Proceedings of the Operational Research Society Simulation Workshop 2016 (SW16)
A.Anagnostou, K. Hoad, M. Kunc eds.

ARGUMENTS FOR AND AGAINST THE USE OF MULTIPLE COMPARISON CONTROL IN STOCHASTIC SIMULATION STUDIES

Dr. Thomas Monks

NIHR CLAHRC Wessex
University of Southampton
thomas.monks@soton.ac.uk

Dr. Christine Currie

University of Southampton
Southampton
Christine.Currie@soton.ac.uk

Dr Kathryn Hoad

University of Warwick
Coventry
Kathryn.Hoad@wbs.ac.uk

ABSTRACT

Pick up any of the standard discrete-event simulation textbooks and you will find that the output analysis section includes a note on multiple comparison control (MCC). These procedures aim to mitigate the problem of inflating the probability of making a single type I error when comparing many simulated scenarios simultaneously. We consider the use of MCC in stochastic simulation studies and present an argument discouraging its use in the classical sense. In particular, we focus on the impracticality of procedures, the benefits of common random numbers and that simulation is very different from empirical studies where MCC has its roots. We then consider in what instances would abandoning MCC altogether be problematic and what alternatives are available. We present an argument for medium to large exploratory studies to move their attention away from classical Type I errors and instead control a subtlety different quantity: the rate of false positives amongst all 'discoveries'.

Keywords: Output Analysis; Comparison Procedures; Multiple Comparison Control

1 INTRODUCTION

This paper is about making mistakes in stochastic simulation studies. The type of mistake here refers to incorrect interpretation of the stochastic output of a discrete-event simulation (DES) model, as opposed to a mistake in the coding of a model. We focus on mistakes in correctly identifying differences between performance measures in competing configurations (scenarios) of a model when there are several or many alternatives to compare or more formally making a Type I or II error in inference. This is the multiple comparison problem (MCP) that is well known and studied in the empirical sciences. The MCP applies to the plethora of comparison procedures found in the DES literature including more advanced selection procedures. However, most evidence from simulation practice illustrates that basic comparison procedures are dominant (Hoad and Monks 2011; Hoad et al. 2014) and when the MCP is tackled it is done so using multiple comparison control (MCC) advocated in standard DES text books (Hoad and Monks 2011). We consider the use of MCC in stochastic simulation studies in practice and present an argument discouraging its use in the classical sense. In particular, we focus on the impracticality of

procedures, the benefits of common random numbers and that simulation is very different from empirical studies where MCC has its roots. We then consider in what instances would abandoning MCC altogether be problematic and what alternatives are available. We present an argument for medium to large exploratory studies to move their attention away from classical Type I errors and instead control a subtlety different quantity: the rate of false positives amongst all 'discoveries'.

We present our argument as follows. First we formally define the MCP and how MCC has been approached in DES in both standard comparison and more advanced selection procedures. Second we detail an arguments for and against the use of MCC in practice with final arguments presenting an alternative formulation of the MCP.

2 THE MULTIPLE COMPARISON PROBLEM

At the heart of the classical multiple comparison problem (MCP) lies the Familywise Error Rate (FWER). The FWER represents the probability of making a single type I error (incorrectly rejecting the null hypothesis). For example, when simulating the weekly throughput of two competing configurations (scenarios) of a manufacturing line a Type I error is equivalent to incorrectly concluding that two configurations have different throughputs when in fact they are the same. In stochastic simulation studies an often, but not exclusively, used method of comparison is to construct a 95% confidence interval of the difference between mean throughputs of the two scenarios. In this instance the probability of making a type I error is 5%.

Now consider a manufacturing line where we have five competing scenarios. If we are aiming to choose the best system a simple procedure is to and compare all scenarios against each other in a pairwise fashion. This requires us to construct ten simultaneous confidence intervals. The MCP is the inflation of the FWER when making multiple simultaneous comparisons. The probability of a making a single type I error is given by [1].

$$P(Making\ at\ least\ one\ type\ I\ error\ in\ m\ tests) = 1 - (1-\alpha)^m \quad (1)$$

The consequence of [1] is that if we perform ten pairwise comparisons using 95% confidence intervals then the chance of at least one false positive result is 40%. If we perform 100 comparisons the chance of a false positive result is almost certain at 99.4%. Figure 1 illustrates the inflation of the FWER as the number of comparisons ranges between one and 100.

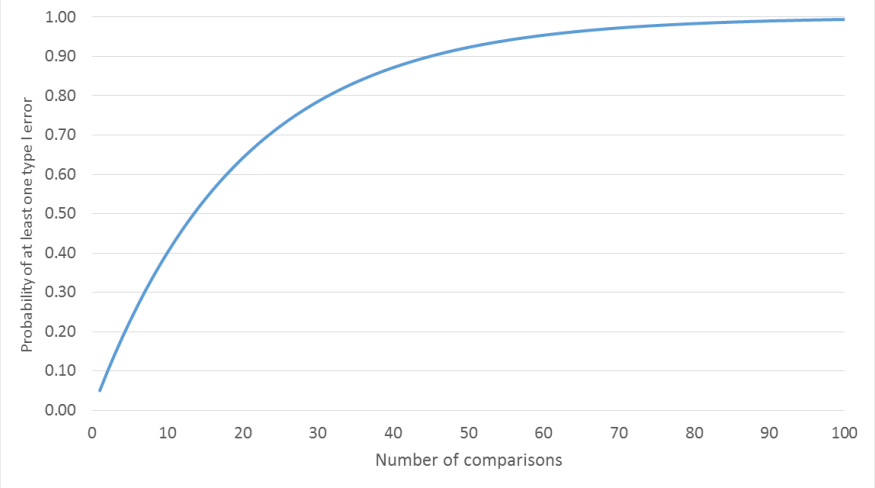

Figure 1 *FWER inflation as the number of comparisons increases. α= 0.05*

3 MULTIPLE COMPARISON CONTROL IN STOCHASTIC SIMULATION STUDIES

We now provide an overview of multiple comparison procedures and control in discrete-event simulation studies. It is worth noting that the use of MCC should depend on the objectives of the simulation study. If the objectives are to optimise a simulation model then selection procedures are more appropraite than MCC. We therefore detail a 'textbook' procedure - the Bonferroni Correction; other classical MCC procedures and an overview of selection procedures..

3.1 The Bonferroni Correction

Given the MCP described our standard discrete-event simulation (DES) texts advise on the use of multiple comparison control (MCC) using the *Bonferroni Correction* to adjust the α level used (see Banks 2005 p449; Law 2006 p537; Robinson 2004 p180). A Bonferroni Correction ensures that the overall probability of making a Type I error remains at level α. This is achieved by adjusting the individual α level of the confidence intervals to α/m. So in the case of six comparisons an individual α level of 0.05/6 would be used in each comparison to maintain the FWER. The Bonferroni Correction has the nice property that it makes no assumptions about the independence of scenarios.

3.2 Selection Procedures

The aim of most simulation projects in which the multiple comparison problem arises is to identify the optimal scenario with respect to a particular outcome measure. This can be thought of as optimization via simulation, where the number of alternatives is finite, and the associated methods are often described as selection procedures. When employing simulation optimization it is then necessary to consider the particular algorithm to use and it is also necessary to decide how to decide whether one system is significantly better than the other. As discussed Bonferroni is one way of doing. There are other ways for calculating the probability of correct selection. For example, these can be approximated using the Slepian inequality (Branke et al., 2007) Branke et al. (2007) provide a comparison of different selection procedures for simulation and although they only discuss the situation when the outputs are independent and normally distributed, it still provides an excellent description of the best-known selection procedures: indifference zone (IZ); expected value of information; and optimal computing budget allocation (OCBA).

Of relevance to this article is the discussion of how to measure the quality of a selection procedure. Three measures are worth considering and are discussed in more detail by Branke et al. (2007):

1. Zero-one loss function: $L_{0-1}(D, \boldsymbol{w}) = \mathbf{1}\{w_D \neq w_{[k]}\}$, i.e. the loss-function equals 1 if the wrong selection is made and 0 if the correct selection is made.
2. The opportunity cost: $L_{OC}(D, \boldsymbol{w}) = w_{[k]} - w_D$, which equals 0 if the correct selection is made and is the difference between the optimal value and the selected value if the incorrect selection is made. OCBA procedures are set up to terminate when a selection *D* has been made and the expected opportunity cost (EOC) associated with this selection *D*, is less than some pre-defined tolerance.
3. In IZ procedures, the measure of performance is the probability of correct selection (PCS) subject to the constraint that the mean of the best solution is at least $\delta > 0$ better than the others. Here, δ is known as the indifference zone parameter and is defined as the smallest difference in the outputs that the decision-maker believes to be significant.

The focus of research into selection procedures tends to be on improving the efficiency of the sampling whilst ensuring that the relevant quality measure (EOC or PCS) is maintained.

Much of the focus of academic research into selection procedures assumes that simulation outputs from different options are independent, which is not the case when common random numbers (CRN) are used; although there are some notable exceptions (e.g. Chick and Inoue 2001; Nakayama 2007; Nelson and Matejcik 1995). The use of CRN induces positive correlation in the outputs of samples coming from

different systems. By exploiting this positive correlation, it is possible to reduce the number of simulation iterations needed to obtain a set level of confidence, hence the recommendation to use them in all of the standard simulation textbooks. However, it is more difficult to exploit this positive correlation structure in complex selection problems. Nakayama (2007) uses an indifference zone method to show how this can be done and he demonstrates through an example how calculating and taking account of the correlation structure improves the efficiency of the sampling over and above methods in which this correlation structure is ignored, e.g. when using the Bonferroni inequalities.

3.3 Other classical MCC procedures

The Bonferroni Correction detailed in DES text books is not the only MCC procedure available. For a review of alternative procedures see Hoad and Monks (2011). Here we note that the Bonferroni Correction we detail in Section 3.1 is the most *conservative* approach to controlling the FWER when constructing multiple simultaneous confidence intervals. Hoad and Monks (2011) found that the sequential step-down Bonferroni procedure (Holm, 1979) was less conservative. See Serlin (1993) and Ludbrook (2000) for the procedure to adapt Holm's original method, based on p-values, to construct simultaneous confidence intervals.

4 AN ARGUMENT AGAINST USE OF MCC

Here we list three arguments against the use of MCC in simulation studies: the practical difficulties in applying the Bonferroni Correction and less conservative FWER MCC to even small scale studies; the use of common random numbers in scenarios; and the differences between empirical and simulation studies.

4.1 The practical difficulties with controlling FWER

A particular practical difficulty arises in medium to large scale studies when attempting to control the FWER. We illustrate this with the Bonferroni Correction procedure outlined in Section 3. Table 1 illustrates the individual confidence intervals needed to maintain a FWER of 5% when comparing up to ten scenarios in a full pairwise fashion. Note how quickly the Bonferroni Correction effects the practical usefulness of results. For instance, beyond five scenarios the individual confidence intervals are above 99.5%. Not only is this incredibly strict, but any useful information of the mean difference between scenarios is lost. Note that even if an overall 10% level of significance was set the individual confidence intervals become extremely strict and difficult to interpret (i.e. approximately 99.5% and above) beyond seven scenarios. It is noteworthy that Hoad and Monks (2011) surveyed 25 simulation practitioners and reported that 36% had conducted one simulation study where the number of scenarios exceeded 100. In these instances using the Bonferroni Correction can lead to many Type II errors.

A further practical difficulty focusing on FWERs is that there is no single statistical definition for what constitutes a family. For example, reconsider our simulation model of a manufacturing line. In addition to throughput we are also interested in comparing the utilization of an expensive resource across the six scenarios. If we are very concerned about FWER then utilization and throughput could be considered as part of the same family. This leads to an individual α level of $0.05/20 = 0.0025$. Moreover, what happens if at a later date a further study using the manufacturing line in conducted where new alternatives are tested? Should a modeler consider these new comparisons as part of the same family and hence apply an even stricter per comparison CI? We might even find that previous results that were 'significant' are no longer so as we must backwards apply our stricter Bonferroni Correction to the original study.

Table 1: *Feasible and Infeasible use of a Bonferroni Correction in full pairwise comparisons (α=0.05)*

Scenarios	Comparisons (m)	α (Bonferroni)	CIs	
3	3	0.017	98.3	Interpretable, but wide CI's
4	6	0.008	99.2	
5	10	0.005	99.5	
6	15	0.003	99.7	Strict and CIs of little use
7	21	0.002	99.8	
8	28	0.002	99.8	
9	36	0.001	99.9	
10	45	0.001	99.9	

4.2 Common Random Numbers

In simple terms common (pseudo) random number (CRN) streams assigned to each simulation activity provide a way to synchronize scenarios. The use of the same pseudo random number (PRN) streams induces a positive correlation between the scenarios to compare (see Law, 2006 and Pidd, 2004 for further details). The general purpose of synchronization is as a variance reduction technique (see equation 2) where the induced correlation reduces results in the variance of the difference between scenarios. In addition to variance reduction we argue that CRN greatly reduce the probability of a Type I error.

$$Var(s_1 - s_2) = Var(s_1) + Var(s_2) - 2Cov(s_1, s_2) \quad (2)$$

When CRN work the observed difference between scenarios is extremely likely to be due to the differences in input parameters and not a Type I error due to random sampling 'noise'. As a simple example, consider a DES model of a single server queue with *s* servers illustrated in Figure 2. If we use separate dedicated PRN streams for the inter-arrival and service times and vary *s* then we know that each entity has the same arrival time and same service time and can deduce that it is only queuing time that varies between scenarios. Now consider the FWER and Figure 1 with respect to CRNs. If we vary *s* so that we conduct 100 comparisons between scenarios with a positive dependency is it valid to conclude that the probability of making a single Type 1 error is inflated to near certainty? If the FWER is a function of random noise between scenarios then it follows that if this noise is removed (or greatly reduced) via CRN then the FWER cannot inflate as is commonly stated. Application of MCC procedures such as the Bonferroni Correction is therefore highly conservative and unnecessary leading to increased Type II errors. This, is however, a simple example where no changes have been made to service or arrival distributions. In section 6 we consider situations where scenarios may not synchronize.

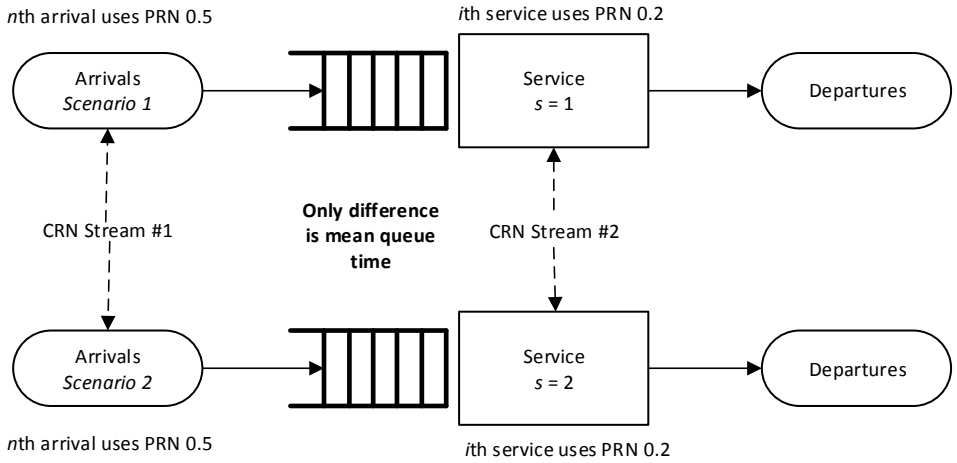

Figure 2 *Example use of CRN to synchronize scenarios and remove the chance of a type I error*
PRN = Pseudo Random Number; scenarios have the same IAT and service distributions

4.3 Simulation is not a classical empirical study

The elephant in the room is, of course, that simulation is quite different from a classical empirical study such as those found in the natural, biological or social sciences. In these cases an experiment is run that simulation modelers might consider a 'black box' where the natural, biological or psychological mechanism might not be accessible. In contrast a DES model is a 'white box': the internal mechanism – the model logic - is completely accessible and understood to the simulation modeler who created it. As such any unexpected findings can be explored in more detail at the mechanism level i.e. the model can be rerun under the same conditions and additional explanatory information can be exported as part of verification and validation of results.

The ability to rerun a model raises another important difference between classic empirical studies and simulation and in fact a main reason to choose simulation over a real world study (Pidd 2004). That is, that replication, even of models that take several days to run, is relatively cheap compared to real world studies.

5 AN ARGUMENT FOR THE USE OF MCC

We now consider in what instances would abandoning MCC altogether be problematic and what alternatives are available. We first illustrate an instance where CRN are unsuccessful.. We then present an argument for medium to large exploratory studies to move their attention away from classical Type I errors and instead control a subtlety different quantity: the rate of false positives amongst all 'discoveries'.

5.1 Common random numbers do not guarantee synchronization

In Section 5 we outlined the role that CRNs and synchronized scenarios play in removing the need to control the FWER in stochastic simulation studies. Unfortunately there is no way of guaranteeing that CRN will work. We now consider the case where scenarios may not perfectly synchronize, as there may be differences in service time distributions, inter-arrival rates and entity routing logic; and as such not all difference in a point estimate may be explained by the different input parameter sets. Moreover there maybe input parameter sets where CRN fail to work at all, may not be used, or in some cases 'backfire' (Law 2006) and increase the variance between scenarios.

As an illustration of what can happen without synchronization we recreate and adapt a simple M/M/S example from Law (2006). We have two competing designs. The first design has a single server and a

traffic intensity $\rho = 0.9$ ($\lambda = 1$ min^{-1}; $\mu = 0.9$). The second design has two servers and also a traffic intensity of $\rho = 0.9$ ($\lambda = 1$ min^{-1}; $\mu = 1.8$). Solving this analytically we know that design two has the lower average queuing time. Using CRN and 100 replications leads to only *three* individual runs where the incorrect decision would have been made, i.e. design one appears superior. On average these three runs had a difference, the wrong direction, of 0.2 minutes. If we force the pseudo random number streams out of sync then the number of individual runs where an incorrect decision would be made rises to 40. On average these 40 runs had a difference of 2.2 minutes. Such an example is often used to illustrate variance reduction i.e. CRN reduce the number of runs required to accurately estimate a mean difference. Of equal importance is that rates of Type I errors will increase without synchronization and an inflation of the FWER can be expected; particularly in cases where multiple simultaneous comparisons are made.

5.2 Reframing the MCP

Thus far we have defined the MCP in terms of controlling the FWER: the probability of making a single Type I error in *m* comparisons. The importance of avoiding Type I errors is built on the assumption that the role of the simulation study is to assist a decision maker is to pick the best system scenario out of n scenarios. However, if the role of the simulation study is more exploratory or if the decision maker is equally concerned about Type II errors then the MCP can be reframed in terms of Benjamini and Hochberg (1995) false discovery rate (FDR).

Firstly, let us clarify the difference between FDR and FWER. Consider carrying out m statistical hypothesis tests. Of these m individual tests, there are m_0 true null hypotheses and m_1 (= m-m_0) false null hypotheses. All the possible outcomes of these m tests are shown in Table 2 (Benjamini and Hochberg, 1995). Obviously, in reality, only m and r (the total number of significant results) are known. The FDR is defined by Benjamini and Hochberg (1995) as the expected proportion of type I errors (V) among all the significant results (r), i.e. E[V/r]. The FWER, however, is defined as the probability that the number of type I errors (V) is greater than or equal to one, i.e. $P(V \geq 1)$. That is, the FWER refers to the probability of making a single type I error in m comparisons.

Table 2 *Outcomes of a scenario comparison*

Actual result	Decision		Total
	H_0 not rejected	H_0 rejected	
H_0 true	U	V	m_0
H_0 false	T	S	m_1
Total	m-r	r	m

V = the number of type I errors; T = the number of type II errors. Only m, r and m-r are observable. U, V, T, S and m_1 are unknown.

A useful property of FDR is that it also controls the FWER in the *weak sense* (Benjamini and Hochberg, 1995). Consider an experiment where *M* comparisons between scenarios are made. If all of the *M* null hypotheses are true (i.e. there is no evidence to suggest differences between scenarios) then FDR is equivalent to Bonferroni. These claims are backed up by several simulation studies (e.g. Benjamini and Hochberg 1995; Benjamini and Hochberg 2000). It has also been shown that the FDR approach can be a more powerful method than the Bonferroni method (Benjamini and Hochberg, 1995; Hoad and Monks, 2011). Since that first paper many 'improved' or otherwise connected methods have been published (e.g. Benjamini and Hochberg 2000; Benjamini and Yekutieli 2001; Genovese and Wasserman 2002; Storey 2002; Storey and Tibshirani 2003; Verhoeven et al. 2005).

This concept of FDR control has gained popularity in a number of disciplines e.g. evolution, ecology, biology, genetics (Benjamini and Yekutieli 2001; Garcia 2004; García 2003; Verhoeven et al. 2005), especially those that are more exploratory in nature, where large numbers of hypotheses are required to be

tested, but where the strict control of the FWER can be relaxed (Black 2004). It is this idea of exploration that is particularly compelling. In (DES) experimentation it can be argued that two main objectives exist: to explore the solution space in order to learn more about the 'important' factors and solution possibilities and/or to find the optimal solution or 'best of a subset' of possible solutions. It is important when your aim is to find the 'best' scenario, that the probability of making a type I error, an erroneous discovery, is kept to a minimum. The Bonferroni type methods strictly control the probably of making one or more type I errors, but at a cost of power, hence increasing the risk of overlooking 'real' differences between scenarios. It can also be argued that when the main aim is to explore the solution space by comparing large numbers of scenarios, it is important to try to reduce the number of type II errors, while still keeping some control of type I errors. The family of FDR controlling methods achieve this aim. Appendix 1 details the procedure for the basic FDR.

6 CONCLUSION

Pick up any of the standard DES textbooks and you will find that the output analysis section includes a note on MCC. The procedures detailed - most typically the Bonferroni Correction - aim to mitigate the problem of inflating the probability of making a *single* type I error when comparing many simulated scenarios simultaneously. Use of conservative procedures such as Bonferroni fail to take account of the difference between classical empirical research and simulation. DES studies provide enhanced experimental control, for example through the use of common random numbers, compared to a classical experiment. Moreover, the mechanism studied in the experiment - the model - is accessible and understood by its coder; therefore greatly reducing the chances of accepting an erroneous result at face value. In practice it would appear that such procedures are largely ignored (Hoad and Monks, 2011); however, we argue that it is possible to make a Type I error in a stochastic simulation study albeit at a lower rate than seen in empirical studies. In practice then we propose a reframing of the MCP with the use of FDR based procedures that provide a better trade-off between Type I and II errors.

We acknowledge that we have primarily focused on simple comparison procedures. Further work will consider complex problems that may involve ranking and selection procedures or large complex models.

ACKNOWLEDGMENTS

TM is funded by the National Institute for Health Research (NIHR) Collaborations for Leadership in Applied Health Research and Care (CLAHRC) Wessex. The views expressed in this publication are those of the author and not necessarily those of the National Health Service, the NIHR, or the Department of Health.

APPENDICES

A. FALSE DISCOVERY RATE PROCEDURE

1. Carry out m hypothesis tests (i.e. scenario comparisons) and calculate m corresponding p-values.
2. Rank the p-values in ascending order: $p^{(1)} \leq p^{(2)} \leq p^{(3)} \leq \ldots \leq p^{(m)}$
3. Define $H^{(i)}$ as the null hypothesis associated with the p-value $p^{(i)}$
4. For i = m, m-1, m-2, … , 1, let n be the largest i for which $p^{(i)} \leq i\alpha/m$.
5. Reject all null hypotheses from $H^{(1)}$ up to and including $H^{(n)}$ (i.e. reject the null hypothesis corresponding to $p^{(n)}$ as well as all those having smaller p-values).

REFERENCES

Banks, J., J.S.Carson II, B.L.Nelson, D.M.Nicol, . 2005. *Discrete-Event System Simulation,* . 4th ed. NJ: Prentice Hall Int.,

Benjamini, Y., and Y. Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1):289-300.

Benjamini, Y., and Y. Hochberg. 2000. "On the Adaptive Control of the False Discovery Rate in Multiple Testing with Independent Statistics." *Journal of Educational and Behavioral Statistics* 25 (1):60-83.

Benjamini, Y., and D. Yekutieli. 2001. "The Control of the False Discovery Rate in Multiple Testing under Dependency."1165-1188.

Black, M. A. 2004. "A Note on the Adaptive Control of False Discovery Rates." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66 (2):297-304.

Branke, J., S. E. Chick, and C. Schmidt. 2007. "Selecting a Selection Procedure." *Management Science* 53 (12):1916-1932.

Chick, S. E., and K. Inoue. 2001. "New Procedures to Select the Best Simulated System Using Common Random Numbers." *Management Science* 47 (8):1133-1149.

Garcia, L. V. 2004. "Escaping the Bonferroni Iron Claw in Ecological Studies." *OIKOS* 105:657-663.

García, L. V. 2003. "Controlling the False Discovery Rate in Ecological Research." *Trends in Ecology & Evolution* 18 (11):553-554.

Genovese, C., and L. Wasserman. 2002. "Operating Characteristics and Extensions of the False Discovery Rate Procedure." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 (3):499-517.

Hoad, K., and T. Monks. 2011. "A Note on the Use of Multiple Comparison Scenario Techniques in Education and Practice." In *Proceedings of the 2011 Winter Simulation Conference (WSC)*, edited by R. R. C. S. Jain, J. Himmelspach, K.P. White, and M. Fu, Institute of Electrical and Electronics Engineers,Inc, Piscataway, New Jersey.

Hoad, K., T. Monks, and F. O'Brien. 2014. "The Use of Search Experimentation in Discrete-Event Simulation Practice." *Journal of the Operational Research Society* IN PRESS.

Law, A. M. 2006. *Simulation Modelling and Analysis*. Boston: McGraw-Hill International

Ludbrook, J. 2000. "Multiple Inferences Using Confidence Intervals." *Clin Exp Pharmacol Physiol* 27 (3):212-215.

Nakayama, M. K. 2007. "Fixed-Width Multiple-Comparison Procedures Using Common Random Numbers for Steady-State Simulations." *European Journal of Operational Research* 182 (3):1330-1349.

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.

Pidd, M. 2004. *Computer Simulation in Management Science*. London: John Wiley and Sons

Robinson, S. 2004. *Simulation: The Practice of Model Development and Use*. London: John Wiley and Sons

Serlin, R. C. 1993. "Confidence Intervals and the Scientific Method: A Case for Holm on the Range." *The Journal of Experimental Education* 61 (4):350-360.

Storey, J. D. 2002. "A Direct Approach to False Discovery Rates." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 (3):479-498.

Storey, J. D., and R. Tibshirani. 2003. "Statistical Significance for Genomewide Studies." *Proceedings of the National Academy of Sciences* 100 (16):9440-9445.

Verhoeven, K. J. F., K. L. Simonsen, and L. M. McIntyre. 2005. "Implementing False Discovery Rate Control: Increasing Your Power." *Oikos* 108 (3):643-647.

AUTHOR BIOGRAPHIES

THOMAS MONKS is a Senior Research Fellow in Operational Research in the Faculty of Health Sciences, University of Southampton. He holds a BSc in Computer Science and Mathematics, MSc in Operational Research and PhD in Simulation Modelling. He has worked as both a Software Engineer in the private sector and an Operational Research Analyst within the public sector. His specialty is simulation of healthcare systems of unscheduled and emergency care. He is co-chair of the UKs Simulation Workshop 2016 and co-chair of the UKs simulation special interest group. http://www.southampton.ac.uk/healthsciences/about/staff/tm3y13.page

CHRISTINE CURRIE is Associate Professor of Operational Research in Mathematical Sciences at the University of Southampton, UK, where she also obtained her Ph.D. She is Editor-in-Chief for the Journal of Simulation. Christine was co-chair of the Simulation Special Interest Group in the UK Operational Research Society until September 2013. Her research interests include mathematical modelling of epidemics, Bayesian statistics, revenue management, variance reduction methods and optimization of simulation models. See http://www.southampton.ac.uk/maths/about/staff/ccurrie.page for more details.

KATHRYN HOAD is a Senior Teaching Fellow at Warwick Business School, University of Warwick. Katy has research and teaching experience in forecasting, statistical analysis and discrete event simulation. She spent many years as co-organiser of the UK Operational Research Society Special Interest Group in Simulation, and over many years has taken on various roles in the running of conferences such as the UK OR Society Annual Conferences, Winter Simulation Conference and Simulation Workshop 16. http://www.wbs.ac.uk/about/person/kathryn-hoad