Proceedings of the 1997 Winter Simulation Conference
ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson

IT IS A FAR, FAR BETTER MEAN I FIND...

Susan M. Sanchez

School of Business Administration
University of Missouri–St. Louis
8001 Natural Bridge Road
St. Louis, Missouri, 63121–4499, U.S.A.

ABSTRACT

An important problem often faced by simulation analysts is that of choosing (or selecting) one of several potential system designs. This can be accomplished by comparing output from simulation models of these systems. In order to keep from making poor decisions, appropriate statistical models should be used. We present an overview of the types of questions that can be effectively answered using selection and ranking or multiple comparison procedures, and provide references for specific details regarding their use.

1 INTRODUCTION

A simulation analyst is often called on to compare two or more systems. Sometimes the differences between the systems can be expressed in terms of different levels of certain quantitative factors, such as the time between scheduled maintenance for a specific piece of manufacturing equipment. In such cases experimental designs (such as factorial or fractional factorial experiments) can be used to build metamodels of the relationship between the factors and the responses, leading the analyst to a better understanding of the system and indicating good designs. In many instances, however, the systems can be viewed as a set of discrete, *qualitative* alternatives. For example, the difference between a LIFO queue and a FIFO queue is qualitative. At other times, even if quantitative factors (such as service times, arrival rates, times between maintenance) give rise to the different potential systems, the analyst may be restricted to a small set of potential designs without the flexibility (or need) to model explanatory relationships. This simpler setting is amenable to analysis using selection and ranking or multiple comparison techniques.

In this paper we begin with a brief summary of both selection and ranking procedures and multiple comparison techniques. We describe four basic scenarios where these methods would be appropriate, and conclude with a discussion of some of the benefits the analyst may receive, as well as pitfalls to watch out for in setting up selection experiments. Our goal is to provide the simulation reader with an appreciation of the utility of these approaches, rather than detailed descriptions of specific procedures.

1.1 Selection and Ranking

Selection and ranking refers to a field of statistics that arose from directly addressing questions of interest to the analyst. A selection procedure has the following elements:

- a specific statement of the selection criteria (e.g., large mean, small variance, etc.)
- rules for determining sample sizes
- an intuitive rule for the selection decision
- a probability guarantee associated with the decision

Underlying these elements are assumptions regarding the distribution of the performance measures (parametric or nonparametric), as well as sampling assumptions (e.g., independent, identically distributed vs. common random numbers).

There is a rich literature on selection and ranking in the field of statistics. While the application of these techniques to simulation is more recent, simulation is particularly amenable to this type of analysis because it gives the analyst a great deal of control over the experimental setting. Also, many selection procedures require two-stage or multi-stage procedures. These can be problematic in real-world settings where data collection is slow, as in awaiting the results of cancer treatments in clinical trials. However, since discrete-event simulation typically allows one to collect a great deal of data quickly, multiple stages do not pose the same problem in the simulation context.

1.2 Multiple Comparison Methods

At times it is important not only to select a system (or systems) from among the k of interest, but also to compare the systems' performances quantitatively. One method of accomplishing this is by estimating the differences between each system's performance and the best performance of the rest, e.g., bounding the differences $\mu_i - \max_{j \neq i} \mu_j$ for $i = 1, \ldots, k$. The resulting sets of confidence intervals are known as *multiple comparisons with the best* (MCB). MCB procedures were first developed by Hsu (1984). More recently, selection procedures which simultaneously allow estimation have been examined by Matejcik and Nelson (1993, 1995), Nelson and Matejcik (1995), Nelson and Yuan (1993), and Chen and Zhang (1993).

These procedures give the analyst the best of both worlds: an intuitive method of determining the best system (within some practical difference), along with detailed information about how much better it is. Hsu and Peruggia (1994) have proposed a graphical technique called the *mean-mean scatter plot* for easy display of the results of multiple comparison procedures. For more thorough treatments of multiple comparison methods, see Hochberg and Tamhane (1987) or Hsu (1996).

2 SAMPLE SCENARIOS

We describe four basic scenarios for selection and ranking problems: choosing adequate systems, choosing good systems, choosing better systems, and choosing the best system. Several of the procedures we discuss are detailed in Goldsman and Nelson (1994). Discussions written for simulation analysts can be found in chapter 13 of Banks, Carson and Nelson (1996), chapter 10 of Law and Kelton (1991), or chapter 7 of Thesen and Travis (1992). For broader coverage of selection and ranking procedures, see Bechhofer, Goldsman and Santner (1995), Gibbons, Olkin and Sobel (1977), Gupta and Panchepakesan (1979), or Mukhopadhyay and Solanky (1994).

It is important to note that although the terms *adequate, good, better* and *best* have an implicit ordering, your goal should not be to force every problem into one of selecting the best system! Philosophically, selection procedures have been developed to provide appropriate methods for addressing real-world problems. As the examples show, each of the four scenarios is of interest in itself.

2.1 Choosing Adequate Systems

Sometimes you may have some threshold value which can be used to classify a system as acceptable (meeting or exceeding this threshold) or unacceptable (failing to meet the minimum standard). Consider the following example:

Example 1a: *Several different investment strategies will be simulated to evaluate their expected rate of return. You intend to implement the strategy with the highest expected rate of return, but only if its expected return is larger than a zero-coupon bond that offers a known, fixed return.*

The goal is to select the best investment strategy (system) only if it is better than a standard—if no better strategy is found, you will continue with the standard. The output of the standard in this example is a known (fixed) quantity, not a random variable. This problem can be modeled as an *indifference zone* selection problem. If the true best system is chosen by the analyst, we say a *correct selection* has been made. Indifference zone methods guarantee that the probability of correct selection is sufficiently high as long as the difference between the performance of the best and second best system is at least some user-specified amount—the smallest practically-significant difference. (If the performances of the best and second-best systems differ by less than this amount, they are considered equivalent for practical purposes.)

Bechhofer and Turnbull (1978) devised a two-stage procedure for selecting the best system relative to a given standard (e.g., Example 1a) when the responses are normal with a common unknown variance. Their procedure requires that a sample two or more be taken from each system in the initial stage for variance estimation purposes. Law and Kelton (1991) recommend that a sample size of at least twenty be utilized to obtain a sufficiently accurate estimate during the first stage without wasting observations unnecessarily.

The slight variation in the problem statement of Example 1b (Goldsman and Nelson, 1994) indicates that the indifference zone approach is not the only way to formulate the problem.

Example 1b: *Several different investment strategies will be simulated to evaluate their expected rate of return. The strategy ultimately chosen may not be the one with the largest rate of return—since factors such as risk could be considered—but none of the strategies will be chosen unless its expected return is larger than a zero-coupon bond that offers a known, fixed return. You are interested in identifying all strategies which outperform the zero-coupon bond.*

Here you are asked to identify a subset of the potential systems (strategies) which you believe contains *all* those better than the standard. (This is one type of *subset selection* procedure.) A correct selection is said to have been made only if the selected subset indeed contains all adequate strategies, and the procedures are constructed so that the probability of a correct selection is high. Gupta and Sobel (1958) considered several variations of this problem, including subset selection based on means for normal populations (either known or unknown common variance) as well as gamma scale parameters and binomial success probabilities.

2.2 Choosing Good Systems

Sometimes you may be faced with a large number of alternatives. Rather than performing detailed analyses of all alternatives, you wish to first pare down this list to a more manageable size by identifying those designs worthy of further scrutiny. This is often referred to as a *screening problem*: you are interested in identifying which designs have the better performance with respect to a certain criteria, which have similar performance, and which are clearly inferior and can be eliminated from further consideration.

Screening has its place in simulation settings in two basic scenarios. First, it may be that several criteria will influence the final decision, and trade-offs among these criteria may be difficult to quantify. In this situation the screening procedure represents a 'first cut' at the decision process: 'good' alternatives will be identified (for future study) and unfavorable alternatives will be stricken from further consideration. However, since the final decision may be made based on other criteria, you are not interested in unnecessarily restricting the result to the identification of a single system. Consider the following example:

Example 2a: *You are designing a data warehouse for the municipal government of a large city where the database will distributed across several geographic locations. A team of MIS analysts has identified 15 potential alternatives for allocating portions of the database to various locations. Fast response times are desirable, but issues of cost, ease of implementation, and political considerations also play a role. Your goal is to identify designs with low response times from which the ultimate decision will be made.*

Second, it may be that the number of potential systems is so large that doing careful analyses of each is prohibitively costly or time-consuming. A pilot study will indicate which alternatives are worth further investigation, which may include more detailed simulation modeling in addition to a more comprehensive analysis of the performance output.

Example 2b: *The result of a brain-storming session has left you with 20 possible physical layouts for a job shop. You are interested in finding a design which yields a high system throughput as the primary criteria. However, the final choice of the design will depend on secondary factors such as cost, ease of implementation, etc. Your charge is to simulate the alternatives and identify a small set of candidate designs for further consideration.*

Either of these scenarios lends itself to the screening approach called *subset selection*, which was pioneered by Gupta (1965). The goal is to select a subset of the systems which contains the true best system (e.g., that with the lowest response time in Example 2a). A correct selection is said to have been made only if the selected subset indeed contains the true 'best' system. Subset selection procedures guarantee that the probability of success is sufficiently high for *any* configuration of the means. Note that the size of the selected subset is a random variable, and depends on the actual values of the underlying means and the common variance. The size of the selected subset will tend to be small if a few alternatives are much better than the rest, and tend to be large if all alternatives have roughly equal performances.

The basic procedure of Gupta (1965) uses balanced data collection—a common sample size from each system. Gupta and Huang (1976) propose a procedure which works in the unbalanced case. A single-stage procedure for binomial systems was proposed by Gupta and Sobel (1960). Sanchez (1987a, b) proposed a sequential procedure which yielded the same subset more efficiently. Applications for observational data, where sample sizes are unknown and uncontrollable, were considered by Sanchez and Higle (1992) and Kannan and Sanchez (1994).

2.3 Choosing Better Systems

One reason simulation is such a valuable tool for decision-making is its use in studying the performance of prospective systems. At times a system is already in place, so you may not be interested in selecting the best system (or a subset of the best) unless it is better than the status quo (also called a default or control). This is similar to the screening problem of Section 2.2. However, instead of systems being defined as 'good' relative to other alternatives, there is a default system which must be beaten. In this case the alternatives under consideration must yield sufficient proof that they are better than the status quo in order for us to recommend a switch.

Example 3a: *A manufacturer of high-density pressure laminates is examining the operation of one of their presses. Carriers for the laminates cycle through the press operation, but periods of blocking* *and starving occur because of material handling issues and imbalances on the line. Management is considering installing additional conveyors and/or altering several of the work-in-process buffers, but these changes would require shutting down production. Your simulation models will determine the recommended action.*

An indifference zone view of Example 3a means you would NOT be interested in selecting a new system if it is considered equivalent in performance to the default. The indifference zone procedure of Paulson (1952) (described in Bechhofer *et al.*, 1995, or Goldsman and Nelson, 1994) allows one to select the best but assumes that the common variance among the populations is known.

As Example 3b illustrates, at times one may be interested in identifying all designs better than the control, rather than a single best design.

Example 3b: *A mail-order firm has a simulation model of their current order-filling and distribution process. They are considering a move to a different distribution system. You have been asked to model ten alternative designs which are under consideration. These result from looking at alternative distributors (common carriers, less-than-truckload distribution) and dispatching rules. Management is interested in improving the average customer wait time. They will keep their current system unless one or more of the alternatives result in better customer service: if so, the choice of the system to implement may depend on secondary factors such as cost and ease of implementation.*

Gupta and Sobel (1958) describe a procedure for normal systems (with a common, known or unknown variance), as well as a procedure for binomial systems.

Examples 3a and 3b may also be analyzed using multiple comparison methods. When comparing to a control (or default) system (say, k) we construct a total of $k-1$ simultaneous confidence intervals for $\mu_i - \mu_k$ for all $i \neq k$ rather than (k choose 2) simultaneous confidence intervals for all pairwise comparisons. Furthermore, if we know that only differences in a specified direction are of interest then one-sided confidence intervals can be formed. (In Example 3b we are interested only in systems with smaller customer wait time.) Procedures for multiple comparisons with a control (MCC) when variances are equal can be found in Miller (1981). A procedure valid when variances are unequal appears in Tamhane (1977). These cases and others are also discussed in Hochberg and Tamhane (1987), Hsu (1996), and Miller (1981).

2.4 Choosing the Best System

Problems with a single well-defined objective may fall naturally into the category of 'choosing the best.' Consider the following scenario:

Example 4: *For the purpose of evaluation prior to purchase, a multinational company has developed simulation models of six different computer network infrastructures. The single measure of system performance is the time to failure (TTF), so that larger TTF is better, and a highly reliable system is desired. The six systems arise from variations in parameters affecting the TTF and time-to-repair distributions and the internet service providers. From a practical standpoint, differences of less than about 12 hours are considered equivalent. The goal is to select the system with the largest TTF and/or estimate the differences in the TTF's associated with the systems.*

Bechhofer (1954) considered the case of known, common variances in the first published work describing the indifference-zone approach. A procedure for unknown common variances is given in Bechhofer, Dunnett and Sobel (1954). Dudewisz and Dalal (1975) and Rinott (1978) considered the case of unknown, potentially unequal variances. Damerdji *et al.* (1996) have begun looking at this problem in the transient system context by considering the case of selecting the best system when the performance variances are known (possibly unequal) but the performance means and distributions are unknown.

Example 4 can be analyzed using a combined selection and ranking/MCB approach. In the simulation context, Matejcik and Nelson (1993, 1995) discuss the case for independent sampling. They extend these results in Nelson and Matejcik (1995), where common random number streams can be used to improve the efficiency and reduce total sample size requirements. (Both these procedures are provided in detail in Goldsman and Nelson, 1994.) Yang and Nelson (1991) and Nelson and Hsu (1993) considered control variate models for multiple comparisons with the best. Goldsman, Nelson and Schmeiser (1991) compare selection, multiple comparison, and an interactive approach to identifying the best system.

A nonparametric approach to the network TTF problem is also possible. Rather than attempting to select the system with the largest mean TTF, we might wish to select the system that is *the most likely* to yield the largest actual TTF. In this case, the problem is formulated mathematically as finding the multinomial category associated with the largest probability of occurrence. (Since we are dealing with probabilities, the indifference zone is expressed in ratio rather than difference terms.) A single-stage indifference-zone procedure was proposed by Bechhofer, Elmaghraby and Morse (1959). More efficient procedures were proposed by Bechhofer and Goldsman (1986) and by Miller, Nelson and Reilly (1996). Bechhofer and Goldsman use closed, sequential sampling to allow experimentation to end when one population becomes sufficiently ahead of the other, while Miller *et al.* make use of pseudo-replications to improve efficiency.

3 BENEFITS

Selection and ranking is useful because it directly addresses questions of interest to the analyst. This is best illustrated in contrast to the most widely-used analysis method for analyzing multiple systems: the Analysis of Variance (ANOVA). Consider the scenario in Section 2.4 (Selecting the Best). If n observations (e.g., n batch means) are collected for each of the k systems, then an ANOVA could be conducted to test the following hypothesis:

$$\begin{array}{ll} H_0: & \text{mean performances are all equal} \\ & \quad \text{for the } k \text{ systems} \\ H_A: & \text{mean responses are NOT all equal} \\ & \quad \text{for the } k \text{ systems} \end{array}$$

How interesting are the results? If you reject the null hypothesis, you are still faced with the problem of determining which system to recommend. The fact that H_0 is rejected is not in itself surprising—the systems the simulations are modeling differ in real terms, and in most situations the time and effort required to build the simulation models would not have been allocated if it was felt the systems were interchangeable.

Sample size also complicates matters. If H_0 cannot be rejected, it may be that the sample size is too small to detect differences of reasonable interest. On the other hand, for sufficiently large sample sizes we are likely to reject H_0 even for tiny differences in the mean responses, even if they are not of practical importance. The problem with this hypothesis testing approach is that it does not directly address the types of questions raised in Section 2: which system (or systems) should be chosen? This can make it difficult to explain the results to management or other decision-makers.

In contrast, both the indifference zone and subset selection approaches provide statistical underpinnings for decision rules which are quite intuitive. For example, if your goal is to select the single system with the largest mean, a natural decision rule to accomplish this is: 'Choose the system which yields the largest sample mean as best.' In the subset selection context, the corresponding decision rule would be 'Choose the systems which yield sample means sufficiently close to the largest sample mean,' where 'sufficiently close' is determined by the sample size and desired probability guarantee.

The selection procedures are also typically easy to implement, in the sense that determining the sample sizes may require only a few table look-ups. Some of these are standard statistical tables. The sources mentioned at the beginning of Section 2 also have tables needed for the selection and ranking procedures they describe.

4 POTENTIAL PITFALLS

Poor Timing. The time to think about selection and ranking is *before* collecting the data, not after. This will insure you that your sample sizes are sufficient to allow you to make a reasonable probability guarantee without wasting resources unnecessarily by simulating longer than necessary.

Loss of information. Not all problems relating to comparing multiple systems should be analyzed using one of the above approaches. As we mention in Section 1, experimental designs can be used in efficient ways if the relationships between system performance can be expressed as a function of several quantitative factors. Earlier in this volume, Kelton (1997) provides an overview and list of references relating to experimental design in the simulation context.

Ignoring system variance. Another pitfall to be aware of is the reliance on comparisons of means for most of these procedures. Complex simulation models often exhibit highly non-linear behavior, and this often translates into quite different variances for different system configurations. In such cases, selecting systems with respect to mean performance alone is questionable.

One alternative is to use a nonparametric procedure to select the best, such as the multinomial procedure for selecting the system with the largest probability of yielding the largest outcome (Bechhofer and Goldsman, 1986). Alternatively, procedures of Section 2.1 (selecting adequate systems) or Section 2.2 (selecting good systems) could be used to identify a subset of systems with adequate/good means with the final choice based on low variance, or vice versa. Recent literature in the industrial quality field pioneered by Taguchi (1986) suggests that combining the mean and variance into a loss function is often more appropriate. However, this so-called *robust design* approach has concentrated on choosing appropriate levels for quantitative factors, rather than selecting from among qualitatively different systems. (See Ramberg *et al.* 1991, Sanchez, 1994, or Sanchez *et al.*, 1997 for robust design methods in the simulation context.)

Using the wrong method. We have cited only a small portion of the literature in the selection and ranking and multiple comparisons area. Procedures have been developed for numerous other situations. Some of these make use of other goals, such as completely ranking the systems, selecting the best k of n systems, or selecting all best systems in the case of ties. Other procedures make use of different underlying assumptions, such as gamma or exponential distributions. Bayesian procedures have also been proposed. This means that even if your problem does not exactly fit one of those described above, an appropriate procedure may already have been developed. However, the large number of procedures available also means there are many opportunities to use the wrong one, so care must be taken in specifying the problem.

5 SUMMARY

We have described a framework for selection and ranking in the simulation setting. These procedures are easy to describe, intuitive in nature, and their implementation often requires only a minimal amount of analysis in conjunction with some table look-ups. Simulation is a setting amenable to the use of these techniques provided that the performance evaluation criteria are clearly specified *a priori* so that appropriate procedures can be used.

REFERENCES

Banks, J., J. S. Carson and B. L. Nelson. 1996. *Discrete-event system simulation*, 2nd ed. Englewood Cliffs, New Jersey: Prentice-Hall.

Bechhofer, R. E. 1954. A single-sample multiple decision procedure for ranking means of normal populations. *Annals of Mathematical Statistics* 25:16–29.

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

Bechhofer, R. E., S. Elmaghraby and N. Morse. 1959. A single-stage multiple decision procedure for selecting the multinomial event which has the highest probability. *Annals of Mathematical Statistics* 30:102–119.

Bechhofer, R. E., and D. M. Goldsman. 1986. Truncation of the Bechhofer-Kiefer-Sobel procedure for selecting the multinomial event which has the largest probability (II): extended tables and an improved procedure. *Communications in Statistics–Simulation and Computation* B15:829–851.

Bechhofer, R. E., D. M. Goldsman and T. J. Santner. 1995. *Design and analysis for statistical selection, screening and multiple comparisons.* New York: John Wiley & Sons, Inc.

Bechhofer, R. E., and B. W. Turnbull. 1978. Two $(k+1)$-decision selection procedures for comparing k normal means with a control. *Journal of the American Statistical Association* 73:385–392.

Chen, H. J., and H. M. Zhang. 1993. Procedures for simultaneous estimation and selection of the best population with a preliminary test. *American Journal of Mathematics and Management Science* 13:195–216.

Damerdji, H., P. W. Glynn, M. K. Nakayama and J. R. Wilson. 1996. Selecting the best system in transient simulations with variances known. *Proceedings of the 1996 Winter Simulation Conference*, ed. J. M. Charnes, D. M. Morrice, D. T. Brunner and J. J. Swain, 281–286. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.

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

Gibbons, J. D., I. Olkin and M. Sobel. 1977. *Selecting and ordering populations: A new statistical methodology.* New York: John Wiley & Sons, Inc.

Goldsman, D., and B. L. Nelson. 1994. Ranking, selection and multiple comparisons in computer simulation. In *Proceedings of the 1994 Winter Simulation Conference*, ed. J. D. Tew, M. S. Manivannan, D. A. Sadowski, and A. F. Seila, 192–199. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.

Goldsman, D., B. L. Nelson and B. Schmeiser. 1991. Methods for selecting the best system. In *Proceedings of the 1991 Winter Simulation Conference*, ed. B. L. Nelson, W. D. Kelton, and G. M. Clark, 177–186. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.

Gupta, S. S. 1965. On some multiple decision (selection and ranking) rules. *Technometrics* 7:225–245.

Gupta, S. S., and D.-Y. Huang. 1976. Subset selection procedures for the means and variances of normal populations: unequal sample sizes. *Sankyā* B38:112–128.

Gupta, S. S., and S. Panchepakesan. 1979. *Multiple decision procedures.* New York: John Wiley & Sons, Inc.

Gupta, S. S., and M. Sobel. 1958. On selecting a subset which contains all populations better than a standard. *Annals of Mathematical Statistics* 28:957–967.

Gupta, S. S., and M. Sobel. 1960. Selecting a subset containing the best of several binomial populations. Chapter 20 in *Contributions to Probability and Statistics*, ed. Olkin, Ghurye, Hoeffding, Madow and Mann. Stanford, California: Stanford University Press.

Hochberg, Y., and A. C. Tamhane. 1987. *Multiple comparison procedures*. New York: John Wiley & Sons, Inc.

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*. London, England: Chapman and Hall.

Hsu, J. C., and M. Peruggia. 1994. Graphical representation of Tukey's multiple comparison method. *Journal of Computational and Graphical Statistics* 3:143–161.

Kannan, P. K., and S. M. Sanchez. 1994. Competitive market structures: A subset selection analysis. *Management Science* 40(11):1484–1499.

Kelton, W. D. 1997. Statistical analysis of simulation output. *Proceedings of the 1997 Winter Simulation Conference*, ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.

Law, A. M., and W. D. Kelton. 1991. *Simulation modeling and analysis*, 2nd ed. New York: McGraw–Hill.

Matejcik, F. J., and B. L. Nelson. 1993. Simultaneous ranking, selection and multiple comparisons for simulation. *Proceedings of the 1993 Winter Simulation Conference*, ed. G. W. Evans, M. Mollaghasemi, E. C. Russell and W. E. Biles, 386–692. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.

Matejcik, F. J., and B. L. Nelson. 1995. Two-stage multiple comparisons with the best for computer simulation. *Operations Research* 43(4):633–640.

Miller, R. G. 1981. *Simultaneous statistical inference*, 2nd ed. New York: Springer-Verlag.

Miller, J. O., B. L. Nelson and C. H. Reilly. 1996. Getting more from the data in a multinomial selection problem. *Proceedings of the 1996 Winter Simulation Conference*, ed. J. M. Charnes, D. M. Morrice, D. T. Brunner and J. J. Swain, 287–294. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.

Mukhodpadhyay, N., and T. S. Solanky. 1994. *Multistage selection and ranking procedures*. New York: Marcel Dekker.

Nakayama, M. 1995. Selecting the best system in steady-state simulations using batch means. In *Proceedings of the 1995 Winter Simulation Conference*, ed. C. Alexopoulus, K. Kang, W. R. Lilegdon and D. Goldsman, 362–366. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.

Nelson, B. L. 1993. Robust multiple comparisons under common random numbers. Special issue on variance reduction techniques of *ACM Transactions on Modeling and Computer Simulation* 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. *Management Science* 41(12):1935–1945.

Nelson, B. L., and M. Yuan. 1993. Multiple comparisons with the best for steady-state simulation. Technical note in *ACM Transactions on Modeling and Computer Simulation* 3:66–79.

Paulson, E. 1952. On the comparison of several experimental categories with a control. *Annals of Mathematical Statistics* 23:239–246.

Ramberg, J. S., S. M. Sanchez, P. J. Sanchez, and L. J. Hollick. 1991. Designing simulation experiments: Taguchi methods and response surface metamodels. *Proceedings of the 1991 Winter Simulation Conference*, ed. B. L. Nelson, W. D. Kelton, and G. M. Clark, 167–176. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.

Rinott, Y. 1978. On two-stage selection procedures and related probability inequalities. *Communications in Statistics: Theory and Methods* A7(8):799–811.

Sanchez, S. M. 1987a. A modified least-failures sampling procedure for Bernoulli subset selection. *Communications in Statistics—Theory and Methods* 16(12):3609–3629.

Sanchez, S. M. 1987b. Small-sample performance of a modified least-failures sampling procedure for Bernoulli subset selection, *Communications in Statistics—Simulation and Computation* 16(4):1051–1065.

Sanchez, S. M. 1994. A robust design tutorial. *Proceedings of the 1994 Winter Simulation Conference*, eds. J. D. Tew, M. S. Manivannan, D. A. Sadowski, and A. F. Seila, 106–113. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.

Sanchez, S. M., and J. L. Higle. 1992. Observational studies of rare events: a subset selection approach. *Journal of the American Statistical Association* 87(149):878–883.

Sanchez, S. M., P. J. Sanchez, J. S. Ramberg and F. Moeeni. 1996. Effective engineering design through simulation. *International Transactions on Operational Research* 3(2):169–185.

Taguchi, G. 1986. *Introduction to quality engineering.* White Plains, New York: UNIPUB/Krauss International.

Tamhane, A. C. 1977. Multiple comparisons in model I: One-way ANOVA with unequal variances. *Communications in Statistics–Theory and Methods* A6:15–32.

Thesen, A., and L. E. Travis. 1992. *Simulation for decision making.* St. Paul, Minnesota: West Publishing Company.

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.

AUTHOR BIOGRAPHY

SUSAN M. SANCHEZ is Associate Professor of Management Science and Information Systems at the University of Missouri–St. Louis. Her research interests include the design and analysis of simulation experiments, selection, and data-intensive statistical methods in business and manufacturing. She is an associate editor for *Operations Research* and *Naval Research Logistics*.