumed to follow an autoregressive process with normally distributed residuals. Damerджи et al. (1996) present a procedure for selecting the best system for transient simulations with known variances. Nakayama (1996, 1997a) presents a single-stage MCB procedure that is asymptotically valid for steady-state simulations, hence extending the domain of applicability of previous MCB procedures. Kolonja et al. (1993) describes an application of MCB using CRN to a truck dispatching system simulation model.

#### 4 UNIFIED PROCEDURES

Recently, there has been an effort to unify the fields of R&S and MCPs. The first reference to such a movement is Gupta and Hsu (1984). They propose a methodology for simultaneously executing R&S and MCB. Matejcik and Nelson (1993, 1995) establish a fundamental connection between indifference zone procedures and MCB. The idea of combining indifference zone approaches with MCB is appealing to the simulation analyst. Such an approach not only selects the best system with pre-specified confidence, but it provides inferences about the relationships between systems which may facilitate decision-making based on secondary criteria that are not reflected in the output performance measure selected.

Nelson and Matejcik (1995) show that most indifference zone procedures can simultaneously provide MCB confidence intervals with the width of the intervals (whisker length) corresponding to the indifference zone. Therefore, both indifference zone selection and MCB inference can be derived from the same experiment with a pre-specified MCB whisker length,  $w = \delta$ . They describe four R&S-MCB procedures which depend on having normally distributed data, but do not require known or equal variance:

1. Rinott's Procedure (Procedure *R*),
2. Dudewicz and Dalal's Procedure (Procedure *DD*),
3. Clark and Yang's Procedure (Procedure *CY*), and
4. Nelson and Matejcik's Procedure (Procedure *NM*).

Procedure *R* is an extension of Rinott's (1978) two-stage indifference zone R&S procedure as described in Section 2. It requires  $n_0$  (where  $n_0$  is the first-stage sample size) i.i.d. samples from each of the  $k$  *independently-simulated* systems. The marginal sample variance for each system is then computed and used to determine the final sample size for each system,  $N_i$  (for  $i = 1, 2, \dots, k$ ). After taking  $N_i - n_0$  additional i.i.d. observations from each of the  $k$  systems, independent of the first-stage samples and independent of the other second-stage samples, the system with the largest overall sample mean is selected as best. In addition, MCB confidence intervals on  $\mu_i - \max_{j \neq i} \mu_j$  are formed. Likewise, Procedure *DD* (based on Dudewicz and Dalal 1975) is performed in the same manner with the only difference being in the calculation of the sample means. While Procedures *R* and *DD* provide both R&S selection and MCB inference, their requirement for independence among all observations precludes the use of CRN. The total sample size required to obtain the desired confidence level is dependent on the sample variances of the systems. In particular, the larger the sample variance, the more

replications (or batch means) required. For this reason, simultaneous R&S-MCB procedures that exploit CRN should require fewer total observations to obtain the same confidence level and whisker length (Nelson and Matejcik 1995).

Procedure *CY* is based upon Clark and Yang's (1986) indifference zone R&S procedure. As one of the few R&S procedures that allows CRN, Clark and Yang (1986) use the Bonferroni inequality to account for the dependence induced by CRN. It is therefore a conservative procedure that typically prescribes more total observations than are actually necessary to make a correct selection (Nelson and Matejcik 1995). Like Procedure *R*, Procedure *CY* is performed in two stages. In the first stage, i.i.d. samples from each of the  $k$  systems are taken *using CRN across systems*. The sample variances of the differences are then used to compute the final sample size,  $N$  (note that  $N$  does not vary across systems). After taking the remaining  $N - n_0$  i.i.d. observations, again using CRN across systems, the system with the largest sample mean is selected as best and the MCB confidence intervals are formed.

Nelson and Matejcik (1995) observe that Procedure *CY* can be effective in reducing the total number of samples required to make a correct selection in comparison with Procedures *R* and *DD*. However, they also note that the benefit gained from using Procedure *CY* is diminished when the number of systems to be compared,  $k$ , is large. This is because the conservatism of the procedure from the Bonferroni inequality increases as  $k$  increases and, at some point, overwhelms the benefit induced by CRN. To overcome this problem, they present Procedure *NM*.

Procedure *NM* is motivated by Nelson's (1993) robust MCB procedure. Since this procedure will be used to illustrate the application of R&S and MCPs in Section 5, it will be treated in greater detail than the previous procedures. Procedure *NM* assumes that the unknown variance-covariance matrix,  $\Sigma$ , exhibits a structure known as *sphericity*. Sphericity implies that the variances of all pairwise differences across systems are equal, even though the marginal variances and covariances may be unequal. Sphericity generalizes *compound symmetry* (Nelson and Matejcik 1995). Several researchers have proposed that compound symmetry accounts for the variance reduction effects of CRN (see Tew and Wilson 1994, Nozari et al. 1987, and Schruben and Margolin 1978 for more details). Procedure *NM* is valid when  $\Sigma$  satisfies sphericity, however Nelson and Matejcik (1995) show it to be extremely robust to departures from sphericity. The procedure is as follows:

1. Specify  $w$  ( $w = \delta$ ),  $\alpha$ , and  $n_0$ . Let  $g = T_{k-1, (k-1)(n_0-1), 0.50}^{(1-\alpha)}$ , where  $T_{k-1, (k-1)(n_0-1), 0.50}^{(1-\alpha)}$  is the  $(1-\alpha)$ -quantile of the maximum of a multivariate  $t$  random variable with  $k-1$  dimensions,  $(k-1)(n_0-1)$  degrees of freedom, and common correlation 0.50.

2. Take i.i.d. samples  $Y_{i1}, Y_{i2}, \dots, Y_{in_0}$  from each of the  $k$  competing systems using CRN across systems.
3. Compute the sample variance of the difference under the condition of sphericity as

$$S^2 = \frac{2 \sum_{i=1}^k \sum_{j=1}^{n_0} (Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot})^2}{(k-1)(n-1)}$$

4. Compute the final required sample size (constant for all  $k$  alternatives) as

$$N = \max\{n_0, \lceil (gS/w)^2 \rceil\}.$$

5. Take  $N - n_0$  additional i.i.d. observations from each system, using CRN across systems.
6. Compute the overall sample means for each system as

$$\bar{Y}_i = \frac{1}{N} \sum_{j=1}^N Y_{ij} \text{ for } i = 1, 2, \dots, k.$$

7. Select the system with the largest  $\bar{Y}_i$  as the best alternative.
8. Simultaneously, form the MCB confidence intervals as

$$\mu_i - \max_{j \neq i} \mu_j \in [-(\bar{Y}_i - \max_{j \neq i} \bar{Y}_j - w)^-, (\bar{Y}_i - \max_{j \neq i} \bar{Y}_j + w)^+]$$

for  $i = 1, 2, \dots, k$   
 where  $-x^- = \min\{0, x\}$  and  $x^+ = \max\{0, x\}$ .

Note that the value of  $T_{k-1, (k-1)(n_0-1), 0.50}^{(1-\alpha)}$  in Step 1 of Procedure *NM* can be derived from Table 4 of Hochberg and Tamhane (1987) or Table B.3 of Bechhofer et al. (1995). For values that fall outside of the tables, the FORTRAN program of Dunnett (1989) can be used.

Nelson and Matejcik (1995) report results that suggest that Procedure *NM* is superior to Procedures *R*, *DD*, and *CY* in terms of the total observations required to obtain the desired confidence level. Procedure *NM*'s only potential drawback is that the assumption of sphericity may not be satisfied (either exactly or approximately) in many situations (Nelson and Matejcik 1995). To evaluate the procedure's robustness to departures from sphericity, Nelson and Matejcik (1995) performed an empirical study. They found that when the desired  $P\{CS\} = 0.95$ , the actual probability attained ranged from 0.88 to 1.0 with a mean of 0.94. Provided the assumption of the data's normality is not significantly violated, this performance suggests that the procedure is sufficiently robust for use in practice. They suggest that the analyst consider slightly inflating the nominal coverage probability  $(1-\alpha)$  to ensure adequate coverage. They also conclude that even when slightly inflating the nominal coverage probability, Procedure *NM* should still outperform Procedure *CY* in terms of the required sample size.

## 5 APPLICATION OF PROCEDURE *NM*

Combined R&S-MCB procedures are more attractive for use in simulation optimization than using either R&S or MCB individually since combined procedures provide both R&S selection and MCB inference with little or no additional computational overhead. Procedure *NM* (Nelson and Matejcik 1995) is the most efficient of the existing combined procedures. For this reason, Swisher (1999) selects it as the simulation optimization technique to apply to the determination of the optimal clinic design for a two-physician family practice clinic from among a group of seventeen competing alternative designs. Clinic design optimality is defined in terms of a multiattribute performance measure referred to as the clinic effectiveness (CE) measure. This is a scalar measure derived from several clinic output performance measures encompassing both clinic profit (revenues - expenses) and patient service levels (e.g., amount of patient waiting). In general, clinic profit is penalized for poor patient service.

Procedure *NM* allows the specification of an indifference zone (which also specifies the MCB whisker length,  $w$ ). A value of  $\delta = w = 10$  was used to select the optimal clinic design from among the seventeen competing alternatives. Since a  $P\{CS\} = (1 - \alpha) = 0.95$  was desired, a more conservative value of  $\alpha = 0.03$  was used, as suggested by Nelson and Matejcik (1995). Given  $n_0 = 30$ ,  $w = \delta = 10$ , and  $\alpha = 0.03$ ,  $g = T_{16, 464, 0.50}^{(0.97)} = 2.7910$  (as derived from Dunnett's 1989 procedure). The sample variance under the condition of sphericity,  $S^2$ , is 882.80 and  $N = \max\{30, \lceil (gS/w)^2 \rceil\} = 69$ . Therefore, 39 ( $N - n_0$ ) observations were generated for each of the seventeen clinic designs by simulating 417 clinic days (27 deleted transient observations + 390 usable steady-state observations) using CRN across designs and forming 39 batches of size  $b = 10$ . Using the formulas provided in Section 4, the overall sample means for the CE measure of each design alternative was computed and the MCB intervals formed (see Table 1).

Procedure *NM* selects clinic design 4 as the best clinic design. From a R&S perspective, this means that with probability greater than or equal to 0.97, clinic design 4 has mean  $\mu_4$  within  $\delta = 10$  of the design with the true largest mean,  $\mu_{[1]}$ . Examination of the MCB intervals provides inferences on the (assumed) superiority of design 4. Interestingly, four other designs (7, 8, 9, and 12) have MCB intervals that contain 0. This means, from an MCB perspective, there is no one clearly superior design. Designs 4, 7, 8, 9 and 12 are all clearly superior to the remaining systems whose upper MCB bound is 0, however there is no clear winner among them. Note that had one design possessed a lower MCB bound of 0, while the rest were upper-bounded by 0, then that design would have been selected as best by MCB.

Table 1: Overall Sample Means and MCB Results for Procedure *NM*

Design	$\bar{Y}_i$	Lower MCB Limit	$\bar{Y}_i - \max_{j \neq i} \bar{Y}_j$	Upper MCB Limit
1	388.17	-28.00	-18.00	0
2	392.41	-23.76	-13.76	0
3	394.18	-21.99	-11.99	0
4	406.17	-9.76	0.24	10.24
5	380.70	-35.47	-25.47	0
6	386.07	-30.10	-20.10	0
7	405.93	-10.24	-0.24	9.76
8	398.74	-17.43	-7.43	2.57
9	405.32	-10.85	-0.85	9.15
10	382.72	-33.45	-23.45	0
11	381.11	-35.06	-25.06	0
12	401.98	-14.19	-4.19	5.81
13	392.53	-23.65	-13.65	0
14	392.07	-24.10	-14.10	0
15	393.78	-22.39	-12.39	0
16	378.67	-37.50	-27.50	0
17	378.28	-37.90	-27.90	0

One of the benefits of using a combined R&S-MCB procedure is that the analyst gains inferences on systems other than the best, which may lead to the selection of an inferior system (if it is not inferior by much) based on some secondary criteria not reflected in the performance measure of interest (Matejcik and Nelson 1993). Although profit is a component of the performance measure used (CE measure), no real inference on clinic profit can be made from examining the CE measure. Therefore, a decision-maker would likely be interested in examining clinic profit as a measure separate from the CE measure for the five clinic designs whose MCB interval covers zero. Table 2 provides the mean daily clinic profit (without any patient service penalties) for each of the five best clinic designs.

Table 2: Mean Daily Clinic Profit for the Five Best Clinic Designs

Clinic Design	Mean Daily Profit (\$)
4	851.92
7	990.78
8	931.89
9	833.88
12	972.08

Note that clinic design 7 produces approximately \$140 per day more clinic profit than the design selected as the best (clinic design 4). In addition, clinic design 7's overall sample mean is less than 25 cents less than design 4's overall sample mean (see Table 1). In short, the MCB inference provided by Procedure *NM* would lead the

clinical decision-maker to choose clinic design 7, despite the fact that clinic design 4 was selected as the best by Procedure *NM*'s R&S result. If only a R&S approach had been used to evaluate the clinic designs, the clinical decision-maker would have selected an excellent design in terms of CE. However, that choice may cost the clinic \$140 per day in profit compared to an equally good (from an MCB perspective) choice. In this case, the value of the application of a combined R&S-MCB procedure is obvious.

## 6 CONCLUSIONS

Ranking and selection and multiple comparison procedures provide excellent tools for selecting the best of  $k$  competing alternatives in discrete-event simulation. R&S approaches allow the simulation analyst to screen alternatives so that he/she may fully study a subset of size  $m$ . They also allow the analyst to choose the best of  $k$  alternatives, given an indifference zone, at or above a user-specified probability. MCPs provide inference about the relationships among  $k$  alternatives. The typical simulation analyst would likely benefit most from applying the unified R&S-MCB procedures proposed by Matejcik and Nelson (1993, 1995) and Nelson and Matejcik (1995). These procedures allow the analyst to select the best system and gain insight about how much better the best is in comparison to the rest of the alternatives. Unified methodologies such as these should dominate the R&S and MCP literature in the future. In addition, further research in the area of multivariate R&S procedures would be beneficial for those situations in which the analyst cannot easily reduce the performance measures to a univariate measure. Unfortunately, it appears there is no current significant research being done in this area.

## REFERENCES

Bechhofer, R.E. 1954. A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Annals of Mathematical Statistics* 25:16-39.

Bechhofer, R.E., J. Kiefer, and M. Sobel. 1968. *Sequential Identification and Ranking Procedures*. The University of Chicago Press, Chicago, Illinois.

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

Bofinger, E. and G.J. Lewis. 1992. Two-stage procedures for multiple comparisons with a control. *American Journal of Mathematical and Management Sciences* 12(4): 253-275.

- Bratley, P., B.L. Fox, and L.E. Schrage. 1987. *A Guide to Simulation, Second Edition*. Springer-Verlag, New York.
- Chen, C.H. 1995. An effective approach to smartly allocate computing budget for discrete-event simulation. *Proceedings of the 34th IEEE Conference on Decision and Control*, 2598-2605.
- Chen, C.H., H.C. Chen, and L. Dai. 1996. A gradient approach for smartly allocating computing budget for discrete event simulation. *Proceedings of the 1996 Winter Simulation Conference*, 398-405.
- Chen, C.H., Y. Yuan, H.C. Chen, E. Yücesan, and L. Dai. 1998. Computing budget allocation for simulation experiments with different system structures. *Proceedings of the 1998 Winter Simulation Conference*, 735-741.
- Chen, H.C., C.H. Chen, L. Dai, and E. Yücesan. 1997. New development of optimal computing budget allocation for discrete event simulation. *Proceedings of the 1997 Winter Simulation Conference*, 334-341.
- Chen, P. 1988. On selecting the best of k systems: An expository survey of subset-selection multinomial procedures. *Proceedings of the 1988 Winter Simulation Conference*, 440-444.
- Chick, S.E. 1997. Selecting the best system: A decision-theoretic approach. *Proceedings of the 1997 Winter Simulation Conference*, 326-333.
- Chick, S.E. and K. Inoue. 1998. Sequential allocations that reduce risk for multiple comparisons. *Proceedings of the 1987 Winter Simulation Conference*, 669-676.
- Clark, G.M. and W.N. Yang. 1986. A Bonferroni selection procedure when using common random numbers with unknown variances. *Proceedings of the 1986 Winter Simulation Conference*, 313-315.
- Damerджи, H. and M.K. Nakayama. 1996. Two-stage procedures for multiple comparisons with a control in steady-state simulations. *Proceedings of the 1996 Winter Simulation Conference*, 372-375.
- Damerджи, H., P.W. Glynn, M.K. Nakayama, and J.R. Wilson. 1996. Selecting the best system in transient simulations with known variances. *Proceedings of the 1996 Winter Simulation Conference*, 281-286.
- Dudewicz, E.J. 1976. Statistics in simulation: How to design for selecting the best alternative. *Proceedings of the 1976 Winter Simulation Conference*, 67-71.
- Dudewicz, E.J. and S.R. Dalal. 1975. Allocation of observations in ranking and selection with unequal variances. *The Indian Journal of Statistics* 37B(1):28-78.
- Dudewicz, E.J. and V.S. Taneja. 1978. Multivariate ranking and selection without reduction to a univariate problem. *Proceedings of the 1978 Winter Simulation Conference*, 207-210.
- Dunnett, C.W. 1955. A multiple comparisons procedure for comparing several treatments with a control. *Journal of the American Statistical Association* 78:965-971.
- Dunnett, C.W. 1989. Multivariate normal probability integrals with product correlation structure. *Applied Statistics* 38:564-579. Correction 42:709.
- Fu, M. 1994a. Optimization via simulation: A review. *Annals of Operations Research* 53:199-247.
- Fu, M. 1994b. A tutorial review of techniques for simulation optimization. *Proceedings of the 1994 Winter Simulation Conference*, 149-156.
- Goldsmán, D. 1983. Ranking and selection in simulation. *Proceedings of the 1983 Winter Simulation Conference*, 387-393.
- Goldsmán, D. 1984a. A multinomial ranking and selection procedure: Simulation and applications. *Proceedings of the 1984 Winter Simulation Conference*, 259-263.
- Goldsmán, D. 1984b. On selecting the best of k systems: An expository survey of indifference-zone multinomial procedures. *Proceedings of the 1984 Winter Simulation Conference*, 107-112.
- Goldsmán, D. 1985. Ranking and selection procedures using standardized time series. *Proceedings of the 1985 Winter Simulation Conference*, 120-123.
- Goldsmán, D. 1986. Tutorial on indifference-zone normal means ranking and selection procedures. *Proceedings of the 1986 Winter Simulation Conference*, 370-375.
- Goldsmán, D. 1987. Ranking and selection tutorial: 2-factor normal means procedures. *Proceedings of the 1987 Winter Simulation Conference*, 52-57.
- Goldsmán, D. and B.L. Nelson. 1990. Batch-size effects on simulation optimization using multiple comparisons. *Proceedings of the 1990 Winter Simulation Conference*, 288-293.
- Goldsmán, D. and B.L. Nelson. 1994. Ranking, selection and multiple comparisons in computer simulation. *Proceedings of the 1994 Winter Simulation Conference*, 192-199.
- Goldsmán, D. and B.L. Nelson. 1998. Statistical screening, selection, and multiple comparison procedures in computer simulation. *Proceedings of the 1998 Winter Simulation Conference*, 159-166.
- Goldsmán, D., B.L. Nelson, and B. Schmesier. 1991. Methods for selecting the best system. *Proceedings of the 1991 Winter Simulation Conference*, 177-186.
- Gray, D. and D. Goldsmán. 1988. Indifference-zone selection procedures for choosing the best airspace configuration. *Proceedings of the 1988 Winter Simulation Conference*, 445-450.
- Gupta, S.S. 1956. On a decision rule for a problem in ranking means. *Mimeograph Series No. 150, Institute of Statistics*. University of North Carolina, Chapel Hill.



- Gupta, S.S. and J.C. Hsu. 1977. Subset selection procedures with special reference to the analysis of two-way layout: Application to motor-vehicle fatality data. *Proceedings of the 1977 Winter Simulation Conference*, 81-85.
- Gupta, S.S. and J.C. Hsu. 1984. A computer package for ranking, selection, and multiple comparisons with the best. *Proceedings of the 1984 Winter Simulation Conference*, 251-257.
- Gupta, S.S. and T.J. Santner. 1973. On selection and ranking procedures – a restricted subset selection rule. *Proceedings of the 39<sup>th</sup> Session of the International Statistical Institute, Volume 1*.
- Ho, Y.C., R. Sreenivas, and P. Vakili. 1992. Ordinal optimization of discrete event dynamic systems. *Discrete Event Dynamical Systems*, 2(2), 61-88.
- Hochberg, Y. and A.C. Tamhane. 1987. *Multiple Comparison Procedures*. John Wiley and Sons, New York.
- Hsu, J.C. 1984. Constrained simultaneous confidence intervals for multiple comparisons with the best. *Annals of Statistics* 12:1136-1144.
- Hsu, J.C. 1996. *Multiple Comparisons: Theory and Methods*. Chapman & Hall, London.
- Hsu, J.C. and B.L. Nelson. 1988. Optimization over a finite number of system designs with one-stage sampling and multiple comparisons with the best. *Proceedings of the 1988 Winter Simulation Conference*, 451-457.
- Inoue, K. and S.E. Chick. 1998. Comparison of bayesian and frequentist assessments of uncertainty for selecting the best system. *Proceedings of the 1998 Winter Simulation Conference*, 727-734.
- Jacobson, S.H. and L.W. Schruben. 1989. A review of techniques for simulation optimization. *Operations Research Letters* 8:1-9.
- Jacobson, S.H., L.W. Schruben, J.R. Swisher, and P. Hyden. 1999. Discrete-event simulation optimization: A survey of recent advances. Submitted for publication.
- Koenig, L.W. and A.M. Law. 1985. A procedure for selecting a subset of size m containing the l best of k independent normal populations, with applications to simulation. *Communications in Statistics* B14(3):719-734.
- Kolojna, B., D.R. Kalasky, and J.M. Mutmansky. 1993. Optimization of dispatching criteria for open-pit truck haulage system design using multiple comparisons with the best and common random numbers. *Proceedings of the 1993 Winter Simulation Conference*, 393-401.
- Law, A.M. and W.D. Kelton. 1991. *Simulation Modeling and Analysis, Second Edition*. McGraw-Hill, Inc., New York.
- Matejcik, F.J. and B.L. Nelson. 1993. Simultaneous ranking, selection and multiple comparisons for simulation. *Proceedings of the 1993 Winter Simulation Conference*, 386-392.
- Matejcik, F.J. and B.L. Nelson. 1995. Two-stage multiple comparisons with the best for computer simulation. *Operations Research* 43(4):633-640.
- Morrice, D.J., J. Butler, and P.W. Mullarkey. 1998. An approach to ranking and selection for multiple performance measures. *Proceedings of the 1998 Winter Simulation Conference*, 719-725.
- Nakayama, M.K. 1995. Selecting the best system in steady-state simulations using batch means. *Proceedings of the 1995 Winter Simulation Conference*, 362-366.
- Nakayama, M.K. 1996. Multiple comparisons with the best in steady-state simulations, *Proceedings of the Second International Workshop on Mathematical Methods in Stochastic Simulation and Experimental Design*, 230-235.
- Nakayama, M.K. 1997a. Multiple-comparison procedures for steady-state simulations, *Annals of Statistics* 25:2433-2450.
- Nakayama, M.K. 1997b. Using common random numbers in two-stage procedures for multiple comparisons with the best for steady-state simulations. *Proceedings of the 11th European Simulation Multiconference*, 155-159.
- Nakayama, M.K. 1999. Multiple comparisons with the best using common random numbers in steady-state simulations. *Journal of Statistical Planning and Inference*, to appear.
- Nelson, B.L. 1993. Robust multiple comparison procedures under common random numbers. *ACM Transactions on Modeling and Computer Simulation* 3(3):225-243.
- Nelson, B.L. and J.C. Hsu. 1993. Control-variate models of common random numbers for multiple comparisons with the best. *Management Science* 39(8): 989-1001.
- Nelson, B.L. and F.J. Matejcik. 1995. Using common random numbers for indifference-zone selection and multiple comparisons in simulation. *Management Science* 41(12):1935-1945.
- Nozari, A., S. Arnold, and C. Pegden. 1987. Statistical analysis for use with Schruben and Margolin correlation induction strategy. *Operations Research* 35:127-139.
- Paulson, E. 1964. A sequential procedure for selecting the population with the largest mean from k normal populations. *Annals of Mathematical Statistics* 35:174-180.
- Rinott, Y. 1978. On two-stage selection procedures and related probability-inequalities. *Communications in Statistics* A7(8):799-811.

- Sanchez, S.M. 1997. It is a far, far better mean I find ... *Proceedings of the 1997 Winter Simulation Conference*, 31-38.
- Santner, T.J. 1975. A restricted subset selection approach to ranking and selection problems. *Annals of Statistics* 3:334-349.
- Schruben, L.W. and B. Margolin. 1978. Pseudorandom number assignment in statistically designed simulation and distribution sampling experiments. *Journal of the American Statistical Association* 73:504-525.
- Sullivan, D.W. and J.R. Wilson. 1989. Restricted subset selection procedures for simulation. *Operations Research* 37:52-71.
- Swisher, J.R. 1999. Evaluation of the design of a family practice healthcare clinic using discrete-event simulation. Master's Thesis. Virginia Tech, Blacksburg, Virginia.
- Tew, J.D. and J.R. Wilson. 1994. Estimating simulation metamodels using combined correlation-based variance reduction techniques. *IIE Transactions* 26:2-16.
- Tukey, J.W. 1953. The problem of multiple comparisons. Unpublished manuscript.
- Wen, M.J. and H.J. Chen. 1994. Single-stage multiple comparison procedures under heteroscedasticity. *American Journal of Mathematical and Management Sciences* 14(1,2):1-48.
- Yang, W.N. and B.L. Nelson. 1989. Optimization using common random numbers, control variates, and multiple comparisons with the best. *Proceedings of the 1989 Winter Simulation Conference*, 1112-1120.
- Yang, W.N. and B.L. Nelson. 1991. Using common random numbers and control variates in multiple-comparison procedures. *Operations Research* 39(4):583-591.
- Yuan, M. and B.L. Nelson. 1993. Multiple comparisons with the best for steady-state simulation. *ACM Transactions on Modeling and Computer Simulation* 3(1):66-79.
- Zinger, A. 1961. Detection of the best and outlying normal populations with known variances. *Biometrika* 48:457-461.
- Zinger, A. and J. St. Pierre. 1958. On the choice of the best amongst three normal populations with known variances. *Biometrika* 45:436-446.

## AUTHOR BIOGRAPHIES

**JAMES R. SWISHER** is Director of Analysis and Design for Lotus Biochemical Corporation. He has also served in the Management Engineering Unit and as Director of Call Center Operations for Lotus' subsidiary, Biological & Popular Culture, Inc. (Biopop). Prior to joining Biopop, he was employed as an Industrial Engineer at the Joint Warfare Analysis Center where he was involved in the

beta-testing of the Visual Simulation Environment (VSE). He holds a B.S. and a M.S. in Industrial & Systems Engineering (Operations Research option) from Virginia Tech. Mr. Swisher is a member of INFORMS, IIE, and Alpha Pi Mu.

**SHELDON H. JACOBSON** is an Associate Professor in the Department of Mechanical and Industrial Engineering at the University of Illinois at Urbana-Champaign. Before joining the University of Illinois, he served on the faculties of Virginia Tech and Case Western Reserve University. He has a B.Sc. and M.Sc. in Mathematics from McGill University and a Ph.D. in Operations Research from Cornell University. His research interests include simulation optimization and sensitivity analysis, frequency domain approaches to analyzing simulation outputs, issues related to the complexity of analyzing structural properties of discrete-event simulation models, and stochastic algorithms for discrete optimization problems. His research has been applied in the manufacturing, service, and health care industries.