BEST-SUBSET SELECTION FOR COMPLEX SYSTEMS USING AGENT-BASED SIMULATION

by

Yu Wang

B.S., Mechanical Engineering, Tsinghua University, 2000

M.S., Mechanical Engineering, Tsinghua University, 2003

M.S., Industrial Engineering, University of Pittsburgh, 2007

Submitted to the Graduate Faculty of

the Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This dissertation was presented

by

Yu Wang

It was defended on

November 21, 2011

and approved by

Larry J. Shuman, Ph.D., Professor, Department of Industrial Engineering

Louis K. Luangkesorn, Ph.D., Research Assistant Professor, Department of Industrial Engineering

Bopaya Bidanda, Ph.D., Professor, Department of Industrial Engineering

Carey D. Balaban, Ph.D., Professor, School of Medicine

Bryan A. Norman, Ph.D., Associate Professor, Department of Industrial Engineering

Bruce Y. Lee, MD, MBA, Assistant Professor, School of Medicine

Dissertation Directors: Larry J. Shuman, Ph.D., Professor, Department of Industrial Engineering,

Louis K. Luangkesorn, Ph.D., Research Assistant Professor, Department of Industrial Engineering

BEST-SUBSET SELECTION FOR COMPLEX SYSTEMS USING AGENT-BASED SIMULATION

Yu Wang, PhD

University of Pittsburgh, 2011

It is difficult to analyze and determine strategies to control complex systems due to their inherent complexity. The complex interactions among elements make it difficult to develop and test decision makers' intuition of how the system will behave under different policies. Computer models are often used to simulate the system and to observe both direct and indirect effects of alternative interventions. However, many decision makers are unwilling to concede complete control to a computer model because of the abstractions in the model, and the other factors that cannot be modeled, such as physical, human, social and organizational relationship constraints. This dissertation develops an agent-based simulation (ABS) model to analyze a complex system and its policy alternatives, and contributes a best-subset selection (BSS) procedure that provides a group of good performing alternatives to which decision makers can then apply their subject and context knowledge in making a final decision for implementation.

As a specific example of a complex system, a mass casualty incident (MCI) response system was simulated using an ABS model consisting of three interrelated sub-systems. The model was then validated by a series of sensitivity analysis experiments.

The model provides a good test bed to evaluate various evacuation policies. In order to find the best policy that minimizes the overall mortality, two ranking-and-selection (R&S) procedures from the literature (Rinott (1978) and Kim and Nelson (2001)) were implemented and compared. Then a new best-subset selection (BSS) procedure was developed to efficiently select a statistically guaranteed best-subset containing all alternatives that are "close enough" to the best one for a pre-

specified probability. Extensive numerical experiments were organized to prove the effectiveness and demonstrate the performance of the BSS procedure.

The BSS procedure was then implemented in conjunction with the MCI ABS model to select the best evacuation policies. The experimental results demonstrate the feasibility and effectiveness of our agent-based optimization methodology for complex system policy evaluation and selection.

Keywords: Agent-based simulation, Statistical selection procedure, Ranking-and-selection, Best-subset selection, Incident response simulation, Mass casualty incident response, Complex system, Complex adaptive system, Optimization via simulation.

TABLE OF CONTENTS

PREFACE . . . xiv
1.0 INTRODUCTION . . . 1
1.1 Research motivation . . . 2
1.2 Problem statement . . . 3
1.3 Contributions . . . 6
1.4 Overview of the dissertation . . . 8
2.0 LITERATURE REVIEW . . . 10
2.1 Complex adaptive system (CAS) . . . 10
2.2 Overview of emergency management . . . 14
2.3 Operations research (OR) models for emergency management . . . 16
2.3.1 Overview . . . 16
2.3.2 Analytical models . . . 17
2.3.3 Simulation models . . . 17
2.3.3.1 Discrete-event simulation (DES) . . . 18
2.3.3.2 Agent-based simulation (ABS) . . . 19
2.3.4 Discussion . . . 20
2.4 Agent-based simulation (ABS) modeling . . . 21
2.4.1 Overview of agents . . . 22
2.4.2 Agent-based modeling . . . 23
2.5 Optimization-via-simulation methods . . . 27
2.5.1 Ranking-and-selection (R&S) . . . 28
2.5.1.1 Indifference-zone selection (IZS) . . . 29

2.5.1.2 Subset selection (SS) 30
2.5.2 Multiple comparison procedures (MCP) 31
2.5.2.1 All-pairwise comparison approaches 31
2.5.2.2 Multiple comparisons with a control (MCC) 32
2.5.2.3 Multiple comparisons with the best (MCB) 32
2.5.3 Summary 33
2.5.4 Discussion 33
3.0 AGENT-BASED SIMULATION MODELING FOR MASS CASUALTY INCIDENT RESPONSE 35
3.1 Overview 35
3.1.1 General operations of MCI response 35
3.1.2 Discussion 37
3.1.3 Simulation platform selection 37
3.1.4 Highlights of the modeling 38
3.1.4.1 Generic agent types 38
3.1.4.2 GIS integration 41
3.1.5 System structure of MCI response 41
3.1.6 Incident site 43
3.1.6.1 On-site emergency medical technician (EMT) 43
3.1.6.2 Casualties 43
3.1.6.3 Casualty degradation 44
3.1.7 Pre-hospital sub-system 47
3.1.7.1 Ambulance 47
3.1.7.2 Incident command 49
3.1.8 In-hospital sub-system 50
3.1.9 Performance measure 52
3.1.10 Model validation 52
3.2 Case study 53
3.2.1 Assumptions, constraints and parameter settings 53
3.2.2 Transportation network construction 54

3.2.3 Evacuation policy 54
3.2.4 Parameter setting 55
3.2.5 Numeric experiments 55
3.2.5.1 Different number of casualties 57
3.2.5.2 Different injury severity distribution 57
3.2.5.3 Different percentage of specialized patients 58
3.2.5.4 Different degradation model 59
3.2.6 Summary 61
4.0 POLICY RANKING AND SELECTION 64
4.1 Introduction 64
4.2 Policy description 65
4.3 Simulation results 67
4.4 Select the best evacuation policy by R&S procedures 72
4.4.1 Rinott procedure 72
4.4.2 KN procedure 74
4.5 Discussion 77
5.0 BEST-SUBSET SELECTION PROCEDURE 79
5.1 Overview 79
5.2 Literature review 80
5.3 Problem formulation 82
5.4 Best-subset selection (BSS) procedure 84
5.5 Statistical validity proof 85
5.6 Numerical experiments 91
5.7 Comparing BSS with MCB 95
5.7.1 Multiple comparisons with the best (MCB) 95
5.7.2 MCB-based method for best-subset selection 98
5.8 Robustness analysis 102
5.8.1 Input parameters 102
5.8.1.1 Impact of indifference-zone width 102
5.8.1.2 Impact of different initial samples size 104

5.8.2 Changing variances 105
5.8.3 Slippage configurations (SCs) 107
5.8.4 Different distributions of competitive alternatives in each region 109
5.9 Summary 111
6.0 SELECT BEST EVACUATION POLICIES USING BSS 113
6.1 Overview 113
6.2 Experiment configuration 113
6.3 Experimental results 116
6.4 Sensitivity analysis 118
6.4.1 Impact of changing required precision of selection 118
6.4.2 Casualty setting 120
6.4.2.1 Degradation models 120
6.4.2.2 Different number of casualties 124
6.4.2.3 Different distributions of injury severity 127
6.4.2.4 Different percentage of specialized casualties 130
6.4.3 Hospital setting 133
6.4.3.1 Different number of Beds 133
6.4.3.2 Different admission criteria 135
6.5 Summary 138
7.0 CONCLUSIONS AND FUTURE RESEARCH 139
7.1 Summary and conclusions 139
7.2 Limitations discussion and future research directions 144
7.2.1 Possible extensions on simulation modeling 144
7.2.2 Possible directions on output analysis 146
APPENDIX. ACRONYMS 148
BIBLIOGRAPHY 150

LIST OF TABLES

2.1 Pros and Cons of ABS [1, 2] . . . 26
3.1 Delphi Estimates of degraded RPM scores in 30-Minute Intervals [3] . . . 46
3.2 Ambulance loading rules . . . 49
3.3 Simulation parameter settings . . . 56
3.4 Finding g by Fitting . . . 61
4.1 Twelve evacuation policies . . . 65
4.2 Simulation parameter settings . . . 68
4.3 R&S results by Rinott procedure . . . 73
4.4 R&S results by KN procedure . . . 76
4.5 Comparison of the number of samples required by Rinott and KN . . . 77
5.1 Mean and variance configurations for the first experiment (SC) . . . 92
5.2 Replication statistics of the first experiment . . . 92
5.3 Mean and variance configurations for the second experiment . . . 93
5.4 Replication statistics of the second experiment . . . 93
5.5 Means and variances of three competitive systems . . . 96
5.6 MCB Results ($k = 3$, $\alpha = 0.05$) . . . 97
5.7 Comparison between BSS and MCB (5000 replications for each) . . . 101
5.8 Competitive system configurations (input parameters) . . . 102
5.9 Results of different indifference-zone widths . . . 103
5.10 Results of different initial sample sizes . . . 104
5.11 Competitive system configurations (changing variances) . . . 106
5.12 Results of different changing variances . . . 106

5.13 Slippage configurations 108
5.14 BSS results of slippage configurations 108
5.15 System configurations – Different distribution of alternatives in each region 110
5.16 BSS Results – Different distributions in each region 110
6.1 Simulation parameter settings 115
6.2 Result of BSS procedure (key steps & statistics of mortality) 116
6.3 Result of BSS procedure – precision of selection $(\lambda^-, \lambda^+) = (0.01, 0.02)$ 119
6.4 Comparion between different precision settings 120
6.5 Results of BSS procedure – different degradation models 121
6.6 Results of BSS procedure – different number of casualties 124
6.7 Results of BSS procedure – different initial injury severity distributions 127
6.8 Results of BSS procedure – different percentages of specialized patients 130
6.9 Results of BSS procedure – different numbers of beds in a hospital 133
6.10 Results of BSS procedure – different admission thresholds 136

LIST OF FIGURES

1.1 Optimization-via-Simulation for complex system policy selection. 6
2.1 Characteristics of complex systems [4]. 11
2.2 Top-down and Bottom-up methodologies (Nilsson and Darley, 2006) [1] 24
2.3 Canonical view of an agent system (Jennings, 2000) [5] 25
3.1 Class diagram of agents 39
3.2 Structure diagram of MCI response system 42
3.3 RPM vs. Survival probability 46
3.4 Degradation in RPM scores under different initial conditions 47
3.5 Hospital process flow 51
3.6 Comparison among different number of casualties 57
3.7 Different injury severity distributions 58
3.8 Comparison among different injury severity distributions 59
3.9 Comparison among different percentages of specialized casualties 60
3.10 Comparison among different degradation models 62
4.1 Simulation results under 12 evacuation policies 69
4.2 Number of samples needed vs. sample variance 74
5.1 Three regions (Desired, Acceptable and Undesired) 83
5.2 Selection frequency of each alternative in the second experiment 94
5.3 Five possible distributions of confidence intervals 100
5.4 Sample sizes with different indifference-zone widths 103
5.5 Sample sizes with different initial sample sizes 105
5.6 Sample sizes with different changing variances 107

5.7 Sample sizes under different slippage configurations 109

5.8 Sample sizes of different alternative distributions in each region 111

6.1 Transportation network of Pittsburgh, PA 114

6.2 Results of BSS 117

6.3 Results of BSS – $(\lambda^-, \lambda^+) = (0.01, 0.02)$ 119

6.4 Comparison among different degradation models 123

6.5 Comparison among different numbers of casualites 126

6.6 Comparison among different initial injury severity distributions 129

6.7 Comparison among different percentages of specialized patients 132

6.8 Comparison among different numbers of beds in a hospital 134

6.9 Comparison among different admission thresholds used by hospitals 138

PREFACE

First and foremost, I would like to express my great gratitude to my advisors, Dr. Larry J. Shuman and Dr. Louis K. Luangkesorn for their generous guidance and support throughout the dissertation research. Without their inspiration and help, this work could not have been accomplished.

I also appreciate the other committee members, Dr. Carey D. Balaban, Dr. Bopaya Bidanda, Dr. Bryan A. Norman, and Dr. Bruce Y. Lee, for their insightful suggestions and comments, which make this dissertation more sound and accurate.

My gratitude also goes to the faculty and staff of the Department of Industrial Engineering for their warm help and assistance to me. I extend my special thanks to Dr. Michael R. Lovell and Dr. Christian Schunn for their direction and support in my first years of Ph.D. study. And I am grateful to Prof. Ken Sochats and the University of Pittsburgh Center for National Preparedness for funding and supporting my dissertation research work.

Thanks to all of my colleagues in Pitt IE and my friends around Pittsburgh for sharing an exciting and joyful life with me!

Last but not least, I am forever indebted to my cherished grandma and my wonderful parents, Yongzhu Wang and Lihua Shi. Their understanding, encouragement, patience and endless love have been the strongest support to me over these years.

This work is dedicated to them.

1.0 INTRODUCTION

While considerable insights can be gained from simple models, many problem domains are inherently complex, and that complexity needs to be well addressed in models in order to better understand and evaluate alternatives in actual systems. The objective of the research is to develop a set of optimization-via-simulation (OvS) methodologies to help analysts appropriately model a complex system and select the best control policies from a finite number of alternatives in an efficient way.

In this dissertation, we focus our attention on a specific example of a complex system - an emergency response system for a mass casualty incident (MCI). Our interest is to investigate how the system's performance measure (mortality) will change under different evacuation policies and to find a set of best evacuation policies which minimize overall mortality. To do this, we have analyzed a MCI response system and created an agent-based model to appropriately capture the complex interactions of different participants in the system, since these interactions may have great impacts on the system performance.

A simulation model can help in estimating the performances of different policies, but cannot provide direct answers about which ones are the best. To select the best policies (the ones that lead to the least mortality in this case), we investigated existing OvS techniques and identified their limitations for our problem of selecting a best-subset that contains all alternatives that are "close enough" to the unknown best. We extended fully-sequential ranking-and-selection (R&S) procedures to develop a new best-subset selection (BSS) procedure to help analysts compare and select the best policies efficiently with guaranteed statistical precision. The BSS procedure also provides an effective control mechanism to run the simulation model in a more scientific and efficient manner.

It should be noted that, although we have chosen an emergency response system as the study case, it does not imply our OvS methodologies are only valid for emergency response systems. Instead, the methodologies presented in this dissertation have general applicability and can be used for almost any complex system analysis and policy evaluation and selection.

1.1 RESEARCH MOTIVATION

Emergency managers are charged with planning, preparedness, response and mitigation to disasters within their jurisdiction. To do this well, managers must prepare plans in advance based on the resources available to them. Similarly, based on their plans, they can report to their jurisdictions that they have a resource shortfall. These plans and estimates for required resources are usually done by use of expert judgment and intuition and are then subject to acceptance by their jurisdictions. These can be augmented by the use of models to demonstrate the effects of different policy options or resources available by responders providing validity in the view of those who would approve any capital purchases or changes in policies.

Models are often used in enterprises for exploring various resource management policies. In particular, models are often used to explore policy options to react to unplanned disruptions. Applications include managing disruptions to commercial supply chains [6], reacting to disruptions in airline routes [7, 8], location of military equipment for global deployment [9], and others. These models are used to test procedures, challenge assumptions and explore new ideas more efficiently and rapidly than experimenting in the real world system [10]. While historically the operations research community has developed techniques that can be applied to homeland security topics, Wright et al. (2006) [11] find there are many rich opportunities that are still available, especially in the emergency response domain.

This research aims to investigate the MCI response system behaviors under different response policies by agent-based simulation model, and to find the best response policies using appropriate statistical analysis techniques. The motivation behind this research is to first demonstrate how a complex system (such as an emergency response system) could be conceptualized as a complex adaptive system and then modeled by agent-based simulation models, and then to provide decision makers a set of methodologies to efficiently run the simulation model, compare different configurations (policies) and find the best alternatives.

1.2 PROBLEM STATEMENT

As a specific instance of a complex system, the emergency response system is essentially a complex adaptive system (CAS) (see Section 2.1 for detail definitions of CAS), since it involves multiple interrelated sub-systems and interactive actors (e.g., injured casualties, responders, ambulances, and hospitals), and the outcome of system is affected by all of these participants' behaviors and their interactions, which leads to highly non-linear system behaviors and makes it difficult to predict the consequences of various control policies/protocols. For a better illustration, an example about dispatching policy selection in emergency response is given below.

Policy Selection: Destination hospital selection in evacuation

A common task of emergency response is casualty evacuation, especially to disaster events that involve a large number of victims. An efficient evacuation plan could effectively save lives and reduce mortality. In practice, the responders implementing evacuation (e.g., ambulances) usually follow instructions from the incident command (more specifically, the dispatching branch) that indicate which hospital a victim should be sent to. Whether or not an efficient evacuation can be organized directly depends on commanders' dispatching decisions, which requires the commanders to continuously analyze feedback information and make correct decisions in a timely manner.

Usually there exists standard operating procedures to facilitate the decision making for commanders. One possible strategy is to always use the nearest available hospital as the evacuation destination. Obviously, such a policy can effectively save transportation time, but may not be optimal in achieving a lower overall mortality. Assume that the nearest hospital is a specialized trauma center with certain specific capabilities that cannot be provided elsewhere, such a dispatching strategy is very likely to exhaust its capacity in a short period with those mild patients who suffer general injuries that can be treated anywhere, while those later-evacuated severe patients who indeed require specialized treatments have to be sent to faraway hospitals and cannot receive timely care, which greatly increases their risk of dying.

An alternative policy may take this issue into consideration, and reserve the capacity of specialized hospitals for potentially specialized patients. However, some questions may be raised such as

- Which specialized hospitals should be reserved?
- What percentage of capacity should be reserved in those hospitals?
- Should those capacity be reserved from the very beginning or only after certain conditions are satisfied?

Each of these questions may lead to a variant of the policy, and currently it is difficult to tell which one(s) are the "best", especially if there is a large number of alternatives existing.

In general, the difficulty in identifying "good" policies is in part because managers lack the appropriate decision-oriented tools to help them make an objective and confidential judgment. Review of past emergency cases could help people learn lessons and accumulate experiences [12], but it provides little help for predicting the effect of a new policy or identifying the "best" one from available candidates. In order to effectively study and analyze the system, a valid model is necessary. To fulfill our research purpose, a useful model should meet the following criteria:

- Reflect the essential system characteristics while simplifying the complexity effectively;
- Result with sufficient details to answer the questions about the system;
- Can be easily understood and extended

So our first question is: how should the response system be abstracted and modeled? It already has been proven that traditional analytical tools are insufficient for analysis of a complex system such as emergency response system [13, 14, 15]. A sophisticated mathematical model may be built to get some quick answers, but too many assumptions must be made to reduce the level of detail (e.g., the range of allowed interactions between system components), which may over-simplify the system and lead to an impractical conclusion.

Simulation can make explicit many of the abstractions needed for mathematical programming formulations and model the system in more detail. Different types of simulation models have been developed, among them agent-based simulation (ABS) models provide an intuitive way to capture the behaviors of a complex system from the ground-up. It enables researchers to translate their perceptions of individual processes into the knowledge about the complete system. For these reasons, ABS modeling technique will be adapted to simulate the whole response system. By altering input parameters, the model can simulate the dynamics of the system under different conditions and policies, and thus provide information about the effectiveness of the policies under different scenarios.

Many simulation models contain stochastic components to represent the uncertainty involved in the real world systems, therefore it is improper to run the model only once, then make a decision based on the single observation. Many replications must be run in order to accurately evaluate policies, and certain OvS techniques should be utilized to scientifically allocate computational resources to avoid either insufficient or excess observations, so that the samples can be obtained in an efficient way, which is especially important for computationally-intensive procedures, such as ABS.

Two categories of OvS techniques, ranking-and-selection (R&S) and multiple comparison procedures (MCP), are well suited for comparison and selection of competitive designs via simulation. R&S approaches are specifically developed to choose the best design(s) while MCP aim at providing inferences about the relationships among competing alternatives. Our investigation has focused on R&S since our goal is to select a set of best emergency response policies from multiple alternatives.

Within R&S, two classes of problem formulations are indifference-zone formulations and subset selection formulations. Currently available indifference-zone approaches are designed to select only a single best alternative with a guaranteed statistical precision, while subset selection methods either lack statistical guarantees on their selections or require too many input parameters to be practical. So to date there is not yet a satisfactory solution for selecting a statistically guaranteed best-subset of alternatives. To address this issue, our second research problem is to develop an efficient statistical methodology to select the best-subset from a finite number of competing alternatives while guaranteeing a pre-specified correct-selection probability (error) level.

1.3 CONTRIBUTIONS

This dissertation has two major contributions to the optimization-via-simulation (OvS) research area, on both complex system modeling and ranking-and-selection (R&S) methodology, as shown in Figure 1.1.

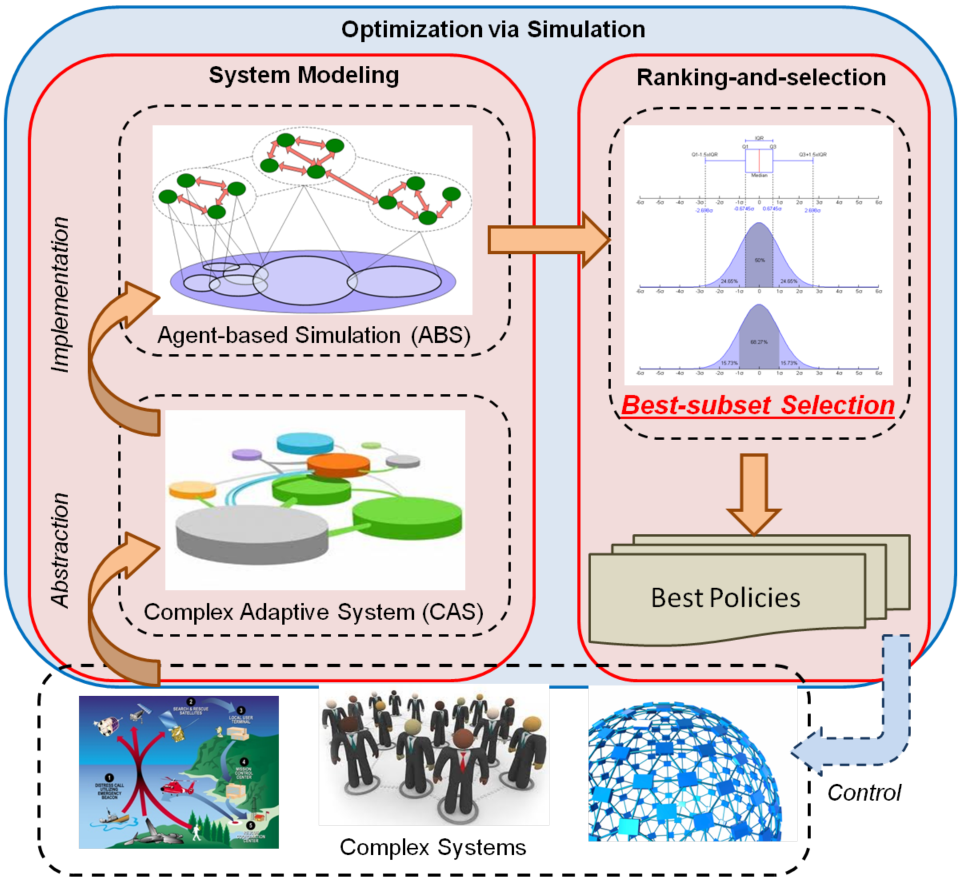


Figure 1.1: Optimization-via-Simulation for complex system policy selection.

For system modeling, after a comprehensive exploration of different modeling techniques, the following procedures are proposed to analyze a complex system (e.g., emergency response system for mass casualty incident). First, the complex system is abstracted as a complex adaptive system (CAS), which is a simplification to the original system. As the result of abstraction, major functional sub-systems are identified, as well as the important agents, their relationships and interaction rules. Those nonessential parts of the original system would be filtered out so that researchers can focus their efforts on those key factors. After all, as mentioned by Lee et al. [16, 17, 18], all computer models are simplifications of reality and can never account for every possible factor or interaction. For the simplified CAS, a flexible and scalable agent-based simulation (ABS) model architecture is proposed and implemented. The ABS model is used to investigate changes in system behavior under different control policies, so that it provides a good tool for managers to identify the best policies when used in combination with proper statistical analysis methods.

In practice, it is a common request from decision makers to have a method that can effectively compare different alternatives and provide a subset containing all "good enough" solutions, so that they can choose their final decision from the selected subset. A review of the literature suggests that there are no efficient procedures for selecting the best-subset with a specified statistical guarantee.

Therefore, the second major contribution of this dissertation is the development of a new fully-sequential R&S procedure to select the best-subset while satisfying the requirement of pre-specified correct-selection probability. The new procedure can select the best-subset efficiently by screening out obviously inferior alternatives in the early stages.

In addition to the two major methodology contributions, there are also some modeling contributions which supplement the emergency response simulation literature, as listed below:

- Integrates the ABS model with a geographical information system (GIS), that is, the pre-hospital transportation network can be automatically constructed based on geographic data generated from a given GIS shapefile [19]; and the ongoing status of the simulation (such as the location of each response vehicle, distribution of evacuated casualties in different hospitals, etc.) can be displayed in a GIS map view dynamically;
- Implements and compares different victim degradation models;
- Adds an in-hospital module to the pre-hospital care phase, so that hospital bed capacity can be included as a constraint; further, the hospital capacity can be broken down into specific tertiary treatment categories (e.g., burn, serve trauma, etc.)

1.4 OVERVIEW OF THE DISSERTATION

The remainder of this dissertation is organized as follows:

Chapter 2 provides a thorough literature review including all major concepts and techniques used in this dissertation. In the first section, an general introduction of complex adaptive system (CAS) is given to help readers understand what is a CAS and why CASs are difficult to analyze and control. The second section presents a brief overview of emergency management, and the third section reviews operations research models developed in emergency response research, and argues why an emergency response system can be conceptualized to a CAS and what modeling technique is appropriate for CAS analysis. The fourth and fifth sections discuss the agent-based simulation modeling and optimization-via-simulation (OvS) techniques in detail, respectively.

Chapter 3 develops an agent-based simulation (ABS) model of the response system for mass casualty incidents. The implementations of the model are explained in detail. Specific issues regarding emergency response simulation such as casualty degradation are addressed. Multiple experimental results are presented to verify and validate the model.

The MCI ABS model provides a good test bed for policy evaluation but can not help researchers find the best alternatives unless combined with certain OvS techniques. Chapter 4 implements two well-known ranking-and-selection (R&S) procedures (the Rinott and the KN procedures) and applies them to the ABS model developed in Chapter 3 to select the best evacuation policy. It also discusses the limitations of existing statistical selection procedures in selecting a subset that contains all alternatives that are "close enough" to the best.

To address this problem, Chapter 5 develops a new fully-sequential R&S procedure – the best-subset selection (BSS) procedure to address the inadequacy. The BSS procedure realizes efficient selection of the best alternative subset and provides an effective simulation control mechanism. The procedure is explained in detail, and theoretical proof of its statistical validity is provided as well. A series of numerical experiments are also given to test the procedure and demonstrate its effectiveness.

Chapter 6 shows how the new best-subset selection procedure can be applied to the ABS model to solve the emergency response policy evaluation and selection problem. Comprehensive computational results are provided to confirm the effectiveness of the methodology. Multiple sensitivity analysis experiments are organized to investigate the impacts of different factors to the policy selection results.

As the last chapter, Chapter 7 presents the summary and conclusions for the dissertation, and discusses some future research directions.

2.0 LITERATURE REVIEW

2.1 COMPLEX ADAPTIVE SYSTEM (CAS)

The concept of CAS originates in the life and physical sciences, and has been developed and widely used in the engineering and social science research, such as strategic organizational design, supply chain management and innovation management [20].

According to Ahmed et al. (2005) [21], almost all biological, economic and social systems can be conceptualized as complex adaptive systems (CASs). Examples of CAS include the biosphere and the ecosystem [22], industrial businesses [23, 20, 1], supply chain network [24, 25, 26, 27, 28], the stock market [29], and any human social group-based systems [30].

As a relatively new research field, there has not yet established a unified definition on the term CAS. North and Macal (2007) [14] presented the following definition for CAS in their book.

"*A complex adaptive system is a collection of interacting components with each of these components having its own rules and responsibilities. Some components may be more influential than others, but none completely controls the behavior of the complete system. All of the components contribute to the results in large or small ways.*"

Similarly, John H. Holland [31] defined CAS as a dynamic network of many agents (which may represent cells, species, individuals, firms, nations) which are constantly acting in parallel, and reacting to what the other agents are doing. Any coherent behavior in the system has to arise from competition and cooperation among the agents themselves. The overall behavior of the system is the result of a huge number of decisions made every moment by many individual agents.

In general, CAS is a special case of complex system. Chaffee and McNeil (2007) [4] provided a nice figure to depict typical complex systems, as reproduced in Figure 2.1, which presents many special characteristics of complex systems such as dynamically interacting, emergence, self-organization, evolution, etc.

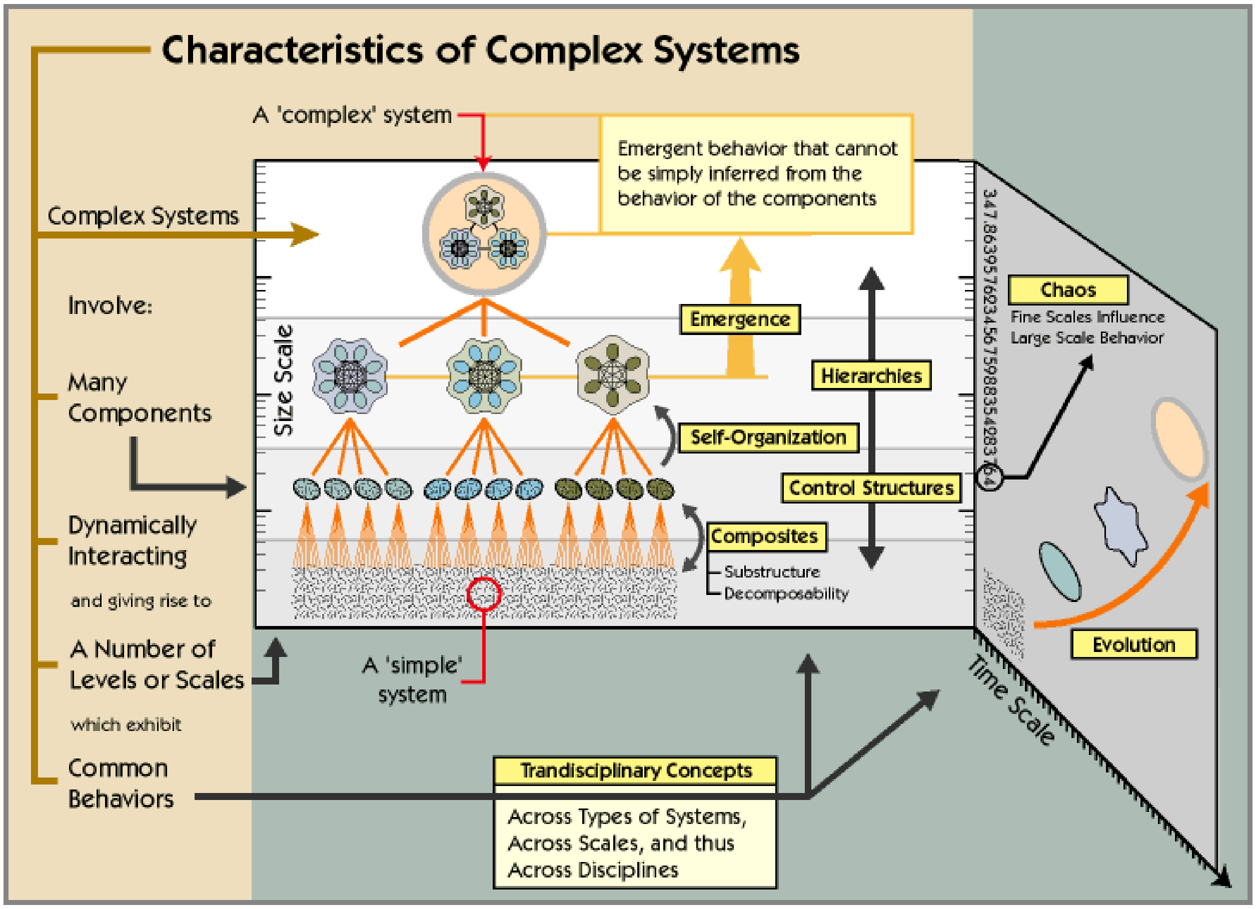


Figure 2.1: Characteristics of complex systems [4].

The key characteristic that differentiates CAS from general complex system is its agency, which refers to the ability to learn from experiences and to adapt to external changes. Not all complex systems have agency. For instance, water is a complex system but not a CAS, since its interacting objects (e.g. oxygen and hydrogen atoms) lack agency [25].

Although most CASs are complicated, CASs are not equivalent to complicated systems as well. McCarthy et al. (2006) [20] proposed a framework to distinguish CASs from complicated systems. In their framework, a system is defined as a set of elements with attributes that are connected to each other and to the environment by certain relationships. Four dimensions are also defined in the framework to identify the category of a system: (1) the number of elements that make up the system; (2) the attributes of the elements; (3) the number and type of interactions among the elements; and (4) the degree of organization inherent in the system.

System elements are the basic constitution units of a system. Each element has attributes reflecting its properties and characteristics and thus determines the heterogeneity of the system. System relationships are the interactions connecting the elements, which can be symbiotic, synergistic, or redundant. The environment refers to any other system or element whose changes in attributes would have an effect on the interested system. A system as a whole is a meaningful family of elements, relationships, and attributes. There is natural purpose and a degree of organization governing the system's existence.

With this framework, a linear complicated system, such as a mechanical clock, may have a large number of elements, but the attributes, relationships and interactions of elements are relatively fixed and unchanging, so that the system is highly structured and tightly coupled, which leads to relatively high levels of stability and predictability, but low levels of adaptability. Such features make it possible to understand, to model, and to reproduce the linear complicated system by decomposing the system to its constituent elements, known as reductionism.

With a complex adaptive system, the system is still complicated, but the system elements have the ability to change their individual attributes and interactions to produce new system configurations and behaviors. It is this ability of adaption that distinguishes a CAS from a linear complicated system.

On the other hand, CASs are not chaotic systems either. Chaotic systems are relatively unstructured and loosely coupled, resulting in outcomes that appear so random and disorganized that it is not possible for the system to adapt.

Instead, CASs are somewhere between linear and chaotic systems, with partially connected agents whose decision making and interactions produce behavior and outcomes that are neither fully controlled nor arbitrary. It produces system behavior that lies between order (no change or periodic change) and chaos (irregular change) and leads to the zone of system adaptability known as the edge of chaos [20].

Rouse (2008) [32] concluded the following characteristics for CASs.

- CASs are nonlinear and dynamic, and are composed of independent agents whose behaviors are based on physical, psychological, or social rules rather than the demands of system dynamics.
- Agents' needs or desires, reflected in their rules, are not homogeneous, their goals and behaviors are likely to conflict. In response to these conflicts or competitions, agents tend to adapt to each other's behaviors.
- Agents are intelligent. As they experiment and gain experience, agents learn and change their behaviors accordingly. Thus overall system behavior inherently changes over time. Adaptation and learning tend to result in self-organization. Behavior patterns emerge rather than being designed into the system. The nature of emergent behaviors may range from valuable innovations to unfortunate accidents.
- There is no single point(s) of control. System behaviors are often unpredictable and uncontrollable, and no one is "in charge." Consequently, the behaviors of complex adaptive systems can usually be more easily influenced than controlled.

From these characteristics, we can see that it is difficult to control or even to predict a CAS since the system keeps redesigning itself. Unlike the common systems studied in Engineering or Physics, a CAS has no single governing equation or rule that controls the whole system. Instead, it has many distributed, interacting parts (agents) which are governed by their own rules. Each of these rules may influence the actions of other agents, and may affect the system outcome. In such a manner, a CAS exhibits an aggregate behavior that can not be simply derived from the actions of the agents. [33]

For CAS, it is often true that the most precise way to predict how the system will behave in the future is to "wait literally for the future to unfold" [13]. Because the behavior of CAS stems from the complex interaction of many loosely coupled variables, the system behaves in a non-linear fashion, which means a given magnitude change in the input to the system is not matched in a linear way to a corresponding change in the output. Therefore, in a non-linear system, large changes in input may lead to small changes in outcome, and small changes in input may lead to large changes in outcome. As a result, the behavior of a complex system can neither be written down in closed form nor be predicted via the formulation of a parametric model, such as a statistical forecasting model.

The intrinsic unpredictability of CAS may cause some ”seemingly wise” decisions to have harmful side effects in practice. Chu et al. (2003) [34] presented an example to illustrate such phenomena. In the example, a new species of fish called Nile perch was introduced into Lake Victoria, which is expected to be more economically profitable to the local people. However, the following unexpected results were observed:

- The local fishermen did not benefit much since they lack the capital and tools for large scale Nile perch-fishing;
- The original fish (the cichlid fish) was quickly eaten up by the new predators (Nile perch), which made the local people lose an important source of daily protein since they could not afford the high price of Nile perch;
- An explosive increase of mosquitoes was found due to the extinction of the cichlid fish, which used to eat the larva of mosquitoes.

As a result, the life quality of the locals has deteriorated instead of improved as expected.

However, although the future behavior of a CAS can't be predicated in an exact manner, it does not imply that the future is random [25]. Although small variations may lead to drastically changes, there still are recognizable behavior patterns exhibited in a CAS. Therefore, our predictive capacity, although limited to the exact prediction at a future point in time, can still benefit from the knowledge of these patterns, which means that we can enhance our control ability to CAS using effective policies or strategies, especially when the system is under some extreme or catastrophic situations.

2.2 OVERVIEW OF EMERGENCY MANAGEMENT

During the past decade, civil conflicts, terrorist attacks, and natural disasters in the world have caused significant loss of life and property. In 2005, the catastrophe caused by Hurricane Katrina in New Orleans impacted all aspects of that city including its assets, population and economy. Of the city's 180,000 structures, 125,000 were flooded; one year later New Orleans population had been reduced by nearly 60% [35]. The disaster influence is profound: some serious issues still remain in the recovery of housing and public healthcare in New Orleans even after four years (in 2009) [36].

Although different emergency events have distinct characteristics in terms of scale, complexity and treatment, all significant emergency events share certain features: they happen suddenly and often unexpectedly and require immediate responses – unlike other common events, emergency response do not allow responders to learn the situation leisurely and take time before making a decision. Besides that, most emergency responses involve many individuals/organizations, without a rational guidance, it is very likely that the whole system would run into chaos. In fact, ineffective management and lack of preparedness are two main reasons for most unsuccessful emergency responses. As a lesson one should never forget, the mismanagement of Katrina responses cost more than $100 billion and over 1,300 lives [37]. How to respond to emergencies appropriately is a major challenge for all emergency managers.

There are considerable efforts made to improve the ability to respond to various types of emergencies. Department of Homeland Security (DHS) [38] lists 15 National Planning Scenarios, which include various types of emergencies/disasters, from potential terrorist attacks to natural catastrophes. They form the basis for coordinated federal planning, training, exercises, and grant investments needed to prepare for emergencies of all types.

In response to these emergencies, a large amount of protocols, standards and policies have been established at different levels. Among them, the *National Incident Management System* (NIMS) [39] provides a systematic, proactive approach framework to guide departments and agencies at all levels to work seamlessly to prevent, protect against, respond to, recover from, and mitigate the effects of incidents. NIMS works hand in hand with the *National Response Framework* (NRF) [40]. NIMS provides the template for the management of incidents, while the NRF provides the structure and mechanisms for national-level policy for incident management.

Based upon the national standards, local governments and agencies establish specific emergency plans for responding to potential local incidents (e.g., Emergency operations plan from Boulder County, Colorado [41]). The general purpose of such plans is to define task assignments and responsibilities for emergency responders in order to best alleviate suffering, save lives and protect property. Checklist, chart and table methods are commonly used in local response plan to assist decision making, guide command flows and regulate appropriate responses.

These response plans provide general instructions, but most of them are not completely prescriptive so that the actual executions are highly dependent on the individual judgments of emergency managers. Although emergency managers are usually experienced personnel with expertise to handle certain types of emergencies, it is still problematic by only relying on subjective intuition and expert judgment of managers. Further, emergency incidents are rare-events, which makes it impractical for emergency managers to master all necessary knowledge to correctly determine the best response strategies, especially when facing peculiarly extreme situations.

It is widely agreed that well-established response policies are indispensable in supporting emergency managers to make timely decisions correctly during the response phase. However, without good understanding of the response system, it is impossible to prepare effective response policies in advance due to the uncertainty of the event – where it might occur; what might be the cause; and what would be the extent of injuries. Due to its expense, it is impossible to perform real-life experiments to verify the effectiveness of a particular policy. Under such circumstances, researchers have developed lots of OR models to study emergency management in a quantitative way.

2.3 OPERATIONS RESEARCH (OR) MODELS FOR EMERGENCY MANAGEMENT

2.3.1 Overview

Wright et al. [11] provide an overview of the use of models in homeland security and classify the models using the four phases of the disaster life cycle: planning, prevention, preparedness and response, combined with the countermeasures and component support portfolios of the U.S. Department of Homeland Security (DHS). The countermeasures portfolios are chemical, biological, radiological, and high explosives. The component support portfolios of DHS are border and transportation security, critical infrastructure protection, cyber security, emergency preparedness and response, and threat analysis. Using their classifications, the work in this dissertation falls into the emergency preparedness and response portfolio and could be used for planning or preparedness purpose by analyzing the effects of resource levels. It could also be used for response by testing a few preselected policy alternatives and the current resource levels to identify likely issues or identify critical resource needs.

2.3.2 Analytical models

Mathematical programming is generally used to find solutions to the optimization problems in emergency management, such as maximal zone coverage or minimizing response time. Toregas et al. (1971) [42], Weaver and Church (1985) [43], and Marianov and Revelle (1994) [44] used set covering models while Schilling et al. (1980) [45] Revelle et al. (1997) [46] and Badri et al. (1998) [47] used goal programming methods. As an early OR models for emergency medical service (EMS) deployment, the hypercube model was first introduced by Larson (1974) [48]. In the Hypercube Model, the whole response system is modeled as an expanded, spatially distributed, multi-server queuing system. The Hypercube Model has been used in other EMS base location studies [49, 50, 51]. In recent years, more analytical models have also been developed for emergency preparedness and response for applications such as vehicle dispatching and routing [52], logistics coordination [53, 54], evacuation planning [55], etc. The advantage of mathematical models is that usually they are relatively lightweight in computational resources consumption and faster to solve, while the disadvantage is that they rely on many assumptions that may oversimplify the system studied, causing the application domain of the model to be tightly constrained and making it unsuitable to model a complex system.

2.3.3 Simulation models

In contrast to analytical models, simulation models can capture behaviors of individual entities, which allows analysts to analyze transient effects such as those occurring during the initial stages of a disaster event. The secion lists a few of simulation models that are relevant to emergency response.

2.3.3.1 Discrete-event simulation (DES)

Goldberg et al. (1990) [56] built a comprehensive DES model to evaluate the response time of the emergency system in Tucson, AZ. The model simulates the response to emergency calls using a multi-server-queuing system. Inside the model, the entire area of interest is divided into zones, and the calls are responded to by the closest idle vehicle on a first-come-first-served basis. The travel time is estimated by the base-zone distance. The model was extensively validated against the actual data and it was found that the zone structure is crucial to build a valid simulation.

Shuman et al. (1992) [57] developed a discrete event simulator (RURALSIM) for designing and evaluating rural EMS systems. RURALSIM could generate multi-type and multi-severity distributed emergency incidents, which are then responded according to a set of pre-defined operational rules. A number of measures of effectiveness output by RURALSIM can provide decision makers more insights into the system evaluation. Several successful implementations of RURALSIM were reported in the states of Maine, Missouri, Oklahoma and Nebraska.

Haghani et al. (2004) [52] presented a simulation model to evaluate a real-time emergency medical service vehicle response system. The model uses real-time travel time information as input and is designed to assist the emergency vehicle dispatchers in assigning response vehicles and guiding those vehicles through non-congested routes. Different response strategies are evaluated with this simulation model.

DES models are also widely used to simulate operations in hospitals. Hirshberg et al. (1999) [58] developed a discrete-event computer model of the emergency room and related hospital facilities to analyze the utilization of surgical staff and facilities during an urban terrorist bombing incident.

Su and Shih (2003) [59] constructed a computer simulation model to evaluate the existing EMS system, tested potential operating policies and suggested improvements for pre-hospital care to decrease casualty mortality and morbidity.

Hung et al. (2007) [60] reported a DES-based patient flow model to test simulated pediatric emergency department staffing scenarios in order to alleviate the pressures that result from increased census and overcrowding. Boginski et al. (2007) [61] introduced a DES model built in Rockwell ARENA, to study the process of patient flow through the hospital system and identify potential sources and locations of delays associated with equipment utilization. Kolker (2008) [62] used discreet event simulation to establish a quantitative relationship between emergency department (ED) performance characteristics and the upper limits of patient length of stay (LOS).

2.3.3.2 Agent-based simulation (ABS) An agent-based simulation model contains a collection of autonomous agents which can perceive their environment, exchange information, make operational decisions, and act based on these decisions [14]. Many ABS models have been developed in emergency management research.

Carley et al. (2003) [63] built a multi-agent simulation model (BioWar) to simulate biological and chemical attacks. BioWar incorporates several sub-models including agent-level disease, diagnosis, treatment, social networks, environmental and attack models. Narzisi et al. (2007) [64] developed PLAN-C to study the performance of populations under catastrophe scenarios due to terrorist attacks. Their research provides particular insight into the dynamics that can emerge in this complex system.

Massaguer et al. (2006) [65] developed DrillSim, a micro-simulation environment for disaster response, in which every agent simulates a different type of real person taking part in the activity. Khalil et al. (2009) [66] compared DrillSim with four other Agent-based crisis response systems (DEFACTO, ALADDIN, RoboCup Rescue, and FireGrid). Their analysis includes architecture and methodology of different systems.

Chen and Zhan (2008) [67] used an agent-based model to simulate the traffic flows and the collective behaviors of response vehicles to investigate the effectiveness of simultaneous and staged evacuation strategies under three different types of road network structures.

Schoenharl et al. (2009) [68] developed an agent-based simulation model using RePast [69] as part of the WIPER system (Wireless Integrated Phone-based Emergency Response). WIPER uses a stream of cellular network activity to detect, classify and predict crisis events. The simulation models human activity, both in movement and cell phone activity, in an attempt to better understand crisis events.

Lee et al. (2010) [70] employed an agent-based simulation model of Allegheny County, Pennsylvania, to explore the effects of various school closure strategies on mitigating influenza epidemics of different reproductive rates.

For hospital simulation, Zhu et al. (2007) [71] proposed R-CAST-MED, an intelligent agent architecture built on Recognition-Primed Decision-making (RPD) and Shared Mental Models (SMMs), to alleviate the issues arising from ineffective information management in emergency medical services. Daknou et al. (2008) [72] studied the application of multi-agent systems for emergency department, and proposed a tool to assist the patient care decision-making process at the emergency department.

In summary, ABS models are suitable for simulating large-scale complex systems since they are sufficiently flexible and extensible, which means different types of agents can be easily added and modified over a wide range of scenarios of varying scope and fidelity. However, they are computationally intensive and require lots of computational time and resources, so it is necessary for analysts to employ efficient methods for designing simulation experiments to run the ABS model in an efficient way.

2.3.4 Discussion

From a systematic view, the emergency response system is a large network of communicating subsystems, with each subsystem adapting its behavior to collaborate with other subsystems in the network. Multiple heterogeneous agents exist in each subsystem, such as emergency medical technicians, police, ambulances, incident command and hospitals. These agents act based on certain rules and interact consciously in nonlinear and dynamic manners. They can collect environmental information, exchange information with each other and adapt their behaviors accordingly. For instance, incident command could stop routing more ambulances to a hospital that has run out of beds as soon as it receives the report of lack of available beds. Multiple decisions and activities involving various actors and organizations take place in parallel. As a result, it is very difficult to understand or control the system.

Comparing these features to the definition and characteristics described in Section 2.1, we can see that an emergency response system is very suitable to be abstracted and studied as a CAS. However, although insights from the CAS can provide increased understanding of emergency response and a helpful formulation for modeling, certain modeling techniques are needed in order to transform such an formulation into tangible and understandable results, particularly from a management perspective. The rationale is that the model should enable managers to test and evaluate different "what-if" scenarios subject to policy changes, so that it helps them compare the effectiveness of different response policies and selecting the proper ones.

The traditional analytical approach of hierarchical decomposition that works for general complicated system (e.g. industrial product design) does not apply for CAS analysis, since decomposition may result in the loss of important information about interactions among the agents of interest [32]. Although a CAS can be reduced to several separate subsystems, we cannot analyze each subsystem independently and then integrate analysis results to understand the system as a whole [73]. Researchers working in this field have argued that a CAS should be modeled and studied by working "bottom up" rather than "top down" [74].

The choice of models is dependent on the nature of the system and critical aspects of interest. In emergency response, responders continuously gather and report information to the incident command (decision makers), and the latter accordingly adjust the action commands to responders as they react to this information. In order to capture these interactions, we use agent-based simulation to model the whole system and to simulate the adaptive behaviors of different agents.

2.4 AGENT-BASED SIMULATION (ABS) MODELING

Agent-based simulation (ABS) modeling is derived partly from distributed artificial intelligence and partly from the science of complexity. According to Luck et al. (2003) [75], agent-based systems has been widely studied in a diverse range, including artificial intelligence, human-computer interaction, distributed and concurrent systems, decision support, information retrieval and management, etc. In ABS modeling, large numbers of actors are simulated as adaptive agents that can adjust their behaviors in response to the changes from environment. Usually, the basic assumptions about the adaption rules are relatively straightforward [24].

2.4.1 Overview of agents

The basic constituents of ABS models are agents, which are defined by North and Macal (2007) [14] as "the decision-making components in complex adaptive systems. Agents have sets of rules or behavior patterns that allow them to take in information, process the inputs, and then effect changes in the outside environment. "

According to Nilsson and Darley (2006) [1], agents distinguish themselves from standard objects in object-oriented programming on the following aspects:

1. Agents embody stronger autonomy than objects; that is, agents are purposeful - "objects do it for free, agents do it for money";
2. Objects are passive while agents are active and have internal mechanism;
3. On the model level, agents are each considered to have their own thread of control whereas in the standard object model, there is a single thread of control.

Macal and North (2010) [76] provide the following characteristics of agents in ABS:

1. An agent is an identifiable, discrete, or modular, individual with a set of characteristics and rules governing its behaviors and decision-making capability.
2. Agents are self-contained. The discreteness requirement implies that an agent has a boundary and one can easily determine whether something is part of an agent, is not part of an agent, or is a shared characteristic.
3. An agent is autonomous and self-directed. An agent can function independently in its environment and in its interactions with other agents for the limited range of situations that are of interest.
4. An agent is social, interacting with other agents.
5. Agents have protocols for interaction with other agents, such as for communication. Agents have the ability to recognize and distinguish the traits of other agents.
6. An agent is situated, living in an external environment with which the agent interacts in addition to other agents.

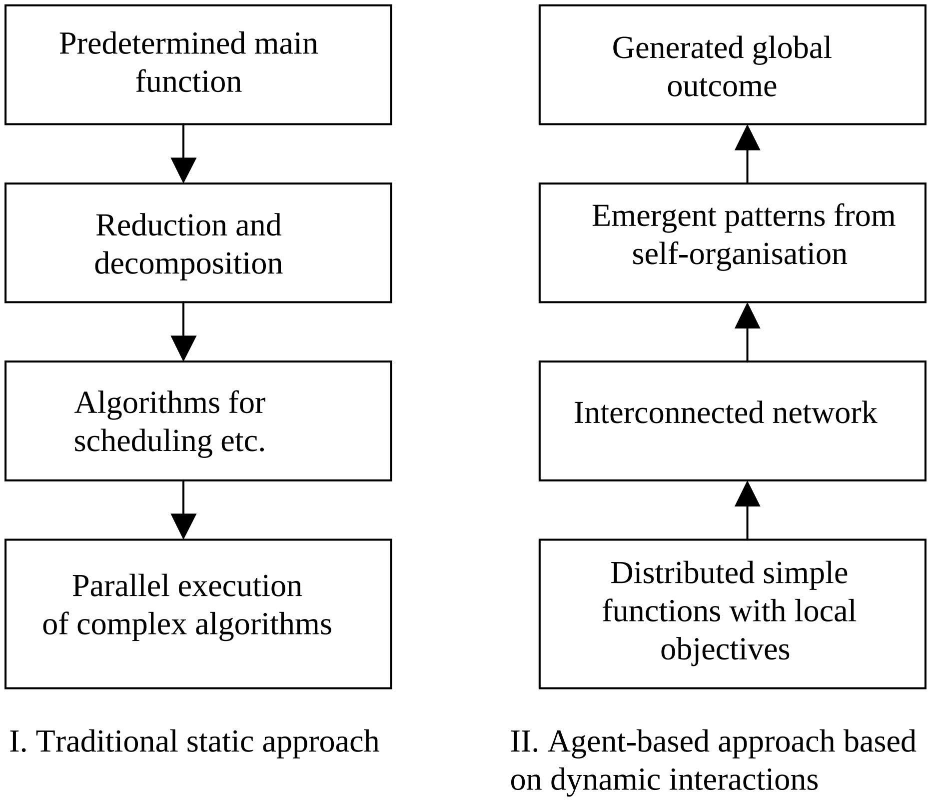
7. An agent may be goal-directed, having goals to achieve (not necessarily objectives to maximize) with respect to its behaviors. This allows an agent to compare the outcome of its behavior to the goals it is trying to achieve.
8. An agent is flexible, having the ability to learn and adapt its behaviors based on experience. This requires some form of memory. An agent may have rules that modify its rules of behavior.

The essential characteristics of an agent include the adaption in environment, autonomy and flexibility [77]. The ability to interact with the environment sets an agent apart from an AI system which has no need of an environment. The capacity for autonomous action enables an agent to have control over its own actions and to function without direct human intervention. An agent achieves flexibility by being responsive to changes in its environment, pro-active in its goal-directed actions, and social in interacting with other agents to reach the pre-defined objective.

Different taxonomy matrices have been introduced by different researchers to classify agents into different categories. One commonly-used classification approach is to describe an agent according to its function (for example a shipping agent or a sales agent). Tu (2008) [77] divided agents broadly according to their architecture with deliberative agents and reactive agents, which are respectively at the stronger and weaker ends of the spectrum of the notion of agency. In this dissertation, we adopt the categorizations presented by North and Macal (2007) [14] by classified simulated agents into two categories: full-functional agents and proto-agents; compared to full-functional agents, proto-agents lack of autonomy and can not make rational decisions by themselves.

2.4.2 Agent-based modeling

In ABS modeling, systems are built from the ground-up in contrast to the top-down manner used in traditional modeling methodologies [78]. According to Nilsson and Darley (2006) [1], the top-down methodologies are based on the assumption that knowledge is outside the "system" and researchers can measure and analyze the observable phenomenon of interest by decomposing the whole system to different sub-units and solving the sub-problems separately. On the contrary, bottom-up methodologies assume that modelers cannot understand the whole phenomenon of interest but they can observe and understand specific activities and processes of individuals on a micro level, and synthesis the whole system by modeling the behaviors and interactions of these individuals. Figure 2.2 compares the difference of these two methodologies.



Source: Derived and modified from Reaidy *et al*. (2003 p.151)

Figure 2.2: Top-down and Bottom-up methodologies (Nilsson and Darley, 2006) [1]

A major characteristic of top-down approaches is the assumption that the future selections can be defined as conditional probability mass functions of past selections. Such dependence complicates the decision making problem, since the construction of probability mass functions quickly becomes an obstacle as the number of possible scenarios combinatorially explodes. This either requires increasing amounts of data to reliably estimate these parameters or forces parameter estimation to rely upon subjective impressions.

As an alternative, the bottom-up modeling approach allows direct imitation of behaviors which may be difficult to replicate solely through probability mass functions over the range of aggregate outcomes. The bottom-up approach provides connections that link the behavior of the individual components to the resulting system effects. The agent-based model thus allows researchers to convert their understanding of individual behaviors or experience with detailed processes into the knowledge about the complete system. As we have discussed in Section 2.1, a CAS can be considered as a multi-agent system that evolves over time and space. Jennings (2000) [5] presented a canonical view to illustrate how such a multi-agent system can be organized at its simplest level, as shown in Figure 2.3.

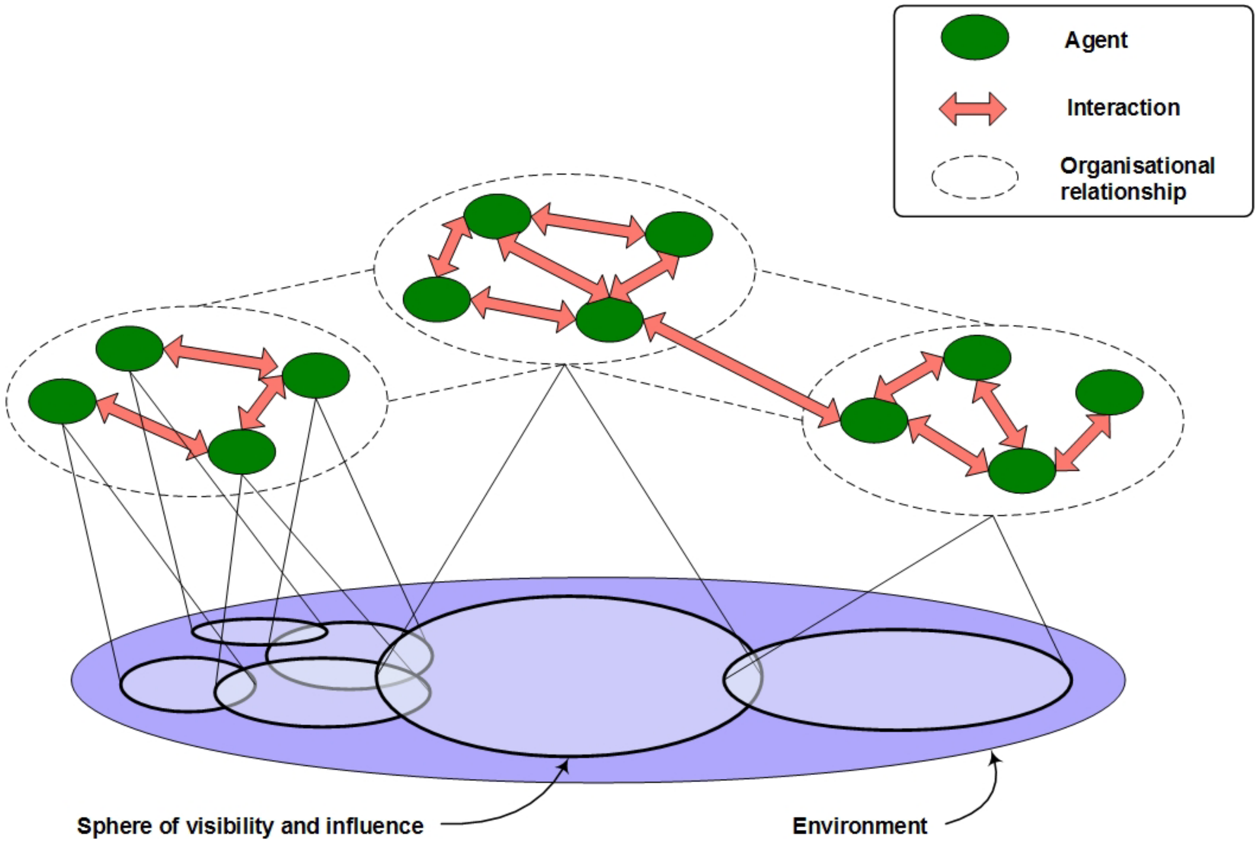


Figure 2.3: Canonical view of an agent system (Jennings, 2000) [5]

From Figure 2.3, we can see that the agents are highly coherent modules and a number of them with related functions may be grouped together in a loose cluster with each agent limited in its view and influence within its activity domain. There is no identifiable central control of the group since this function is distributed among the agents, and embedded within each is its individual limited set of control rules. Their network topology is usually pre-determined and they communicate their requests and intentions with each other by message exchanging. Because agents communicate in this manner, they are more naturally suited to distributed simulation.

According to Luck et al. (2005) [79], agent-based simulation modeling has achieved a comparatively wide degree of acceptance and has been successfully implemented in many real-world systems. In addition to the models developed for emergency research (in Section 2.3.3.2), the examples of ABS application also include epidemic and pandemic prevention [17, 18], disease propagation [80], human movement in a theme park [81], manufacturing shop floor control [82], urban planning [83], water usage policy management [84], network security [85], and product/system design [77, 86, 87, 88, 89]. Readers are also referred to Macal and North (2010) [76] for a more comprehensive review of ABS applications.

Due to substantial public research and development investments, many ABS modeling software environments are now freely available [90]. These include Repast, Swarm, NetLogo and MASON among many others. Proprietary toolkits are also available such as AnyLogic. A detail review and recommendations of ABS development platforms is provided by Railsback et al. (2006) [91]. A recent survey and comparison of agent-based modeling and simulation tools can be found in Allan (2009) [92].

Nilsson and Darley (2006) [1] and Bonabeau (2002) [2] concluded the advantages and disadvantages of ABS modeling, their results are summaried in Table 2.1.

Table 2.1: Pros and Cons of ABS [1, 2]

Advantages	**Disadvantages**
• Provides a natural description of a complex adaptive system; • Increases realism; • Includes heterogeneity and bounded rationality; • Promotes scalability and flexibility;	• High development costs in both time and effort; • Requires more data to be collected than many other approaches; • Computationally intensive;

The major disadvantage that impedes the application of ABS is its low computational efficiency. Furthermore, simulation models must be run for a certain number of replications in order to obtain statistical meaningful conclusions, which consequently aggravates the low-efficiency problem even more. Therefore, it is necessary to find proper statistical-based techniques for efficient simulation control and economic output analysis.

2.5 OPTIMIZATION-VIA-SIMULATION METHODS

Well-developed ABS models only provide the prerequisite for the purposes of further system analysis. In order to obtain correct conclusions, proper optimization-via-simulation (OvS) techniques must be used to control the simulation and analyze the results. The benefits of OvS include:

- Best utilize computational resources for computationally intensive simulation (e.g., ABS);
- Improve efficiency by screening out non-competitive systems in the early stages;
- Identify the "best" design or policy efficiently with given confidence level;
- Gain some insights about applying adaptive control technique to simulation of a large-scale complex adaptive system [93].

Multiple approaches have been applied to different simulation models to address various optimization problems, including genetic algorithms [94], simulated annealing [95], maximum likelihood estimation based methods (e.g., bootstrap methods) [96, 97], tabu search [98], threshold accepting search methods [99, 100], and ant colony optimization [101, 102] among others. The goal of our study is to choose a "good set" of systems from a number of competing alternatives, where the "best" refers to the system with the largest or smallest expected performance measure. This can be accomplished by comparing output from different alternative systems using the appropriate statistical methods.

Kim and Nelson (2007) [103] classify comparison problems arising in simulation studies into four classes: (1) selecting the system with the smallest (or largest) performance measure (selection of the best), (2) comparing all alternatives against a standard (comparison with a standard), (3) selecting the system with the largest probability of actually being the best (multinomial selection), and (4) selecting the system with the largest probability of success (Bernoulli selection). The objective of our simulation is to identify the best response policy(s) that lead to the least amount of mortality, and there are two major categories of OvS methods that could be used to realize the objective, which are ranking-and-selection (R&S) and multiple comparison procedures (MCP).

2.5.1 Ranking-and-selection (R&S)

R&S procedures are specifically developed to select the best population or a subset that contains the best from competing alternatives [104]. Over the last decade, there have been fruitful efforts in developing statistically valid R&S procedures. In general, these procedures can be classified into two large categories: Bayesian procedures and Frequentist procedures.

Bayesian procedures try to maximize the posterior probability of correct selection (Chen et al. 2000) [105] or try to minimize the opportunity cost given a simulation budget (Chick and Inoue 2001) [106], and are usually more efficient than Frequentist methods. However, Bayesian procedures cannot provide a statistical guarantee of correct-selection [107], which is their major disadvantage.

Compared with Bayesian-based approaches, Frequentist procedures, such as Rinott (1978) [108] and Kim and Nelson (2001) [109], are relatively conservative, since they allocate simulation effort to different systems to ensure a probability of correct selection even for the least favorable configuration. But they can provide statistical guarantees of correct-selection, which is preferred in many application cases. For this reason, the focus of this dissertation is on Frequentist procedures, and the procedure proposed here does guarantee a pre-specified level of correct-selection probability.

Frequentist procedures may be single or multi-staged. Based on their objectives, there are two formulations of the problem of comparing alternative systems, which are indifference-zone formulations and subset selection formulations respectively [110].

Indifference-zone formulations provide a guarantee of selecting the single best system, where an indifference-zone parameter δ is defined at the range where the experimenter is "indifferent" to alternatives within δ of the best system. Subset-selection formulations choose a subset of the available alternatives so that there is a defined probability guaranteeing that the subset includes the best system.

2.5.1.1 Indifference-zone selection (IZS) A large set of ranking-and-selection procedures are based on the indifference-zone formulation. These approaches are characterized by two parameters, $\{\delta, P^*\}$, where δ is known as indifference-zone, which indicates a region in which the experimenter would not discriminate among competing systems. The value P^* denotes the threshold of desired probability of correctly selecting the best alternative $P\{CS\}$; it is expected that $P\{CS\} \geq P^*$.

The original indifference-zone R&S procedure was proposed by Bechhofer (1954) [111] as a single stage procedure. From a given $\{\delta, P^*\}$, the procedure can determine the number of required observations for each competing system. A major disadvantage of Bechhofer's procedure is its assumption for common, known variance across all systems, which may not be justified in a given simulation.

To address this issue, Dudewicz and Dalal (1975) [112] presented a two-stage procedure (D-D), in which variances are estimated at the end of first stage and are used to calculate the number of observations required at the second stage. A weighted average of the first and second stage sample means is then used to select the best system. Rinott (1978) [108] modified the D-D procedure to the R procedure, which yields a greater $P\{CS\}$ in some cases, but may require a larger total number of observations. For this reason, it is not appropriate to use the R procedure when the number of competing systems is large, especially when run time is an issue.

In order to handle cases involving a large number of alternatives, Nelson et al. (2001) [113] presented the NSGS (Nelson-Swann-Goldsman-Song) procedure, which uses the data from the first stage sampling to screen out alternatives that are not competitive, and thereby avoid the (typically much larger) second-stage sample for these systems.

The two-stage procedures with screening can be extended to more than two stages or to sequential-stage procedures, where a screening procedure is applied at each stage until only the best alternative is left, such as the fully-sequential procedures KN (Kim and Nelson 2001) [109] and KN+/KN++ (Kim and Nelson 2006) [114]. These procedures are effective in eliminating inferior systems and thus more efficient than the R procedure.

In recent years, researchers put their attentions on improving the applicability and efficiency of indifference-zone selection procedures. Hong and Nelson (2005) [115] proposed sequential procedures (HN) that attempt to balance the cost of sampling and switching to minimize the total computational cost. Hong and Nelson (2007) [107] also presented procedures that are capable of selecting the best alternative in the situations when the alternatives are revealed (generated) sequentially during the experiment. Osogami (2009) [116] proposed a two-stage indifference-zone approach (TSSD) with the goal of reducing both the number of simulated samples of the performance and the frequency of configuration changes. Tsai and Nelson (2010) [117] applied the Control Variates (CV) technique in fully-sequential indifference-zone selections to develop a more efficient R&S procedure.

As a common characteristic, almost all existing procedures (KN/KN+/KN++, HN, TSSD) are designed to select only a single system (which will then be claimed as the best) whose mean performance measure (μ_b) is within an indifference-zone (δ) to the true-best system's mean (μ_B, unknown). All others will be screened out in the early stage or disregarded in the final stage (due to exceeding the computation budget limitation).

2.5.1.2 Subset selection (SS) The other major type of R&S is subset selection, which is first presented by Gupta (1965) [118]. The goal of the Subset Selection procedure is to identified a subset of random size that contains the best system, with user-specified probability P^* and without the specification of an indifference-zone (i.e., $\delta = 0$).

Like Bechhofer's indifference-zone procedures, Gupta's subset selection procedure requires equal and known variances among competing alternatives. To solve this problem, Sullivan and Wilson (1989) [119] develop two subset selection procedures that extend Gupta's work by allowing unknown and unequal variance, and specification of a non-zero indifference-zone.

Many subset selection approaches are designed to select a restricted subset, where the term "restricted" implies that extra input parameters are needed to restrict the selection set. For examples, Koening and Law (1985) [120] developed a two-stage indifference-zone procedure to select a subset of size m containing the v best of k systems; where ($1 \leq v \leq m < k$). If $m = v = 1$, then the problem is to choose the best system. When $m > v = 1$, they are interested in choosing a subset of size m containing the best. If $m = v > 1$, they are interested in choosing the m best systems. Chen (2009) [121] proposed a heuristic two-stage selection procedure (Enhanced Two-Stage Selection procedure) to select a subset of size m containing at least c of the v best of k normal populations with unknown means and unknown variances. They also derived the probability of correct selection lower bound of determining a subset based on the distribution of order statistics.

Compared with indifference-zone based approaches, subset selection procedures are less popular in practice. The early subset selection procedures cannot provide guarantee on the performance of all alternative systems among the selected subset – they only claim that their selection subset contains the best but do not claim that the whole subset selected is “good”. And the restricted subset selection methods require too many input parameters which are usually difficult to justify in applications. These deficiencies make them less be used in practice.

2.5.2 Multiple comparison procedures (MCP)

Unlike R&S procedures, whose objective is only to find the optimal alternative(s), MCP provide not only inference about the best system, but also relationships among all the systems. According to Swisher (2003) [122], MCP can be classified into three general categories: all-pairwise comparisons approaches, multiple comparisons with a control (MCC), and multiple comparisons with the best (MCB).

2.5.2.1 All-pairwise comparison approaches Two sub-categories can be made to classify the all-pairwise comparison approaches: (a) combined paired-t, Bonferroni, and all-pairwise comparisons; (b) all pairwise multiple comparisons (MCA).

The first category is referred as the brute force approach by Fu (1994) [123], since it examines all possible pairwise for k systems, resulting in a total $k(k-1)/2$ of confidence intervals. Due to the Bonferroni inequality, each confidence interval must be constructed at level $\{1-\alpha/[k(k-1)/2]\}$ in order to have a joint confidence level of at least $(1-\alpha)$, which causes extremely wide individual confidence intervals for a large number of alternatives, and consequently, little inference can be obtained from it.

Unlike brute force approaches, MCA (Tukey 1953) [124] obtains an overall simultaneous confidence level $(1-\alpha)$ with shorter confidence half-widths for all $k(k-1)/2$ pairwise comparison, thus it is better for comparison. However, compared with other MCP methods, all-pairwise comparisons usually need the most observations

2.5.2.2 Multiple comparisons with a control (MCC) Sometimes the goal is to compare a set of alternatives to a pre-defined control (e.g., current existing design). Dunnett (1955) [125] proposed the first MCC procedure to construct $(k-1)$ simultaneous confidence intervals in comparison to a fixed control. The traditional MCC is then expanded to a two-stage MCC procedures (Bofinger and Lewis, 1992 [126]), and allows different systems having different probability distributions to be compared against a single (standard) design (Damerdji and Nakayama, 1996 [127]). MCC is usually efficient since it takes the least number of observations.

2.5.2.3 Multiple comparisons with the best (MCB) MCB is used to select the best system and identify those significantly worse than the best. Since the best system is unknown before, the number of observations needed for MCB is usually larger than MCC.

The first MCB procedures were developed by Hsu (1984) [128]. Yang & Nelson (1991) [129] and Nelson & Hsu (1993) [130] describe modifications to the MCB procedure that incorporate two variance reduction techniques (control variates and CRN) to shorten the length of the confidence intervals for a specified level of confidence. Goldsman and Nelson (1990) [131] outline an MCB procedure for steady-state simulation experiments. They also discuss results on how the batch size can impact the probability of correct selection when using the simulation technique of batch means. Nelson and Banerjee (2001) [132] present a two-stage MCB procedure that simultaneously achieves several objectives for a given probability of correct selection.

An important characteristic of MCB is it can be combined with R&S procedures, compared with the individual approach, the combined R&S-MCB procedures not only select the best system with pre-specified confidence but also provide insight about how much better the best alternative is in comparison to the rest of the alternatives, with little or no additional computational overhead.

Gupta and Hsu (1984) [133] first proposed a unified methodology for simultaneously executing R&S and MCB. Nelson and Matejcik (1995) [134] show that most indifference-zone procedures can simultaneously provide MCB confidence intervals with the width of the intervals corresponding to the indifference-zone. They also derive a two stage combined procedure - Procedure NM. Swisher and Jacobson (2002) [135] apply procedure NM to determine the optimal clinic design from seventeen competing alternatives.

2.5.3 Summary

In summary, R&S and MCP are both effective tools for selection of the best alternative(s). R&S approaches allow the simulation analyst to choose the best design at or above a user-specified probability level within an indifference-zone, or to screen alternatives to a smaller subset. MCP provide inference about the relationships among competing alternatives. Both procedures are easily adaptable and statistically valid. R&S and MCP are applicable to comparisons among a finite and typically small number of systems (often less than 100) [136]. When the number of alternatives becomes large, other methods should be considered. For example, if the factors of the studied systems can be parameterized, response surface methods (RSM [137]) could be utilized to find an optimal solution.

2.5.4 Discussion

In our research, it is desired to have a method that examines alternatives and provides them with a subset of alternatives that are close to the best, so that they can choose the final decision from the "best-subset", instead of unconditionally trusting the best solution provided by a computer. Such a requirement is not only due to the fact that people are unwilling to leave their decision making responsibilities to computers, but also has its practical reason - it is usually neither possible and nor necessary to include all system parameters in a computer model. Some constraints on the system may not be quantifiable for inclusion in a mathematical model, but must be considered in practice. Indeed, the simulation model is only an abstraction of the real system, but not a complete representation. Consequently, the best system selected by computer for the abstraction may not be best for the real system, and could even be infeasible or simply unrealistic in practice. Hence, we suggest that a best-subset involving multiple potential alternatives should be more useful than the only one choice, as it allows the decision maker to choose among a set of alternatives based on criteria not in the model, such as social or political feasibility.

Our research aims to solve a practical policy selection problem for a local government agency. We have developed an agent-based simulation model of emergency response and used the model to examine a set of alternative emergency response policies. The ideal deliverable for the customer is a methodology that can select a subset of alternative policies that all have demonstrated good performance in the simulation model, so the customer can choose among them using other factors to select the policy to implement in practice.

However, our review of the literature suggests that there has not yet a method for selecting a best-subset with statistical guarantee (given probability level) from a set of alternatives. Indifference-zone methods select only a single best alternative, which is not acceptable (because of skepticism of the output of computer models) and subset selection methods are not useful as well (because of the lack of guarantees on the subset or the impractical input parameter requirements). Therefore, we have developed the methodology described in Chapter 5 to select the best-subset from a finite number of competing alternatives while guaranteeing a pre-specified correct-selection probability level.

For relevant research, Kim (2005) [138] developed a fully sequential procedures for comparison systems with a standard. Andradóttir, Goldsman, and Kim (2005) [139], Andradóttir and Kim Kim (2010) [140], Batur and Kim (2005) [141] and Batur and Kim (2010) [142] considered the problem of finding a set of feasible or near-feasible systems among a finite number of simulated systems in the presence of stochastic constraints. However, none of these research efforts addresses the best-subset selection problem. Because that the actual best system is unknown, it is difficult to recognize and eliminate the inferior systems during the screening stage, as well to establish appropriate stopping criteria.

This explains the motivation of the development of a fully sequential R&S procedure to select the best-subset while satisfying the pre-specified correct-selection probability requirement. According to Osogami (2009) [116], the KN series are the most efficiient algorithms in terms of the number of samples needed. So we extend KN in order to develop this new procedure that selects the best-subset by efficiently screening out obviously inferior alternatives in the early stages.

3.0 AGENT-BASED SIMULATION MODELING FOR MASS CASUALTY INCIDENT RESPONSE

In this chapter, we represent a specific complex system – mass casualty incident response system – as an agent-based simulation model. We hope that the methodology used in this chapter can serve as a guide for future researchers to model and analyze complex systems.

3.1 OVERVIEW

3.1.1 General operations of MCI response

Mass casualty incidents (MCIs) refer to those large-scale disasters involving relatively large numbers of victims (affected people) with injuries at different severity levels. In a MCI response system, when an incident occurs and is reported, the incident command will assess the situation and dispatch responders to the disaster scene to perform triage, stabilization and evacuation.

Triage is a technical term used widely in the emergency medical literature and practice. It is defined as the process of assessing a group of patients' situations and assigning appropriate medical resources for treatment [143], which is usually performed by the first arriving emergency medical technicians (EMTs). On-site triage is recommended or required in most mass-casualty situations in order to avoid resource waste and manage limited assets better, especially for large-scale incidents where medical resources are usually tight [144, 145]. The first step of triage is to screen and classify injured victims into several categories based on their severity levels [146, 147, 148]. A popular triage coding system for trauma events [146, 143, 3, 148] is presented as follows:

- “Black” or expectant – Non-salvageable/dead on arrival (DOA): Victims who are found to be clearly deceased at the scene with no vital signs and/or obviously fatal injuries.
- “Red” or immediate – Life-threatening injury: Victims who have life-threatening injuries or illness but salvageable (such as head injuries, severe burns, severe bleeding, heart-attack, breathing-impaired, internal injuries). They have the first priority for treatment and transportation.
- “Yellow” or delayed – Severe injury. Victims who have potentially serious but not immediately life-threatening injuries (such as fractures).
- “Green” or minimal – Walking/moderate wounded. Victims who are not seriously injured, quickly triaged, and escorted to a staging area out of the scene for further evaluation and transportation.

As the next step, the on-site emergency medical services (EMS) personnel assess the patients' situation and determine the appropriate actions to take. In severe situations, the EMS responders treat and stabilize the patients and then evacuate them to appropriate medical facilities (hospitals). In less critical situations, the EMS may just treat the patients at the scene and leave them for further medical care to be delivered by other support responders.

Evacuation is usually performed by ambulances traveling from their bases to the scene. When an ambulance arrives, the EMS will load the most critical patients and transport them to an appropriate hospital for more definitive treatment. An evacuation ambulance may travel back and forth between the scene and various hospitals multiple times, depending on management’s decisions.

The above EMS operations are a generic, fundamental response plan, which is extracted from the federal, state and local standards (e.g., NFPA 1561 [149], Boulder County Medical Emergency Response Plan [41]) and are being executed nationwide. Although variations may be made in the details (rules) for treatment or transportation of casualties to fit the special needs in certain situations, the basic response principle is to stabilize the casualties at the scene and then transport them to medical facilities as soon as possible according to their severity priorities.

Besides medical responders, other possible responders might include firefighters, police and hazmat (hazardous materials) teams. They are usually assigned to perform certain specified tasks. For example, firefighters are trained for basic life support and can be the first responders to the scene and work as emergency medical technicians to stabilize victims at the scene; hazmat teams might be needed at the scene to deal with the contaminated materials first before other responders can enter the scene.

3.1.2 Discussion

In an emergency response, responders begin with limited information about the incident and make decisions based on information they gather themselves or through communicating with other responders. Based on protocol and decisions made by incident command, responders operate under a set of rules that may change as incident command gets new information.

For example, in a mass casualty incident, there is an initial call to an emergency number that notifies responders that an incident has occurred. The first units on the scene then provide situational awareness and begin triaging patients. As additional responders arrive casualties are triaged, information is collected and reported to incident command, and patients are evacuated to the appropriate hospitals. As information is reported and the scope of the incident becomes more apparent, incident command adapt the response to the size and type of incident based on the resources available.

The information gathering and processing influence the incident response directly. The concrete action steps are dependent on the information gathered during the response, and responders have to make decisions with incomplete information in a distributed fashion. So agent-based models are especially relevant to modeling emergency response to mass casualty incidents since they provide a nature way to describe the information collection and interpretation for various actors. In this chapter, we build an agent-based model to simulate a mass casualty incident response in an urban area. The model is used to examine the effects of different evacuation policies to the response.

3.1.3 Simulation platform selection

Railsback et al. (2006) [91] compared different agent-based simulation toolkits, including Repast [150], NetLogo [151], MASON [152], and Swarm [153]. According to their review, Repast is the most complete Java platform. Compared to the other platforms, Repast has good execution speed and many other desired capabilities, such as the ability to reset and restart models from the graphical interface, the "Multi-run" experiment manager, and built-in geographical and network supporting functions. Due to these benefits, we chose Repast as our modeling platform.

Repast stands for "Recursive Porous Agent Simulation Toolkit", which is an open-source, cross-platform, agent-based modeling and simulation toolkit that was originally developed by researchers at the University of Chicago. Some attractive features of Repast include:

- Full object-orientation;
- Flexible hierarchically nested definition of space and visualization of 2D, 3D environments;
- Supports 2D and 3D Geographical Information Systems (GIS): ESRI ArcGIS or OpenMap;
- Provides convenient interface to connect with external optimization tools ;
- Available on virtually all computing platforms including Windows, Mac OS, and Linux;
- Good tutorial and documentation [154]; many publications about successful application experiences [155, 156, 150, 69].

3.1.4 Highlights of the modeling

3.1.4.1 Generic agent types North and Macal (2007) [14] classified commonly used agents into two general categories: full-agent and proto-agent. Compared to full-agents, proto-agents are much simpler in both concept and implementation. Proto-agents cannot make any reasoning-based decisions, but just act following given rules or commands; full-agents have the capability to perceive the environment, collect / analyze / exchange information, and make decisions based on the information obtained. In short, full-agents are more intelligent than proto-agents.

Based on this taxonomy, we decided to use proto-agents to model those non-decision-making participants, such as injured casualties and ambulances, and employing full-agents to simulate the decision maker – incident command. Besides that, in order to achieve better extensibility and code reusability, we extended the agent definition architecture by deriving three generic agent classes: Indicator, Performer, and Commander, as shown in Figure 3.1.

From Figure 3.1 we can see that the biggest difference among these three derived classes is that Commander is derived from full-agent so it has the ability to make decisions, while Indicator and Performer can not since they are derived from Proto-agents.

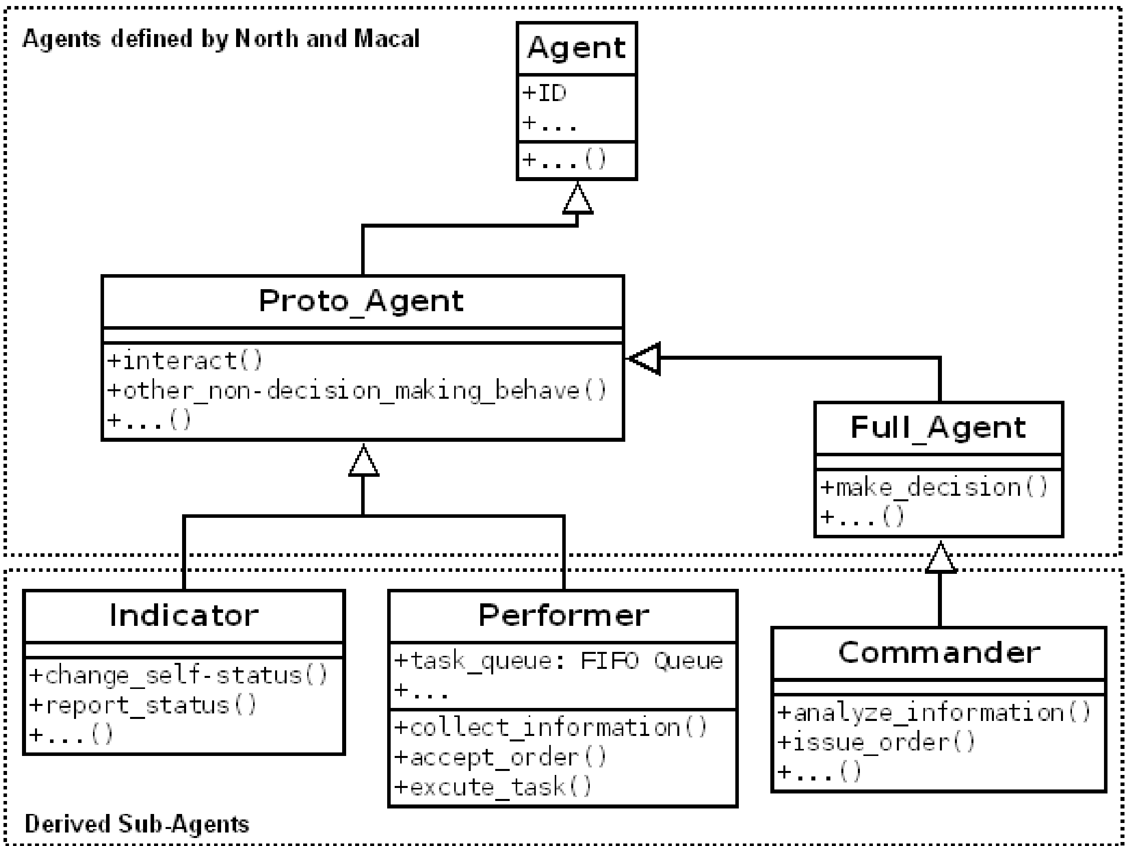


Figure 3.1: Class diagram of agents

The common characteristics shared by all three types are:

- All agents possess attributes and rule-based behaviors, the behaviors include self-action and interaction with others;
- Behavior will change the attributes, which could be either of itself or others. For example, the deterioration behavior of injured casualties (self-action) could decrease their survival probability (self-attribute); the casualty-pickup behavior of an ambulance (interaction) will change the attributes of both agents (the number of passengers on the ambulance, evacuated status for the casualty).

For differences, Indicator agents are entities that do not move through the system by themselves, but can only be moved by other agents. Their state can change in accordance to specified rules and their state can be queried by other agents. Furthermore, an Indicator agent can neither collect outside information nor make decisions, the only information it can provide is about itself. In our model, the injured casualties are modeled as Indicators.

Performer agents can execute tasks which are either generated according to internal rules or assigned by Commanders. Each Performer agent owns a unique task queue, and it will execute received tasks in a first-in-first-out (FIFO) order. They also maintain a state that can be queried by a Commander agent. The triage EMTs and ambulances are modeled as Performers in current model, but they could be modeled as full agents if being endowed decision making autonomy.

Both Performer and Indicator are proto-agents since they are deficient in their decision making capability. Unlike them, as a derivative of full-agent, Commander agents can collect information and make decisions based on the information in combination with certain rules, so that they have autonomic adaptivity to system changes. In our simulation, incident command and hospitals are modeled as Commanders.

The definition of generic agent classes makes it easier to add new agent instances. For example, suppose that we want to add fire trucks as another type of evacuation vehicle besides ambulances, we can simply derive a new sub-class from the Performer – Responder class. Most attributes (e.g., capacity, speed, etc.) and action methods (pickupCasualtes(), transportToHospital()) can be inherited from the parent class, the only modifications that we need to do are specifying some feature parameters. So it would be quite easy to introduce new agents to expand the model.

3.1.4.2 GIS integration In order to simulate the impacts due to traffic during the evacuation, we model the transportation network of the city or area under incident using the data extracted from geographical information system (GIS) shapefiles. In our model, the transportation network is modeled as a set of nodes and arcs, where the nodes are used to represent specific locations, such as the incident site (we assume that the incident site is relatively compact so that it can be modeled as a single node), street cross-sections, medical facilities locations, ambulance bases, etc., and the arcs are used to represent the connecting streets or roads between two locations.

We developed a set of methods [19] to analyze GIS road shapefiles, extract network data and simplify the network. The simplification is necessary since not all nodes are needed to be included in the model. Strategically, finer grids are modeled for the more interesting areas (e.g., street blocks around the incident scene) while cruder grids are built for other less interesting locations. Such an implementation enables a better granularity control to the transportation network simulation.

In the simulation, the evacuation vehicles move along the arcs (roads) through the transportation network, and the ongoing status of evacuation (such as the location of each ambulance, distribution of evacuated casualties in different hospitals, etc.) can be displayed on a GIS map view while the simulation is running, which provides a direct picture to the emergency managers about the evacuation process and can help them identify potential problems.

With the developed tools, our simulation model is no longer location dependent and can be used to model any region (a city or a county) by simply replacing the source GIS data, which provides great flexibility to the model to simulate any urban area wherever GIS data are available. Although the most simulation experiments presented in this dissertation use Pittsburgh (PA, US) as the scenario, we have also applied this methodology to seven other cities in Pennsylvania (US) to simulate the responses to incidents occurred at different places of each city.

3.1.5 System structure of MCI response

Our model simulates the emergency medical response to a mass casualty incident in an urban area. Through the analysis in the Section 3.1.1, we determined the following agents will be simulated in our model:

- Injured casualties;
- Emergency medical technicians (EMTs): perform on-site triage;
- Ambulances: evacuate casualties to hospitals;
- Hospitals: receive casualties and provide definitive care to them;
- Incident command: collect information, make casualty dispatching decisions according to current effective evacuation policy combined with the information reported by others.

The entire response system is decomposed into three sub-models, the incident site, pre-hospital (evacuation), and in-hospital processing, as shown in Figure 3.2.

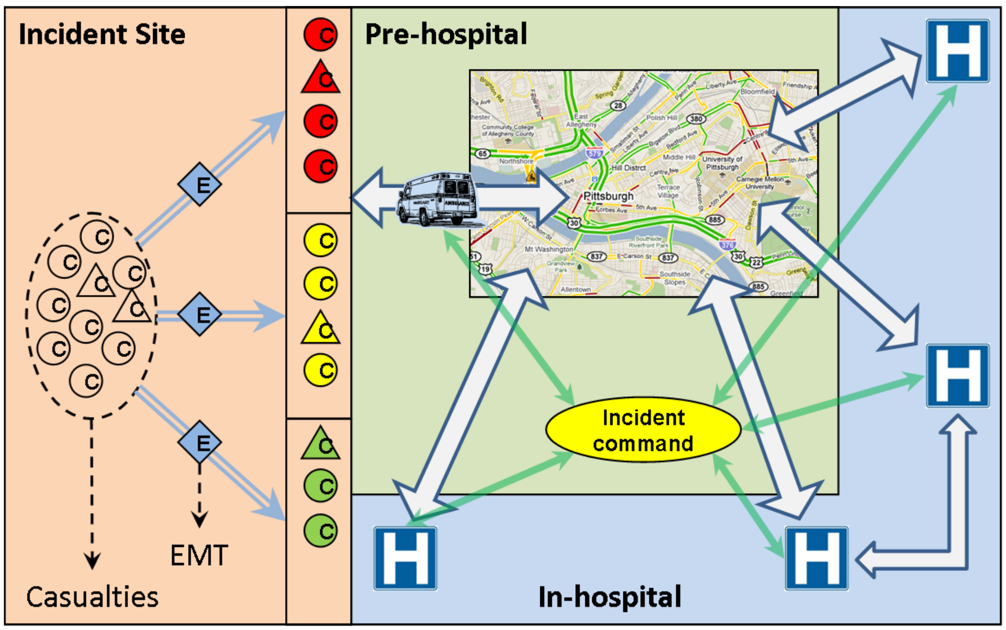


Figure 3.2: Structure diagram of MCI response system

In Figure 3.2, the circles and triangles with a letter "c" inside represent injured casualties, where the triangles correspond to the specialized type of casualties and the circles stand for general casualties. The diamond shapes with a letter "E" inside represent on-site EMTs, who perform on-site triage and classify the triaged casualties into different groups (red, yellow and green), where different colors indicate different injury severities and evacuation priorities.

The Ambulances in the pre-hospital sub-system travel between the incident site and hospitals, which are indicated by the boxs with a capital "H" inside, to evacuate the classified casualties according to their priorities and the orders from the incident command. The incident command collects information from ambulances and hospitals and select a destination hospital for a loaded ambulance upon receiving its request.

3.1.6 Incident site

The major on-site responses include on-scene triage and stabilization performed by EMTs. The triage results determines the priorities of victims in evacuation. During triage, casualties are identified as being red, yellow, or green (or black, which means died and is no longer part of the model), while the red has the highest evacuation priority. In addition, the triage determines whether or not a patient requires special treatments at a specialized hospital. Other agents are only aware of the triage designation instead of practical status of casualties, which allows us to simulate triage errors.

3.1.6.1 On-site emergency medical technician (EMT) On-site EMTs are modeled as Performer agents, which are used to simulate the first arrived emergency medical technicians who perform on-scene triage and stabilization. They classify the casualty into different groups by on their types ("general" or "specialized") (see Section 3.1.6.2), and assigning a color designation (red/yellow/green) to indicate the injury severity of each casualty based on the triage result. As ambulances arrive, triaged patients are loaded according to specified policy (see Section 3.1.7.1). Both triage and patient loading take certain amounts of time that are assumed following Gamma distribution.

3.1.6.2 Casualties Casualties refer to the victims involved in the incident, who are modeled using Indicator agents. For one specific incident, it is possible to observe multiple types of casualties with various injuries at different severity levels. For example, the possible injuries suffered in a bomb blast include blunt, blast, and burnt trauma. And the casualties who were closer to the blast are usually injured more seriously than those further away. In order to represent these differences, two attributes are defined – " casualty type" and "survival probability", where the "casualty type" is used to differentiate different types of patients; and the "survival probability" is used to specify to survival possibility of a casualty in an quantitative way.

In our model, we instantiate two basic casualty types, which are "general" and "specialized". The "general" type is assigned to casualties who are adult (indiscriminate gender), suffering common injuries so that they can be admitted and treated by any hospital. On the contrary, the "specialized" type is used to mark those casualties who have specific requirements in treatments, including but not limited to specialized injury type (burnt, head injury, etc.). For example, child/infant casualties could also be marked as "specialized" since they typically have to be sent to a definitive children's hospital for treatments.

Although very simple, those two abstracted types provide an effective way for us to depict the basic characteristics of different casualties. To anyone who wants to model casualties in a more practical manner, it is quite easy to derive new concrete sub-types based on those two basic types. For example, we could derive a new sub-type called "male-infant-with-head-injury" from the "specialized" type.

Another attribute – " survival probability" is designed to quantify the survival possibility of a casualty by a positive real value within the range of $[0,1]$. A larger value of survival probability represents a good condition (usually observed from a mild-injured patient) while a smaller value corresponds to a bad condition of a patient suffering severe or life threatening injury.

3.1.6.3 Casualty degradation Before definitive care or treatments were received, the health condition of a casualty would deteriorate continuously (especially for those injured seriously), this is called *Casualty Degradation*. The consequence of degradation is the decline of survival probability of the casualty. In this research we studied the casualty degradation by two different models, one is the proportional-hazards based model [157], and the other is Sacco's RPM-based model [158, 143, 3].

The proportional-hazards based degradation model is proposed by Wu [15]. In this model, each injured casualty is assumed to have an initial survival probability (P_0) , which will continuously decrease until definitive care is received or its value reaches zero (which indicates the death of casualty).

The deterioration rate of the survival probability is given by

$$R(t) = g^{-t} \tag{3.1}$$

where g is a real constant greater than one which captures the deterioration characteristic of victims; t is the elapsed time. So the survival probability can be expressed by a monotonically decreasing function of time, as shown in the formula below.

$$P_t = P_0 \times R(t) = P_0 \cdot g^{-t} \tag{3.2}$$

A key parameter in Sacco's RPM-based degradation model is the RPM score, which is designed to measure the victim's injury severity. According to Sacco et al. [143, 3], RPM also provides a good predictor of survival probability that can be easily obtained during the triage. The RPM score takes on integer values from 0 to 12, which is the sum of coded values for respiratory rate, pulse rate, and best motor response. Sacco et al. have provided evidence-based survival probability estimates for each RPM score through logistic regression, as well as deterioration rates that are estimated by experts for each RPM score through the Delphi method.

The logistic function used by Sacco et al. for estimating survival probability is

$$P_s = \frac{1}{(1+e^{-w})} \tag{3.3}$$

where P_s is the survival probability estimate. The parameter w is calculated using

$$w = w_0 + (w_1 \times \mathrm{RPM}) \tag{3.4}$$

where w_0 and w_1 are weights that were determined by Sacco et al. through data analysis.

Figure 3.3 depicts the relationship between RPM and survival probability (based on Table 2 of [3])

Sacco et al. also used the Delphi method to estimate casualty deterioration before a casualty reaches definitive care, and expressed it by the decline of the RPM score, as shown in Table 3.1 (reproduced Table 3 of [3]) and Figure 3.4. For instance to better understand Table 3.1, a casualty with an initial RPM score of 12 may degrade to RPM of 11 after two hours (smaller RPM value corresponds to lower survival probability). Here we assume that the casualty received little or no treatment while awaiting transportation to a higher (more definitive) level of care.

It should be noted that the RPM scores are widely used in this dissertation as a measure of a patient's health condition. For example, when an evacuated casualty arrives at a hospital, a

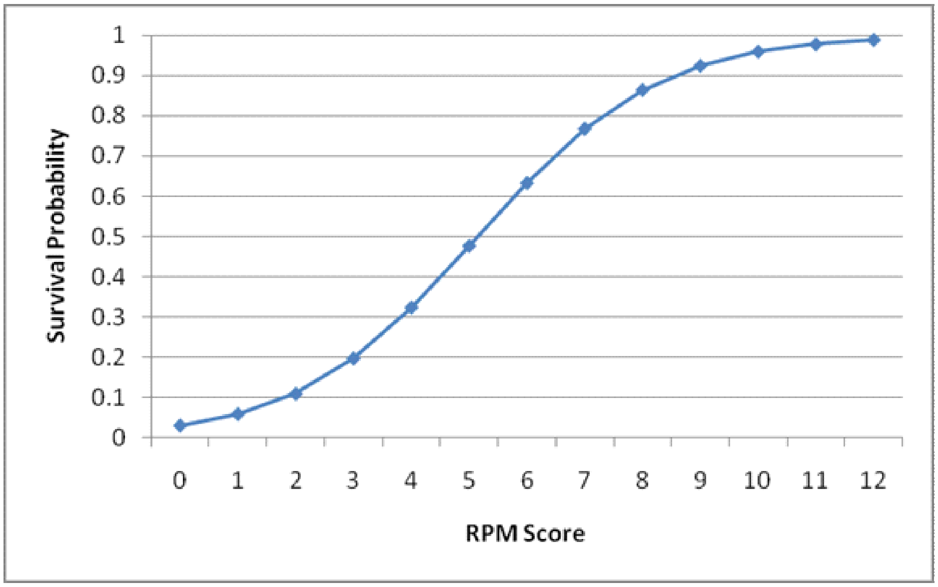


Figure 3.3: RPM vs. Survival probability

Table 3.1: Delphi Estimates of degraded RPM scores in 30-Minute Intervals [3]

Initial RPM	Time Intervals											
	1	2	3	4	5	6	7	8	9	10	11	12
12	12	12	11	11	10	10	9	9	9	8	8	8
11	11	11	10	10	9	8	7	7	7	6	5	5
10	10	9	9	8	8	7	6	5	5	5	4	4
9	9	8	8	7	5	4	3	1	0	0	0	0
8	7	6	4	3	3	2	2	0	0	0	0	0
7	6	5	3	1	0	0	0	0	0	0	0	0
6	4	3	2	1	0	0	0	0	0	0	0	0
5	3	2	0	0	0	0	0	0	0	0	0	0
4	2	0	0	0	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0

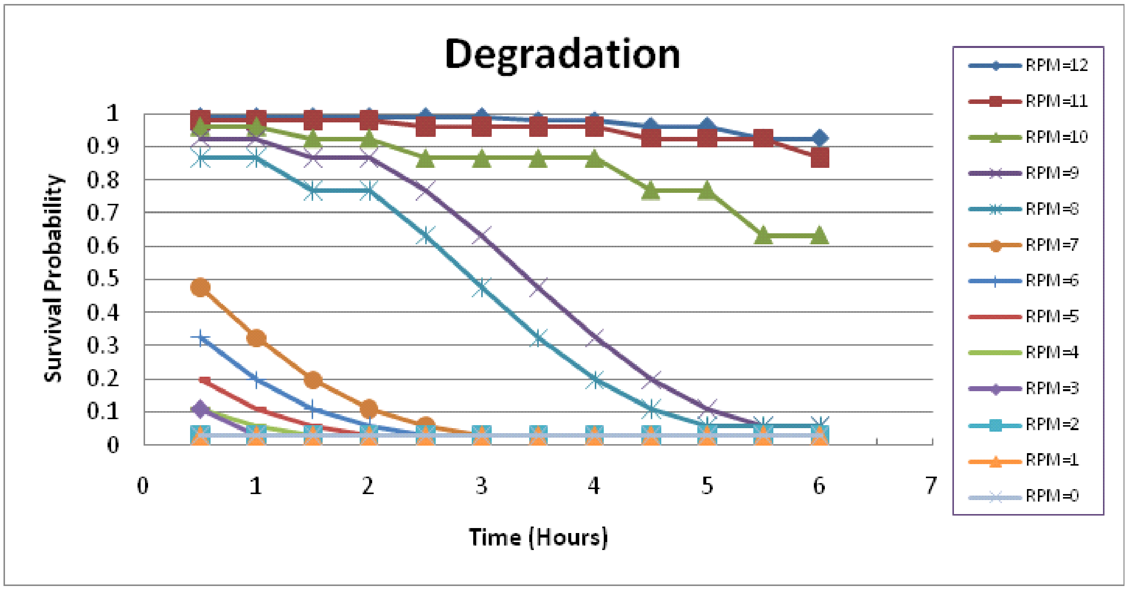


Figure 3.4: Degradation in RPM scores under different initial conditions

(degraded) RPM score is given after triage, which determines whether or not the casualty should be admitted or discharged. (See Section 3.2.1 for a more detail description.)

3.1.7 Pre-hospital sub-system

The pre-hospital sub-model involves ambulances and the incident command. In the beginning of the simulation, ambulances are located at their bases and waiting for orders. At the time that an incident is reported, the ambulances are sent to the incident site. The first arrival ambulance initiates triage as described in Section 3.1.6.1. The subsequent ambulances evacuate patients to different hospitals as directed by the incident command. The incident command decides where the patients should be sent according to the feedback information from ambulances in conjunction with its understanding of the current state of available hospitals.

3.1.7.1 Ambulance Ambulances are modeled by Performer agents. They are responsible for casualty evacuation from the incident site to hospitals, and their major tasks include:

1. Stay-and-wait for orders from incident command;
2. Travel along a calculated shortest path to the disaster site;
3. Evacuate patients from the disaster site to hospitals.

In the beginning of the simulation, all ambulance agents are located at their bases (nodes) and are assumed available. Upon receiving a "go-to-incident-site" order, each ambulance will calculate a shortest path (the one takes the least travel time) from its base node to the incident node, then set out and head to the incident node along the calculated path.

When an ambulance arrives at the incident site, it will begin to load casualties based on the following rules:

1. Triaged Only: only triaged casualties can be loaded;
2. Two passengers at most: one ambulance can take at most two casualties on each trip;
3. "Worst-first" pickup strategy: if there are casualties triaged differently at the scene, an ambulance should load a red casualty first, then yellow, then green;
4. One red casualty per vehicle: once an ambulance loads a red casualty, the other casualty it takes can only be yellow or green;
5. Same type principle: on each trip an ambulance can only take casualties of the same type, determined by the first loaded casualty's type.
6. No waiting at incident site: an ambulance will be informed about available triaged casualties immediately upon its arrival, then it has to make an instant pickup decision and begin loading. Only if there are no triaged casualties ready for transport at the scene (but still have some triage ongoing), an ambulance is allowed to wait at the scene for next triaged casualty;
7. No replacement once loaded: for example, when an ambulance arrives, and there are only two green casualties waiting for evacuation. According to "no waiting as possible" rule, the ambulance should begin to load them. Once the pickup decision was made, no change is allowed even if there is a red casualty being triaged while the ambulance is being loaded.

After casualties are loaded, the ambulance requests instructions from incident command about which hospital it should go to. Incident command then chooses a target hospital following the current evacuation policy and provides the selected hospital to the ambulance.

Table 3.2 lists all possible states for a loaded ambulance.

Table 3.2: Ambulance loading rules

Loading State	Comment
1 red	No same type casualty with triaged yellow/green available when making pickup decision
1 red + 1 yellow/green	Two passengers are the same type
2 yellow/green	Only when no red casualties are present. 2 green casualties would be loaded only when no yellow casualties are present.
1 yellow/green	Only when no red casualties are present and only a single yellow/green casualty of the given casualty type is available

3.1.7.2 Incident command The incident command is modeled by a Commander agent, which is a full-function agent that can exchange information with other agents and make operation decisions.

The incident command can collect information from the incident site, ambulances and hospitals, so that it knows the status of the entire response system. Using the information provided by on-scene triage and hospitals in accordance with the current policy in effect, the incident command then assigns ambulances and casualties to specific hospitals when the ambulance picks up casualties.

The list below concludes major duties of simulated incident command:

- Receipt of information from ambulances regarding the condition of loaded patients;
- Matching of patients with specific injuries to facilities capable of treating these specific problems (e.g., a burn patient to a burn center or facility with burn care capability);
- Indicating the target hospital for loaded ambulances;
- Notification of hospitals of the number of patients they should expect;
- Receipt of information from hospitals about the availability of beds;
- Balancing the loads of hospitals so as to not overload any one hospital.

3.1.8 In-hospital sub-system

Hospitals are the last stage of the response system, each evacuated casualty will be sent to a hospital for definitive care. There are two types of hospital in the system – specialized and regular hospitals. Specialized hospitals are defined as those that can provide the specific treatments required by specialized type of patients. A typical hospital is assumed to consist of the following medical units: emergency department (ED), intensive care unit (ICU), operating room (OR) and general wards (GW).

The size of each unit in a hospital is the number of beds available for injured casualties, after accounting for on-going operations. The model can also track casualties in ambulances en route to the hospital. However, the hospital only reports to the incident command the information that is required for the policy being evaluated.

When a casualty arrives at the destination hospital, the casualty enters the emergency department, where medical staff will perform arrival triage. The casualty is then classified into one of two triage categories: critical and non-critical.

Critical casualties are those who may need resuscitation or urgent surgery. The critical patient will be moved to a bed in the emergency department and receive necessary care and diagnosis from an emergency medical specialist. Upon the diagnosis, the specialist will make a decision whether or not an urgent surgery is needed. If no surgery is needed, the patient will be sent to a bed in either ICU or GW. Otherwise, the patient will be moved to an operating room for surgery.

For the patients who are diagnosed as non-critical during the arrival triage, they wait in a waiting room until a bed becomes available in the ED so that they can receive further examination from medical staff. The staff will then decide whether the patient should be admitted into a general ward or be discharged.

If the patient is admitted as an inpatient, a bed in the relevant ward is assigned. However, if the relevant ward is full, the patient would be prevented from moving into a ward, which would cause a block in the emergency department.

It should be noted that for severe trauma patients who needs to be moved directly into OR, a bed in a ward or ICU typically must be found before admission to the OR is allowed. If there is no bed available, the critical patient will be transferred to another hospital, which requires an ambulance and results in further delay before definitive care.

A brief chart of patient flow in a hospital can be found in Figure 3.5.

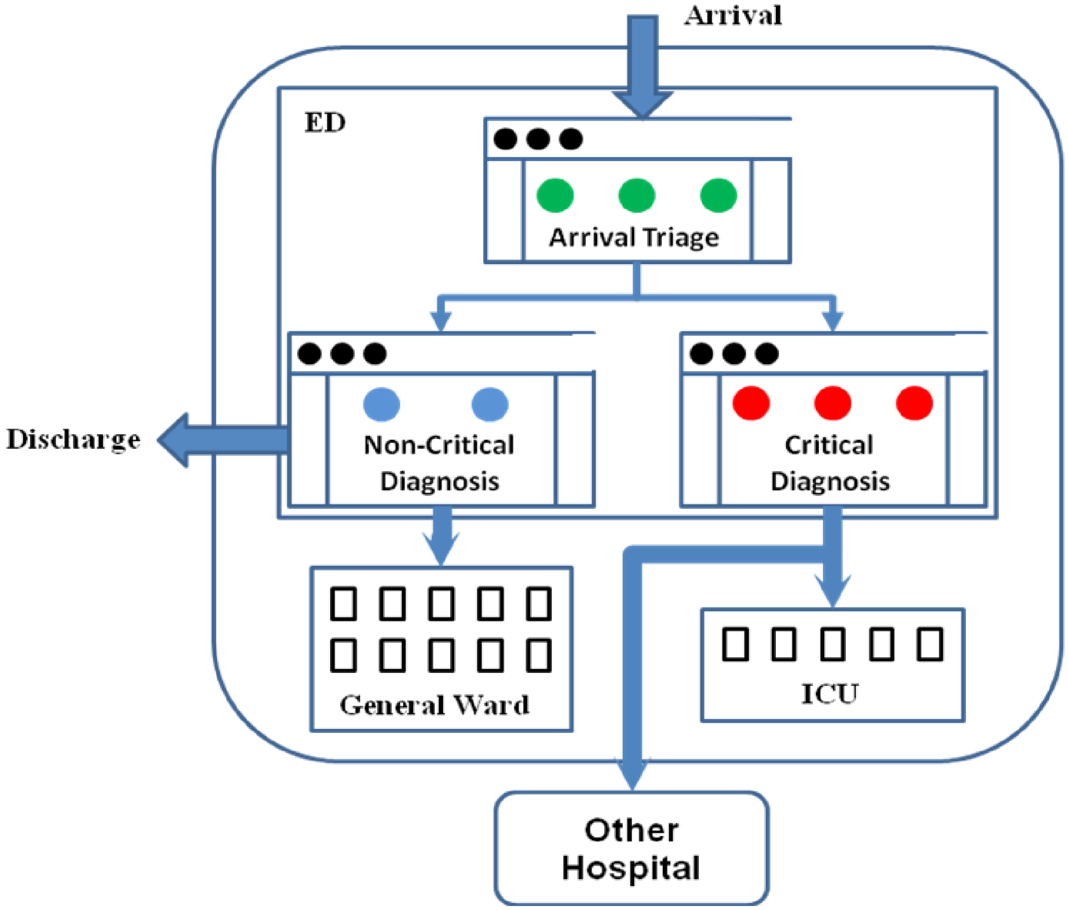


Figure 3.5: Hospital process flow

Based on the analysis above, each hospital is modeled as a single full-functional agent, which contains different medical departments (ED, ICU, and GW). Each department can be considered as a parallel-processing workstation that can process (treat) several workpieces (patients) simultaneously. Each hospital agent instance takes evacuated casualties as its input. The admitted patients will be moved among different "workstations" (medical departments) to be "processed" (examined or treated). Finally, the model will calculate the survival states for those patients who receive definitive cares based on his/her survival probability at that time and estimate the overall mortality thereby.

The definitive cares refer to following treatments:

- A non-critical patient is discharged directly;
- A non-critical patient is assigned a GW bed;
- A critical patient is assigned a bed in ICU; or
- A critical patient who is transferred to other hospital survives to be cared there.

3.1.9 Performance measure

The performance measure for the response model is the overall mortality among the casualties. At the beginning of the incident, each casualty is randomly assigned an initial survival probability based on a distribution chosen to correspond to the incident being modeled. Over the course of the simulation, each casualty's survival probability degrades according to a certain casualty degradation model mentioned in Section 3.1.6.3. This continues until the casualty reaches definitive care (i.e., after the casualty has completed in-hospital triage and has been admitted to the ICU, GW or discharged). When the casualty reaches definitive care, his/her ultimate survival is determined by comparing his/her survival probability at that time with a random number drawn from $[0,1]$. The overall mortality of all the casualties is the mortality of that run of the simulation.

3.1.10 Model validation

According to Brown et al. (2004) [83], there are usually two steps to establish confidence to a computer model: verification and validation. Verification is to verify that the program is free of bugs and correctly implements the conceptual model; and validation is to validate the model by showing it generates output that matches the relevant aspects of the system being modeled.

Verification and validation are critical processes of simulation studies since they provide guarantee that a simulation model can represent the real system and gives realistic results for making reliable decisions. However, it is challenging to validate a complex, large-scale simulation system due to its randomness and numerous internal operations and interactions.

Gass (1983) [159] summarized various validation methods, as listed below:

- Face validation (expert opinion). Ask subject matter experts to review the model and judge if it satisfies with their knowledge.
- Technical validation. See if the model assumptions are plausible and if the outputs are reasonable.
- Structural validation. See if the model operates in the similar way as the real system to produce comparable behaviors.
- Sensitivity analysis. Investigate how the model behaves when its variables and parameters change and compare to the real-world system.
- Replicative validation. See if simulation results match data obtained from the real system.

For this research, the major validation methods used are face validation and technical validation. We asked subject matter experts to check the model assumptions and to review the simulation results. From their feedback we are confident that the model is valid and can be used to compare different operation policies reliably. Besides that, multiple sensitivity analysis experiments are also performed to validate the robustness of model under different parameter configurations (e.g., different casualty degradation models).

3.2 CASE STUDY

3.2.1 Assumptions, constraints and parameter settings

The case study is used to validate the model by showing it can generate reasonable simulation results under given input parameters. For this study, we assume an IED (Improvised Explosive Device) explosion at the Pittsburgh D. L. Lawrence Convention Center in downtown Pittsburgh, PA, United States. There are 150 patients that require medical care. Casualties are of two casualty types: children and adults. Children have to be treated at one of two specialized hospitals: Children's Medical Center or Magee Women's Hospital. For each of the 10 total hospitals, we assume that there are 10 available beds in general wards and 5 beds in ICU in the beginning of simulation. The injury severities are modeled by different initial survival probability values: the larger values indicate mild injuries and the smaller values correspond to severe injuries. We also assume that the injury severity of each victim is independently and identically distributed according to a specified exponential distribution. The EMTs only can estimate the injury status based on the information gathered during triage, namely that a victim is either specialized or general and the severity is triaged red, yellow or green. The actual survival probability for the casualty will deteriorate continuously before definitive care is received at a hospital. After a casualty is evacuated to the hospital, emergency room staff will perform in-hospital triage and use Sacco's RPM score to indicate the casualty's injury severity (after degradation). The RPM score will be used to decide if a casualty should be admitted or discharged by comparing it to a pre-defined threshold; for those being admitted, another threshold value will be used to decide if they are in critical condition or not. In addition, we assume that regionally there are 24 ambulances available to respond to the incident. The ambulances initially start in one of 6 bases that are distributed over the Pittsburgh region.

3.2.2 Transportation network construction

We generate the transportation network by using a simplified version of the Pittsburgh area road network. We then choose 202 nodes, to include the incident site, intersections of major roads, and locations of hospitals and ambulance bases. Then each resulting road segment is assigned a baseline speed which will be used to calculate the shortest path for ambulances. Details of the construction of the model from GIS data are given in Zimmerman et al. (2010) [19].

3.2.3 Evacuation policy

The evacuation policy refers to a dispatch policy that governs the incident command's assignment of triaged casualties to hospitals. As each transport ambulance reaches the scene, it picks up triaged patients based on their triaged priority. Then, incident command provides the ambulance with its destination based on the type of the patient(s) and the status of the hospitals using the policy described as below.

First the incident command identifies the hospitals with corresponding type and having available ICU and GW beds, then it selects a subset from those hospitals that have positive available ED capacity (which means patients can get immediate treatment upon their arrival). Finally it selects one nearest hospital from the subset.

The available ED capacity (AEDC) will be calculated as:

$$\text{AEDC} = \max\{0, (\text{\# of Available ED Beds} - \text{\# of scheduled incoming patient})\}$$

If all hospitals' AEDC is 0, then the incident command selects one nearest hospital from a subset that contains all hospitals which have the equally shortest length of total waiting queue, that is, the hospital should have the shortest length of total waiting queues (including arrival waiting queue at ED, non-critical diagnosis waiting queue at ED, critical diagnosis waiting queue at ED, ICU waiting queue, and GW waiting queue). It should be noted that the number of scheduled incoming patients will also be counted into the total waiting queue length.

3.2.4 Parameter setting

Table 3.3 lists the parameters used in the simulation model. The parameters are classified into different categories according to their characteristics. As part of the model validation, we tested it using a range of input parameters for the incident setting to simulate different emergency situations.

3.2.5 Numeric experiments

We run the emergency response model using different input configurations. Each configuration was run for 300 replications, and the results are shown in the box-plots which identify the mean, 25 and 75 percent quartiles, and the range of mortality among the 300 replications.

The initial simulation scenario includes n_c = 150 injured casualties, with a percentage of specialized type P_s = 0.2. The initial injury severity follows a exponential distribution with Λ = 0.4. Sacco's degradation model is used as the casualty degradation model. The simulation results for this scenario are used as a standard to compare with the result of other input configuration settings, as shown in the following sections.

Table 3.3: Simulation parameter settings

Category	Parameter	Value
Casualty Setting	Num. of Casualties	$n_c \in \{50, 100, \mathbf{150}, 200, 300, 400\}$
	Initial Survival Probability Distribution	Expo(Λ), $\Lambda \in \{0.1, 0.3, \mathbf{0.4}, 0.5, 0.7, 0.9\}$
	Percentage of Specialized Patients (Children)	$P_s \in \{0.0, \mathbf{0.2}, 0.4, 0.6, 0.8, 1.0\}$
	Casualty Degradation Model	{Proportional-hazard based, **Sacco's**}
Ambulance Setting	Num. of Ambulance Bases	$n_b = 6$
	Num. of Ambulances	$n_a = 24$ (4 at each base)
Hospital Setting	Num. of Hospitals	$n_h = 10$
	Initial Available GW Beds	$n_{GW} = 10$
	Initial Available ICU Beds	$n_{ICU} = 5$
	Surge Capacity Ratio	$r_{sc} = 0$
	Num. of Triage Beds at ED	$n_{ar} = 3$
	Num. of Non-critical Beds at ED	$n_{ncd} = 2$
	Num. of Critical Beds at ED	$n_{cd} = 3$
	Admitting threshold (RPM score)	11
	Critical threshold (RPM score)	4
Time Setting	On-site Triage Time (min)	0.5
	ALS Pickup Time (min)	Gamma(μ = 19.15, sd = 13.98)
	BLS Pickup Time (min)	Gamma(μ = 9.27, sd = 6.43)
	Drop-off Time (min)	Gamma(μ = 23.16, sd = 12.56)
	Arrival Triage Time in hospital (min)	Gamma(μ = 5, sd = 0.5)
	Non-Critical Examination Time (min)	Gamma(μ = 7, sd = 0.5)
	Critical Examination Time (min)	Gamma(μ = 9, sd = 0.5)
Stopping Criteria – All living casualties have reached definitive care.		

3.2.5.1 Different number of casualties In the first experiment set, we checked the impact due to the number of casualties. We simulate six different settings, i.e., $n_c \in \{50, 100, 150$ (the standard), $200, 300, 400\}$. The results are compared using a box-plot in Figure 3.6. We can see that the mortality increases with the increasing number of casualties, which follows intuition since it takes longer to evacuate more number of casualties to hospitals, and the average waiting time also becomes longer for a casualty to receive the definitive care, which explains the mortality increase. Besides that, we observed that when the total number of casualties is below the total medical capacity (beds) in the region (CasNum $\leq$ 150), the mortality increases relatively slow. However, when the total number of casualties exceed the medical capacity (CasNum $>$ 150), the mortality increases faster.

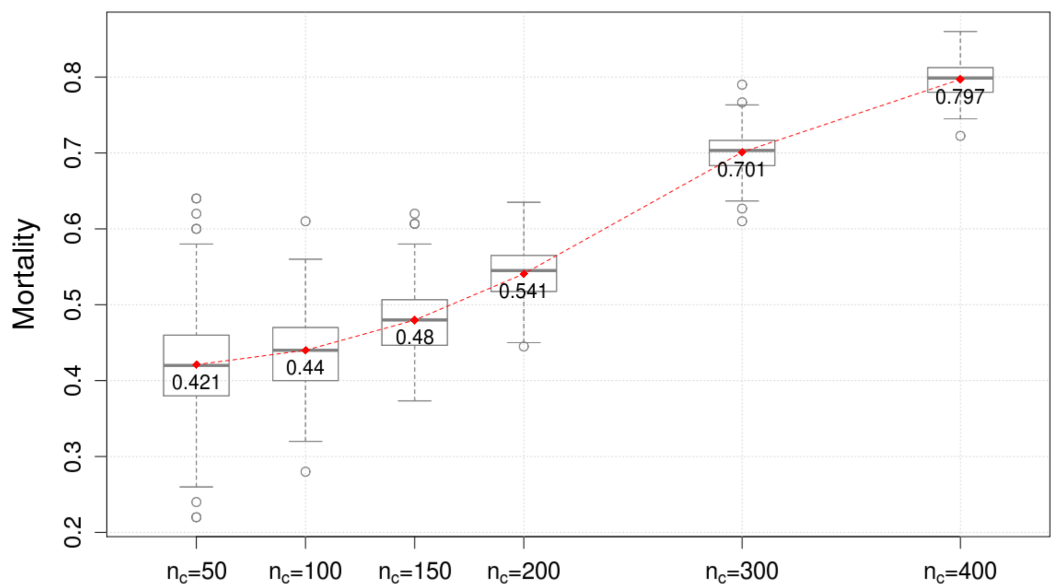


Figure 3.6: Comparison among different number of casualties

3.2.5.2 Different injury severity distribution The second set of experiments is used to check the impact of different injury severity distributions. By changing the severity distribution rate Λ, we can simulate incidents with different levels of scale. In general, a large Λ corresponds to a milder incident since more of the casualties will have higher initial survival probabilities, or larger RPM scores. And along with the decrease of Λ, the proportion of severe casualties (who have smaller survival probabilities or smaller RPM scores) increases so that a smaller Λ corresponds to a more severe incident which contains more seriously injured casualties.

In this set, we check six scenarios, $\Lambda \in \{0.1, 0.3, 0.4$ (the standard), $0.5, 0.7, 0.9\}$, while $\Lambda = 0.1$ corresponds to the most severe incident and $\Lambda = 0.9$ corresponds to the mildest incident (among the set). Figure 3.7 demonstrates different injury severity distributions under different Λ ($\Lambda = 0.1$, 0.4, and 0.9), where the injury severity is measured by Sacco's RPM scores.

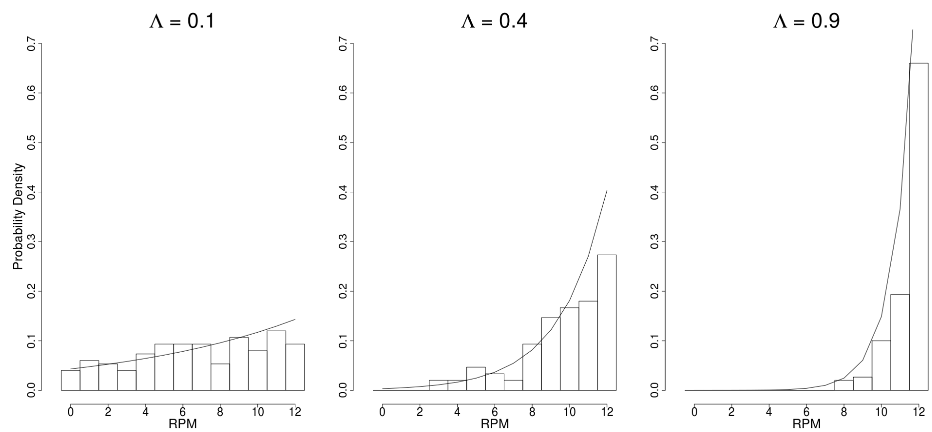


Figure 3.7: Different injury severity distributions

Figure 3.8 displays the simulation results. As expected, an incident with more severely injured casualties leads to a higher mortality than the one with fewer severely injured patients. In addition, the decrease of mortality for smaller values of Λ ($\Lambda \leq 0.4$) is more obvious than that for larger Λ ($\Lambda > 0.4$).

3.2.5.3 Different percentage of specialized patients The third experiment set is to check the impact of different percentages of specialized patients. The specialized percentage P_s is set to 0.0, 0.2 (the standard), 0.4, 0.6, 0.8 and 1.0 respectively, which correspond to the cases that none, 20%, 40%, 60%, 80% and all of the casualty population are specialized type (children). As we have mentioned in the Section 3.1.6.2, specialized type of patients have to be sent to and treated by a specialized hospital. Since there are only two specialized hospitals out of ten (20%), it is expected

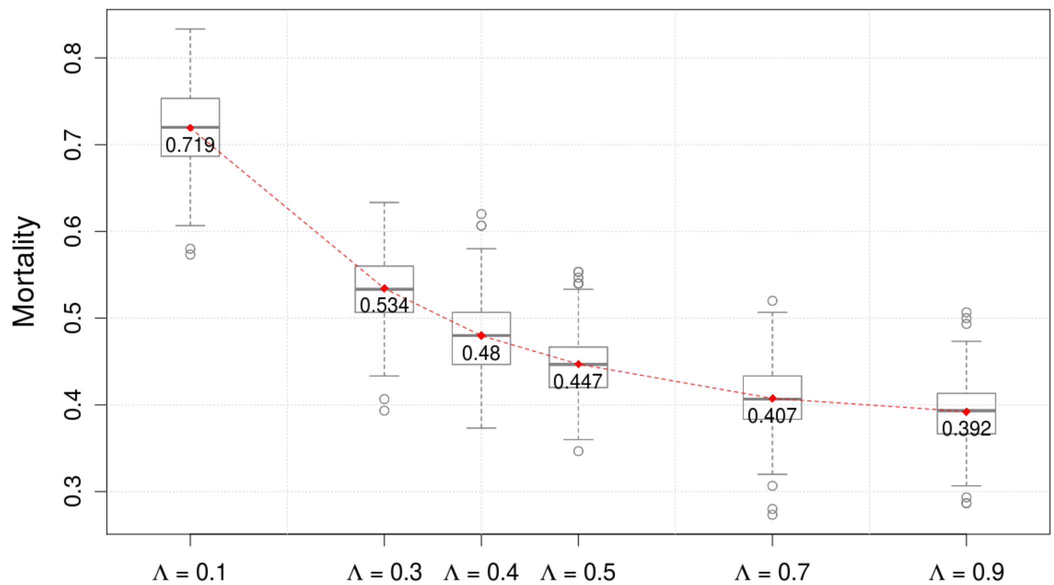


Figure 3.8: Comparison among different injury severity distributions

to observe increasing mortality as the specialized percentage increases, as shown in Figure 3.9. It can be observed the increase of mortality is not linear although the percentage increases in a linear fashion, which reflects the non-linear nature of the system.

3.2.5.4 Different degradation model The last experiment set tests the impact of different casualty degradation models – the Sacco's RPM-based model and the proportional-hazard based model. The Sacco's RPM-based degradation model is used as the standard. In alternative scenarios, the casualties' survival probabilities deteriorate based on the proportional-hazard based model (Formula 3.2, which is re-written as below).

$$P_t = P_0 \times R(t) = P_0 \cdot g^{-t}$$

Where P_0 is the initial survival probability, g is the deteriorate base and t is the elapsed time.

In order to eliminate the impact from irrelevant factors, we tested the two degradation models using the same casualty data sets, which guarantees the initial survival probabilities are identical. In addition, a set of different values for g ($g \in \{1.0, 1.045, 1.196, 2.007\}$) is chosen to investigate

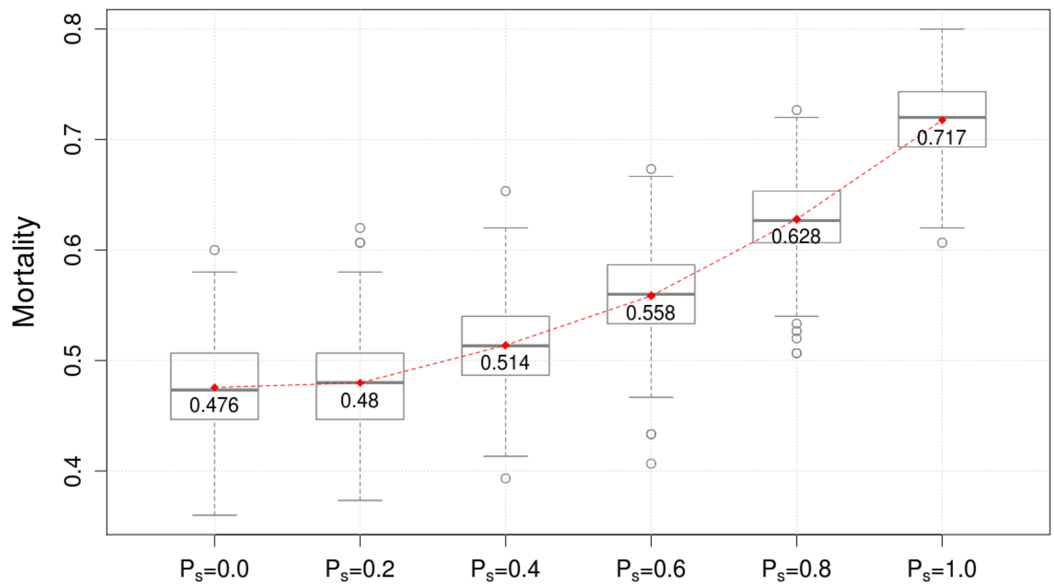


Figure 3.9: Comparison among different percentages of specialized casualties

the impact due to the change of deteriorate base. The reasons of choosing these particular values are explained as below.

The choice of $g = 1.0$ is used to test an extreme assumption that there is no degradation occurring for all casualties. In such a case, the survival of each casualty is solely determined by his/her initial survival probability. The choice of g =1.045 is a result of trial and error, with a purpose to make the mean mortality of proportional-hazard based model close to that of Sacco's RPM-based model.

The choices of g =1.196 and 2.007 are results of fitting the proportional-hazard based model to the data provided by Sacco's model. In Sacco's RPM-based model, the injury deteriorates in different ways for different initial conditions (as shown in Figure 3.4). However, the deterioration rate of proportional-hazard based model disregards the initial condition and solely depends on one parameter g. In order to make the comparison based on certain common foundation, we fit the proportional-hazard based model using the data provided by Sacco's model to determine the value of g, the methods of fitting are summarized in Table 3.4.

Table 3.4: Finding g by Fitting

Result	Method of Fitting
g = 1.196	1. Refer to Figure 3.3 to covert the RPM scores in Table 3.1 to corresponding survival probabilities; 2. Average the survival probabilities (over all 13 initial conditions) at each time point to get a series of paired-data (t, s) = (time, average of survival probabilities at time t); 3. Fit the proportional-hazard based model using the series of paired-data (t, s) to determine g.
g = 2.007	1. Refer to Figure 3.3 to covert the RPM scores under initial RPM=6 (which is a median) in Table 3.1 to corresponding survival probabilities; 2. Record paired-data (t, s) = (time, survival probabilities with initial RPM=6 at time t); 3. Fit the proportional-hazard based model using the series of paired-data (t, s) to determine g.

The simulation results are shown in Figure 3.10. It is not unexpected to find the minimal mean mortality achieved by $g = 1.0$ due to no degradation happening to casualties, and the result of g =1.045 is very similar to that of Sacco's RPM-based model since it was so designed.

From Figure 3.10 we can see that that the proportional-hazard based degradation models with g = 1.196 and 2.007 lead to more mortality than Sacco's degradation model, that is because the proportional-hazard based degradation model assumes any casualty deteriorates in the same pattern despite of his/her initial injury severity, which results much faster degradation rates for those mildly injured casualties than Sacco's degradation model.

3.2.6 Summary

In this chapter, we have developed an agent-based simulation model for emergency medical response to a mass casualty incident in an urban area. Three interrelated sub-systems (incident site, pre-hospital, and in-hospital) and various interactive agents are developed and introduced in details.

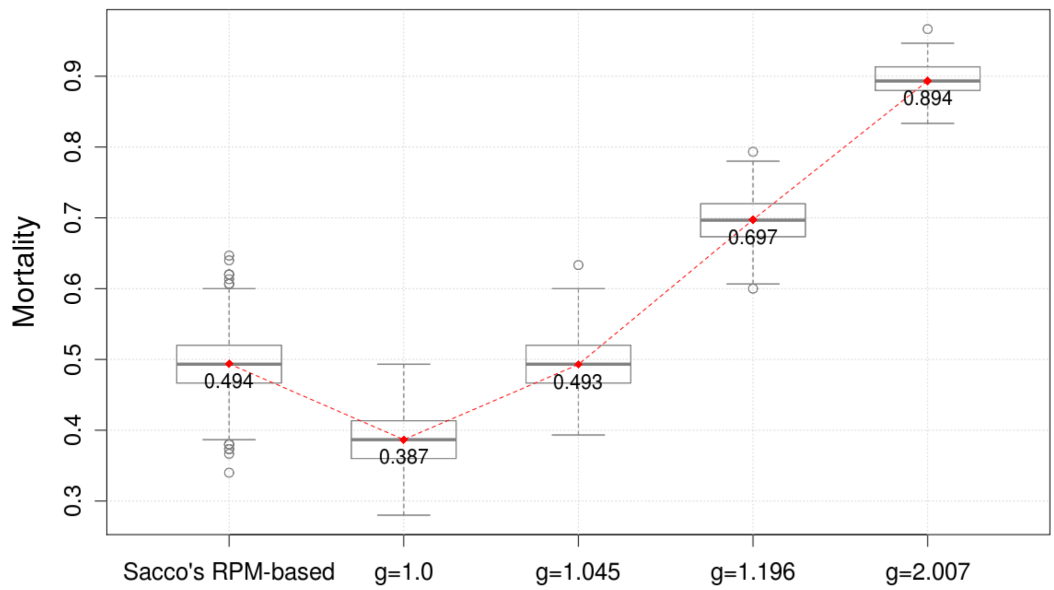


Figure 3.10: Comparison among different degradation models

During the development, we have defined three new generic agent types (Indicator, Performer, and Commander). The new generic agent types provide a set of prototype templates to derive new agents, which makes the model easy to expand to contain more different functional agents. Besides that, the simulation model provides an interface to import processed GIS data to construct the transportation network, which facilitates researchers considering the effects of different traffic to the casualty evacuations. In addition, it also enables displaying the ongoing evacuation status on a GIS view dynamically along with the running simulation, which gives a more direct illustration to the researchers about the evacuation process.

This methodology can be used to build similar models for other cities or areas at a relatively low level of investment of time. This model can also be used to evaluate other decisions such as the effect of increasing the number of ambulances, introducing additional hospital beds, or identifying good locations for additional medical and emergency response facilities.

Like all such models, there are limitations in interpretation. Currently, it only reports a single performance measure – mortality. In cases where there are more complex evaluation criteria, the model could be modified to report other performance measures or combinations of performance measures. For example, the morbidity of injured patients could become another measure to evaluate the MCI response performance. Second, it only reports quantitative results. Decision makers using this model should be aware of other factors that may impact decisions such as negotiated agreements or financial factors that should be considered in conjunction with the results of this model.

4.0 POLICY RANKING AND SELECTION

4.1 INTRODUCTION

The agent-based simulation (ABS) model developed in Chapter-3 provides a natural way to capture the complex behaviors of a mass-casualty incident response system from the ground-up. It eliminates many of the assumptions needed for mathematical programming formulations so that the system can be modeled in a more realistic way. The model is then used to simulate the impacts due to different response policies, and the best policy that leads to the least (minimum) mortality can be identified by comparing the simulation results.

Due to the randomness involved in simulation, it is improper to run the policies for only one round, and make a decision based on the single round of observation. Multiple replications are needed for each policy and certain optimization via simulation (OvS) techniques must be used to analyze the results in order to obtain a statistically confident conclusion.

In this chapter, we show how the best response policy can be selected from a set of alternatives efficiently using ranking-and-selection (R&S) techniques. We implement two R&S procedures (the Rinott procedure and the KN procedure) and compare their efficiencies. Although both procedures are valid in selecting a single best with a specified confidence level, they are both deficient in selecting a subset containing all alternatives that are "close enough" to the best one. Hence, we then argue that a new selection procedure should be developed to help decision makers select the best-subset while providing a statistical guarantee for the relative correctness of that selection set.

4.2 POLICY DESCRIPTION

In this dissertation, the term “policy” refers to a set of pre-defined principles or rules to guide decisions and to achieve rational outcomes. It is exchangeable with the term “action procedure” or “protocol”. For emergency response, there are usually various pre-defined policies to guide the decisions/actions of different agents/sub-systems, such as the protocol for the first responder at the scene, the evacuation policy for ambulances, or the admission policy for hospitals, etc.

In this chapter, we focus on the evacuation policies. These policies govern the incident command routing (assignment) of ambulances at the incident site to hospitals once the ambulance has been loaded with triaged casualties. The policies differ in terms of the information required for the hospital status (space or bed availability at emergency department (ED)) and the thresholds for closing specialized hospitals.

Currently, twelve different evacuation policies (P-1 to P-12) are proposed and employed to guide the casualty evacuation; the details are included in Table 4.1. Our objective is to select the best evacuation policy that leads to the minimal mortality from these alternatives.

Table 4.1: Twelve evacuation policies

P-1 – Random_Dispatching

Description: Select a hospital at random from all hospital candidates

Information Exchanging: (None)

P-2 – Shortest_Arrival_Waiting_Queue (only the arrival waiting queue for triage at ED is considered)

Description: Select the nearest hospital (from the incident site) from a subset which involves those hospitals that have the shortest waiting queue of arrival patients at ED.

1. (Incident command receives the “where-to-go” inquiry from an ambulance);
2. Incident command checks the status of each hospital, identifies a subset which contains those hospitals having the shortest arrival waiting queue;
3. Incident command selects the nearest one from the subset as the target. (A random choice would be made if there is a tie)

Information Exchanging: (One-way)

- Hospitals → Incident command (the length of arrival waiting queue)

(Continued on next page . . .)

(Table 4.1: continued)

P-3 – Shortest_Waiting_Queues (all waiting queues are considered)

Description: Select the nearest hospital from a subset which involves those hospitals that have the shortest length of total waiting queues, that is, the hospital should have the smallest number of total waiting patients (including arrival waiting queue at ED, non-critical diagnosis waiting queue at ED, critical diagnosis waiting queue at ED, ICU waiting queue, and GW waiting queue)

Information Exchanging: (One-way)

- Hospitals → Incident command (the length of total waiting queues)

P-4 – Revised_Shortest_Waiting_Queues (all waiting queues + expected coming)

Description: Similar to P-3, but the number of scheduled incoming patients is also counted into the length of total waiting queues.

Information Exchanging: (Two-way)

- Hospitals → Incident command (the revised length of total waiting queues)
- Incident command → Hospitals (# of scheduled incoming patients)

P-5 – Available_First_otherwise_Shortest_Waiting_Queue

Description: First try to select the nearest hospital from a subset that involves hospitals having positive available ED capacity (which means patients can get immediate treatment upon their arrivals). The available ED capacity (AEDC) is calculated as:

AEDC = max{0, (# of available ED beds – # of scheduled incoming patients)}

If all hospitals' AEDC is 0, then P-4 is used to select one hospital with the shortest revised length of total waiting queues.

Information Exchanging: (Two-way)

- Hospitals → Incident command (AEDC & the revised length of total waiting queues)
- Incident command → Hospitals (# of scheduled incoming patients)

P-6 – Available-Capacity_AEDC__otherwise_Shortest_Waiting_Queue

Description: First identify hospitals with available ICU & GW beds, then use P-5 to select the target.

Information Exchanging: (Two-way)

- Hospitals → Incident command (available capacity of beds & the revised length of total waiting queues)
- Incident command → Hospitals (# of scheduled incoming patients)

(Continued on next page ...)

(Table 4.1: continued)

P-7 – P-6 in Specialized-Hospital-Reserved mode

Description: An ambulance can only pick up same type of patients and send them to corresponding hospitals (general patients to general hospitals, specialized patients to specialized hospitals). Other rules are the same as P-6.

Information Exchanging: (Three-way)

- Ambulances → Incident command (type of casualties loaded)
- Hospitals → Incident command (available capacity of beds & the revised length of total waiting queues)
- Incident command → Hospitals (# of scheduled incoming patients)

P-8 ∼ **P-12** – P-6 in First-Open-Then-Reserved Mode

Description: In the beginning of the simulation, the system is under All-Hospital-Open mode, which means that Policy-6 is used (specialized casualties must go to specialized hospitals, but general patients can be sent to any hospitals, and the target hospital is chosen using P-6). However, after a specified number of specialized type of casualties (n_s) are observed, the system switches to Specialized-Hospital-Reserved mode (P-7), and the thresholds for different policies are:

P-8: $n_s = 3$; P-9: $n_s = 6$; P-10: $n_s = 9$; P-11: $n_s = 12$; P-12: $n_s = 15$;

Information Exchanging: (Three-way)

- Ambulances → Incident command (type of casualties loaded)
- Hospitals → Incident command (available capacity of beds & the revised length of total waiting queues)
- Incident command → Hospitals (# of scheduled incoming patients)

These evacuation policy are used to respond an IED explosion incident described in Section 3.2.1 with the same assumptions and constraints. Table 4.2 presents all parameters used in the simulation study.

4.3 SIMULATION RESULTS

We first performed a pilot study, in which each evacuation policy was run for 300 replications (indexed by 1, 2, ..., 300). In order to highlight the difference due to different policies, we generated 300 casualty data sets corresponding to the 300 replications. Each data set contains 150 random casualty data whose initial RPM scores are drawn from an exponential distribution with a scale parameter $\Lambda = 0.4$. The benefits of using these data sets are

Table 4.2: Simulation parameter settings

Category	Parameter	Value
Casualty Setting	Num. of Casualties	$n_c = 150$
	Initial Survival Probability Distribution	Expo(Λ), $\Lambda = 0.4$
	Percentage of Specialized Patients (Children)	$P_s = 0.2$ (30 among 150)
	Casualty Degradation Model	Sacco's RPM-based model
Ambulance Setting	Num. of Ambulance Bases	$n_b = 6$
	Num. of Ambulances	$n_a = 24$ (4 at each base)
Hospital Setting	Num. of Hospitals	$n_h = 10$
	Initial Available GW Beds	$n_{GW} = 10$
	Initial Available ICU Beds	$n_{ICU} = 5$
	Surge Capacity Ratio	$r_{sc} = 0$
	Num. of Triage Beds at ED	$n_{ar} = 3$
	Num. of Non-critical Beds at ED	$n_{ncd} = 2$
	Num. of Critical Beds at ED	$n_{cd} = 3$
	Admitting threshold (RPM score)	11
	Critical threshold (RPM score)	4
Time Setting	On-site Triage Time (min)	0.5
	ALS Pickup Time (min)	Gamma(μ = 19.15, sd = 13.98)
	BLS Pickup Time (min)	Gamma(μ = 9.27, sd = 6.43)
	Drop-off Time (min)	Gamma(μ = 23.16, sd = 12.56)
	Arrival Triage Time in hospital (min)	Gamma(μ = 5, sd = 0.5)
	Non-Critical Examination Time (min)	Gamma(μ = 7, sd = 0.5)
	Critical Examination Time (min)	Gamma(μ = 9, sd = 0.5)
Stopping Criteria – All living casualties have reached definitive care.		

- For a specific policy, each replication visits a different independent and identically distributed (IID) casualty data set;
- For the replications with same index (but belonging to different policies), they all visit the same casualty set, so that the variation from casualties can be eliminated, which make it more meaningful to compare different policies.

All replications (300*12) in this pilot study are completed on a personal desktop computer with a 2.21 GHz AMD Athlon(tm) 64 CPU and 2.50GB RAM memory. The whole running time is approximately six hours (i.e., six seconds per replication in average). The simulation results are shown in Figure 4.1.

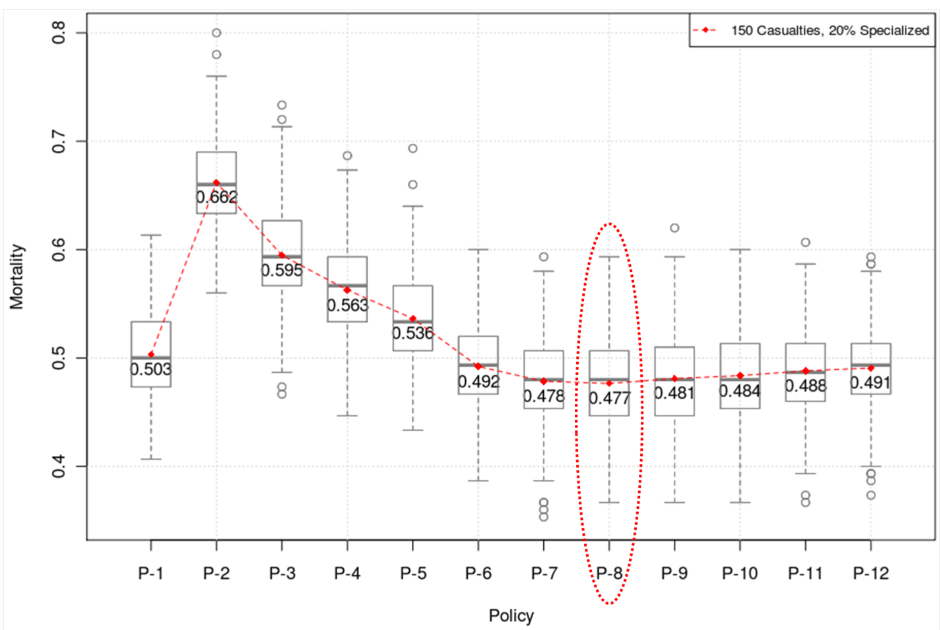


Figure 4.1: Simulation results under 12 evacuation policies

In Figure 4.1, the red dots connected by a dash line indicate the mean mortality for each of the different evacuation policies, and the numbers below the dots mark the concrete values of mean mortality. From the figure, we can observe the following phenomena.

First, P-2 leads to the highest mortality, which is because P-2 only uses the distance and the length of arrival waiting queue at ED as its decision criteria to select the target hospital. However, since the arrival triage is relatively fast, injured casualties tend to be evacuated to those few hospitals nearest to the incident site. After the arrival triage, casualties may have to wait a long time in queues before receiving more definitive care. During this waiting period, the patients' conditions may continue to deteriorate. This is of particular concern for severe patients who may not survive (if they do not receive definitive care in a timely manner), which leads to a higher mortality for the incident.

Besides that, P-2 does not consider the available capacity of beds when making decisions, so that it may send patients to a hospital without sufficient capacity to treat them. As a result, the excess patients will not receive a bed and will have to be sent to other hospitals, which makes the situation even worse.

Compared with P-2, P-3 obtains an improved mortality because it considers all waiting queues instead of only arrival waiting queue, which reduces the negative effect due to patient aggregation. For the same reason, P-4 improves upon P-3 by taking the expected incoming patients into consideration.

P-5 and P-6 differ from previous policies by using the available capacity of medical units as an additional decision criterion. This helps to dispatch patients more reasonably to avoid the "no-beds-for-waiting-patients" situation from occurring, achieving even better results (smaller mortality).

Readers may have noticed that the random dispatching policy (P-1) is a better than P-2, P-3, P-4 and P-5, but worse than P-6. The explanation for this interesting phenomenon is that P-1 balances the load of each hospital evenly, which happens to avoid too long waiting times and "no-beds-for-waiting-patients" circumstances, and therefore achieves a fairly good result. The comparisons between P-1 and P-2 through P-6 suggest that for this special case, decisions based on incomplete information are worse than no information. Only comprehensive information can support a good decision.

P-1 to P6 do not intentionally consider the specialized nature of the injury; in other words, while a certain casualties should only be sent to a hospital with special facilities (e.g., burn patients require a burn unit), a more general type of injury can be treated at either a general type of hospital or a specialized one. However, the medical resources at specialized hospitals are relatively scarce since the number of specialized beds is substantially less than general hospital beds. For our

Pittsburgh example case, two of the ten hospitals are assumed to have specialized beds available. Therefore, it makes sense to reserve some of the medical resources of specialized hospitals for those patients who really need it.

With this in mind, P-7 to P-12 reserve beds for specialized patients, and these policies differentiate each other by their threshold values (n_s), which marks when the Specialized-Hospital-Reserved mode should be triggered. For example, the zero-threshold of P-7 means the system enters the Specialized-Hospital-Reserved mode from the very beginning of simulation. For P-8 through P-12, the system enters the Specialized-Hospital-Reserved mode only after n_s casualties are observed, where n_s are sequentially increasing positive integer series with a constant step equal to 3.

From Figure 4.1, it can be observed that for this case, the minimal average mortality (0.477 or 47.7%) is achieved by P-8, which suggests that P-8 should be selected as the best policy. However, the mean mortality of P-7 is very close to P-8, so it is difficult to determine which one is the true best. The box-plot can provide researchers an intuitive idea about the goodness of each policy, but is unable to present a statistical guarantee about the correctness of selection.

As a summary, we can see that different evacuation policies for MCI response lead to different impacts on mortality. The policies based on comprehensive information achieve better results than those utilizing partial information or no information. Considering that specialized casualties must be treated at specialized hospitals, it is necessary to strategically reserve some capacity of specialized hospitals for those patients who really need it. That is, in anticipation of a major incident, a certain number of beds in specialized units should be made available. However, the cost of reserving these beds may be quite high.

Note that although the box-plot provides good intuition about the goodness of policy, it is incapable to present statistical guarantee for the selection. In order to reach a statistically confident conclusion, the following questions are of interest.

- Are 300 replications per policy enough to make a statistical valid decision to decide the best alternative at certain (saying 95%) confidence level?
- Is there a better way to allocate the computational budget to different alternatives so that the conclusion can be made in a more efficient manner?

4.4 SELECT THE BEST EVACUATION POLICY BY R&S PROCEDURES

As we have discussed in Chapter 2.5, R&S procedures are specifically designed to select the best population from competing alternatives. A R&S procedure can provide a statistical guarantee about the selected result, and can help researchers appropriately allocate simulation resources to reach a decision in an efficient way. In this research, we implement two R&S procedures (the Rinott procedure and the KN procedure) to select the best evacuation policy.

4.4.1 Rinott procedure

The Rinott procedure is a two-stage R&S procedure first presented by Rinott (1978) [108]. As setup, the Rinott procedure requires three input parameters, that is, an indifference-zone parameter δ, a confidence level $P^* = 1 - \alpha$, and a sample size of the first stage n_0. The parameter δ is the smallest actual difference that it is worth detecting. Differences of less than δ are considered insignificant.

In the first stage, n_0 observations are taken from each of the competitive alternatives. The variances calculated from the first stage data are then used to determine the number of observations required in the second stage. Finally, a series of weighted average of sample means (including both the first and second stage) are compared to select the best system. The full Rinott procedure [103] is transcribed as below.

Setup: Select confidence level $1-\alpha$, indifference-zone parameter $\delta > 0$ and first-stage sample size $n_0 \geq 2$.

Initialization: Obtain Rinott's constant $h = h(n_0, k, 1-\alpha)$, (can be calculated by the program provided by [160]).

Obtain n_0 observations X_{ij}, $j = 1, 2, \ldots, n_0$, from each system $i = 1, 2, \ldots, k$.

For $i = 1, 2, \ldots, k$ compute the sample variance of the data from system i:

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

Let $N_i = \max\left\{n_0, \left\lceil \frac{h^2 S_i^2}{\delta^2} \right\rceil\right\}$

where $\lceil \cdot \rceil$ indicates rounding up any fractional part to the next larger integer. Here N_i is the number of observations that will be taken from system i.

Stopping Rule: If $n_0 \geq \max_i N_i$ then stop and select the system with the largest/smallest $\bar{X}_i(n_0)$ as the best. Otherwise, take $N_i - n_0$ additional observations $X_{i,n_0+1}, X_{i,n_0+2}, \ldots, X_{i,N_i}$ from each system i for which $N_i > n_0$.

Select the system with the largest/smallest $\bar{X}_i(N_i)$ as the best.

For this study, we choose as the input parameters: $\alpha = 0.05$, $\delta = 0.01$, $n_0 = 10$, so the Rinott's constant $h = 4.435$. The final R&S results are shown in Table 4.3. Figure 4.2 shows the relationship between the number of samples needed and the sample variance (estimated using the first stage data).

Table 4.3: R&S results by Rinott procedure

Policy ID	Sample variance S_i^2	Total Samples (N_i)	Mortality (Mean)
p1	1.150×10^{-3}	227	0.504
p2	2.106×10^{-3}	415	0.663
p3	1.197×10^{-3}	236	0.594
p4	1.726×10^{-3}	340	0.563
p5	1.523×10^{-3}	300	0.536
p6	1.464×10^{-3}	289	0.492
p7	1.077×10^{-3}	212	0.479
p8	0.775×10^{-3}	153	0.473
p9	0.874×10^{-3}	172	0.478
p10	1.838×10^{-3}	362	0.484
p11	2.835×10^{-3}	558	0.487
p12	1.911×10^{-3}	376	0.491

From Figure 4.2, we can see that the number of samples needed for each policy is proportional to its sample variance estimated by the first stage data. A policy with larger sample variance requires more replications than those with smaller variances. With an indifference-zone with width

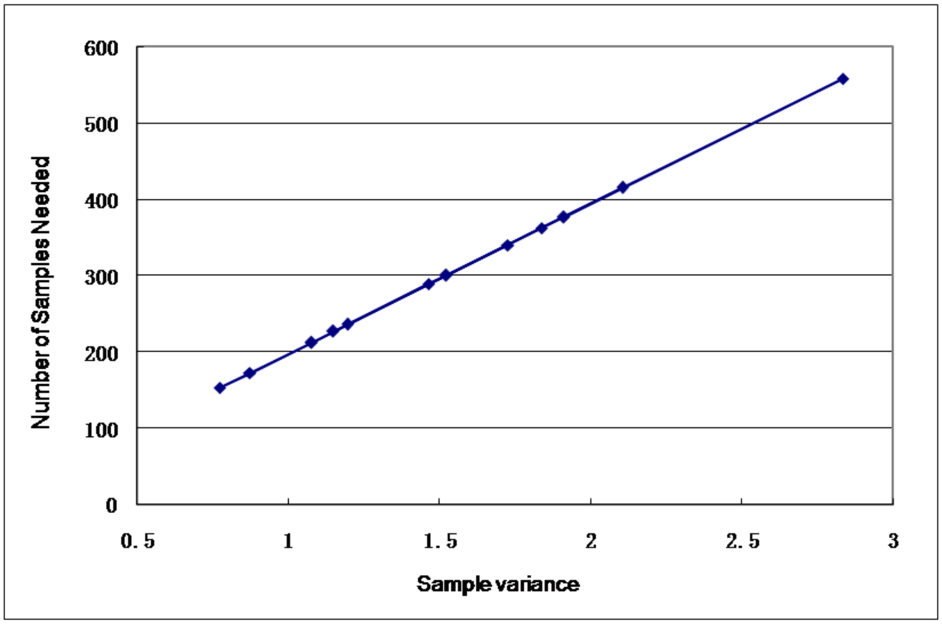


Figure 4.2: Number of samples needed vs. sample variance

equal to 0.01 and at least 95% confidence level, we would select P-8 as the best evacuation policy, since if P-8 is not the true best policy, its mortality differs from the true best by at most 0.01 at the 95% confidence level.

4.4.2 KN procedure

The KN procedure developed by Kim and Nelson (2001) [109] is a fully-sequential R&S procedure. Similar to the Rinott procedure, KN requires δ, $P^* = 1-\alpha$, and n_0 as its input parameters.

During its initialization stage the KN procedure observes n_0 data from each alternative, and uses those data to estimate the sample variances of the differences between the various systems ($S^2_{i\ell}$). Then the KN procedure takes more samples from the more promising systems and in the following stages eliminates those systems that are confirmed as inferiors; such a screening process continues until only one alternative is left.

As a fully sequential procedure, the KN procedure has more opportunities to discard inferior systems, which might not be detected by 2- or 3-stage procedures (e.g., the Rinott procedure) until the final stage. Thus, the KN procedure is expected to be more efficient in the sense that fewer observations and less computer time are needed to find the best.

The full KN procedure [103] is transcribed below (it is assumed that the best alternative refers to the one with the largest sample mean)

Setup: Select the overall desired PCS $1-\alpha$, indifference-zone parameter $\delta > 0$ and common first-stage sample size $n_0 \geq 2$. Set

$$\eta = \frac{1}{2}\left[\left(\frac{2\alpha}{k-1}\right)^{-2/(n_0-1)} - 1\right]$$

Initialization: Let $I = \{1, 2, \ldots, k\}$ be the best of systems still in contention, and let $h^2 = 2\eta(n_0-1)$.

Obtain n_0 outputs $X_{ij} (j = 1, 2, \ldots, n_0)$ from each system $i (i = 1, 2, \ldots, k)$ and let $\bar{X}_i(n_0) = n_0^{-1}\sum_{j=1}^{n_0} X_{ij}$ denote the sample mean of the first n_0 outputs from system i.

For all $i \neq l$ calculate

$$S_{il}^2 = \frac{1}{n_0-1}\sum_{j=1}^{n_0}\left(X_{ij} - X_{lj} - [\bar{X}_i(n_0) - \bar{X}_l(n_0)]\right)^2$$

the sample variance of the difference between system i and l. Set $r = n_0$.

Screening: Set $I^{old} = I$. Let

$$I = \left\{i : i \in I^{old} \text{ and } \bar{X}_i(r) \geq \bar{X}_l(r) - W_{il}(r), \forall l \in I^{old}, l \neq i\right\}$$

where

$$W_{il}(r) = \max\left\{0, \frac{\delta}{2r}\left(\frac{h^2 S_{il}^2}{\delta^2} - r\right)\right\}$$

Stopping Rule: If $|I| = 1$, then stop and select the system whose index is in I as the best.

Otherwise, take one additional output $X_{i,r+1}$ from each system $i \in I$, set $r = r+1$ and go to **Screening**.

Similar to the Rinott procedure, we choose the input parameters for the KN procedure as α = 0.05, δ = 0.01, n_0 = 10. And the R&S results are shown in Table 4.4.

Table 4.4: R&S results by KN procedure

Stage	**Number of rounds**	**Remaining Policy Set (I)**
Initialization	$r = 1 \sim 10$	$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$
Screening	$r = 11$	$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$
	$r = 12 \sim 16$	$\{1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$
	$r = 17$	$\{1, 4, 6, 7, 8, 9, 10, 11, 12\}$
	$r = 18$	$\{4, 6, 7, 8, 9, 10, 11, 12\}$
	$r = 19 \sim 30$	$\{6, 7, 8, 9, 10, 11, 12\}$
	$r = 31 \sim 96$	$\{6, 7, 9, 10, 11, 12\}$
	$r = 97 \sim 102$	$\{7, 9, 10, 11, 12\}$
	$r = 103 \sim 122$	$\{7, 9, 10, 12\}$
	$r = 123 \sim 156$	$\{7, 9, 10\}$
	$r = 157 \sim 188$	$\{9, 10\}$
Stopping	$r = 189$	$\{9\}$

Table 4.4 clearly shows how those non-dominant policies are eliminated gradually from the remaining set. It should be noticed that "a non-dominant policy" is not equivalent to "a policy with worse performance measure". A policy gets eliminated because at some time point there is strong statistical evidence indicating that the policy could not be the best. In other words, with enough statistical confidence, it is believed that there is at least one other policy better than this one, so that it is safe to eliminate this policy from the candidate set.

For example, the policy selected by the Rinott procedure – P-8 – is eliminated after 30 rounds, which means at that time, the KN procedure has collected enough evidence suggesting that P-8 is dominated by one or more other policies so that it is not necessary to keep it in the candidate set.

The final selection of the KN procedure is P-9, which can be described: given an indifference-zone with width equal to 0.01 and at least 95% confidence level, we believe that the P-9 is the policy that leads to the minimal mortality among 12 alternatives; if it is not, then its mortality differs from the true best by at most 0.01 at the 95% confidence level.

4.5 DISCUSSION

Although the final selections of the Rinott procedure and the KN procedure are different, it does not suggest a contradiction due to the existence of the indifferent-zone. From Table 4.3, the three lowest values of mortality are achieved by P-8, P-9 and P-7, which are 0.473, 0.478 and 0.479 respectively. Their pair-differences are all less than the width of indifference-zone $\delta = 0.01$. In others words, according to the definition of indifference-zone formulation, experimenters are "indifferent" to these three alternatives since all of them are within δ-distance to the best. As long as the final choice is made from these three alternatives, it is correct.

Table 4.5 compares the number of samples required by the Rinott and the KN procedures. It is obvious that the KN procedure is more efficient than the Rinott procedure because those inferior alternatives can be screened out in a timely manner in the KN procedure.

Table 4.5: Comparison of the number of samples required by Rinott and KN

Procedure	Rinott	KN
Number of Samples	3640	961

In conclusion, the KN and Rinott procedures are both good statistical methods for selecting a single best alternative. They provide a statistical guarantee for the correctness of selection and help determine the number of replications needed. As a fully sequential procedure, the KN procedure is more efficient than the two-stage Rinott procedure since it can discard inferior systems effectively.

However, in our study, incident managers may also want an approach that can select a subset containing all alternatives that are "close enough" to the best with a pre-specified statistical confidence level, so that they can choose their final decision from the "best-subset".

Existing subset selection methods [161, 162, 118, 163, 164, 119] are incapable for this requirement. They are either unable to provide guarantee for the overall performance of selected subset (such as [118, 163, 119], they only claim that their selection subset contains the best but do not claim that all alternatives in the subset are "good"), or resort to some impractical assumption (such as [161, 162, 164] where they assume a common equal variance over all alternatives, which is very unlikely to be satisfied).

In order to address this problem, we developed a new fully sequential R&S procedure to select the best-subset. Similar to the KN procedure, the new selection procedure can select the best-subset by efficiently screening out the inferior alternatives, and can provide a statistical guarantee for the correctness of selection to satisfy the pre-specified confidence level. The details of the best-subset selection procedure are discussed in the next chapter.

5.0 BEST-SUBSET SELECTION PROCEDURE

5.1 OVERVIEW

The goal of our research is to develop a methodology to help decision makers choose the "best" response policies from different alternatives. For the particular example, "best" refers to the policies leading to the less mortality. Chapter-3 develops an ABS model to simulate the behaviors of a MCI response system under different policies, and Chapter-4 implements two indifference-zone based ranking-and-selection (R&S) procedures (the Rinott and the KN procedures) to demonstrate how to select the best one from twelve evacuation policies by comparing the simulation outputs strategically. The R&S procedures can help analysts appropriately allocate computational resources to reach a statistically guaranteed conclusion in an efficient manner, which is desired for computationally intensive models, such as ABS.

As we have discussed in Section 2.5.1, existing indifference-zone based R&S procedures are designed to select only a single best system instead of a best-subset containing of all alternatives that are close enough to the best. In Section 4.5, we explained why existing subset selection methods are insufficient for best-subset selection (they are either incapable of providing guarantees for the overall performance of the selected subset, or they depend on certain impractical assumptions). However, there is a practical requirement from decision makers to obtain the best-subset, so that they can choose their final decision from the best-subset based on criteria not in the model, such as social or political feasibility.

In order to address this problem, this chapter develops a new fully sequential R&S procedure to select the best-subset. Based on the input from decision makers, this best-subset selection (BSS) procedure can select all desired alternatives, screen out those undesired, and provide a guarantee that the correctness of selection is at or above a pre-specified confidence level.

In order to demonstrate the effectiveness of BSS, we compare the best-subset procedure with the MCB procedure developed by Hsu (1984) [128]. First we demonstrate how simultaneous MCB confidence intervals can be constructed from simulation results, and illustrate a method to select the best-subset based on the information provided by the MCB confidence intervals. Then we use BSS to select the best-subset for the same data configuration. The comparison result shows that the MCB-based method is deficient in providing a statistical guarantee to its selection result and is overly conservative on sampling. We argued that the new BSS procedure is the most suitable for selecting the best-subset from a finite number of alternatives.

In the final part of this chapter, we perform a senstivity analysis to analyze the robustness of BSS from different aspects (input parameters, changes of variances and distributions of competitive systems in different regions). The results of these experiments demonstrate the BSS procedure works robustly in selecting the best-subset from a finite number of alternatives.

5.2 LITERATURE REVIEW

Although most existing R&S procedures focus on identifying the best system [121], a few have examined the problem of best-subset selection. According to Chen (2008) [165], the approaches for selecting a subset of good systems can be roughly classified into two categories. One category considers limited computational budgets and tries to maximize the probability of correctness of selection. A typical representative of this category is OCBA-*m*, which extends the OCBA (Optimal Computing Budget Allocation) procedure by Chen et al. (2000) [105] with a goal to maximize the probability of correctly selecting the top-*m* systems with a given computing budget.

Another category of approaches is designed to select a restricted subset, which attempts to exclude populations that deviate more than a specified indifference-zone from the best. These approaches provide a statistical guarantee of correctness of selection and are more efficient relative to any computing-budget constraint.

In one example of the second category, Koening and Law (1985) [120] developed a two-stage indifference-zone procedure to select a subset of size m containing the v best of k systems; where $(1 \leq v \leq m < k)$. If $m = v = 1$, then the problem is to choose the best system. When $m > v = 1$, they are interested in choosing a subset of size m containing the best. If $m = v > 1$, they are interested in choosing the m best systems.

Sullivan and Wilson (1989) [119] developed a two-stage restricted subset selection procedure that determines a subset of maximum size m containing at least one system that is within a pre-specified distance to the best. Chen (2009) [121] proposed a heuristic two-stage selection procedure (Enhanced Two-Stage Selection procedure) to select a subset of size m containing at least c of the v best of k normal populations with unknown means and unknown variances.

Since our target is to select the "best" response policies from a number of alternatives that lead to the minimal mortality, in order to fulfill this goal, we need a method that satisfies the following requirements:

1. The method can select an unknown best from different alternatives based on a given criteria of measurement;
2. The method can select all alternatives that behave almost "as well as" the best one (which can be implemented by indicating a indifference-zone parameter), and discard any alternatives that are "worse enough" compared to the best one;
3. The method should be able to provide a statistical guarantee to the correctness of selection.

After a cautious investigation of the existing methods, we found that the first class of methods (Bayesian-based) can not provide statistical guarantee about the correctness of selection, they only try to maximize the posterior probability of correct selection under given simulation budgets. And the second category (restricted subset selection) requires too many input parameters which are usually difficult to justify in applications. For example, the Enhanced Two-stage selection procedure by Chen (2009) [121] requires four input parameters (m, c, v, k), however, sometimes decision makers are also not clear beforehand about how large the selection set should be (m) and how many best alternatives the set should contains (c, v).

The inadequacy of existing methods motivates us to develop a new best-subset selection procedure. According to our requirements, the new procedure should be able to select all good enough alternatives, discard those inferior ones and can provide a statistical guarantee about its selection (which excludes Bayesian approaches). According to Osogami (2009) [116], the KN procedure is the most efficient algorithm in terms of the number of samples needed, so we will extend the KN procedure to enable it to select the "best-subset" in an efficient way by screening out the obviously inferior alternatives in its selection stages.

5.3 PROBLEM FORMULATION

In this section, we formulate the best subset-selection problem and define the notation. Assuming that there are in total $k \geq 2$ competing simulation systems, let X_{ij} be a univariate real-valued output data from replication (or batch) j of system i, and the performance measures of different systems are defined as $x_i = \mathbf{E}[X_{ij}]$ $(i = 1, \ldots, k)$. We will assume that a larger mean is better, and we let $x_{[1]} \geq x_{[2]} \geq \ldots \geq x_{[k]}$, where system [1] is the best system (unknown to us) .

Our problem is to find the best-subset I; here we refer to the best system as the system with the largest mean, and any system whose performance measure is within λ-distance to the best will be considered as item belonging to the best-subset, so we can define I as

$$I = \left\{ i : x_i \in \left[\max_{i=1,\ldots,k} x_i - \lambda, \ \max_{i=1,\ldots,k} x_i \right] \right\}$$

For solving this problem, we assume that , $X_{ij} \sim \mathbf{N}\left(x_i, \sigma_i^2\right)$ $(i = 1, 2, \ldots, k)$, that is X_{ij}'s are distributed as normal distributions with mean of x_i and variance of σ_i^2.

It should be mentioned that this assumption is not restrictive since it requires neither common variance nor independent sampling (which implies that our selection procedure allows using common random numbers (CRN) to increase the precision when comparing two or more alternative configurations by simulation). In addition, the assumption of normality is generally plausible when the basic observations of system performance are either within-replication averages (from a transient or steady-state simulation) or batch means with a large batch size (from a steady-state simulation). Although the non-normality of basic observations may be problematic, Kim and Nelson (2001) [109] show that fully sequential R&S procedures tend to be robust to non-normality. In addition, any non-normality can be mitigated by using batches of non-normal data as basic observations (as in Kim and Nelson (2001) [109]).

For stochastic systems, it is not always possible to guarantee that we select the best-subset which contains all systems that are within λ-distance to the best. Instead, we apply the idea of indifference-zone again to the λ-boundary to find a set of best systems. Specifically, we adopt a similar approach to that used in Andradóttir, Goldsman, and Kim (2005) [139] and Andradóttir and Kim (2010) [140] by asking a decision maker to specify a range $[\lambda^-, \lambda^+]$ around the λ-boundary such that $\lambda^+ > \lambda^-$. Then three regions can be defined:

- $x_B - x_i \leq \lambda^-$: This is the definitely best region (S_B). Any system in this range should be retained in the final best-subset. We call any system inside this range as "desired".
- $\lambda^- < x_B - x_i \leq \lambda^+$: This is the transition region (S_T). For any system inside this region, it does not matter whether it is selected into the final best-subset or not. In other words, it is all right to exclude a system in this region from the best-subset (even if $x_B - x_i \leq \lambda$), or to accept a system into the best-subset as long as the system i is within this region. Therefore we call any system inside this region as an "acceptable" system.
- $\lambda^+ < x_B - x_i$: This is the elimination region (S_E). Any system in this region should be screened out in the screen phase and should not be contained in the final best-subset. We use the term "undesired" to refer to any system in this region.

Figure 5.1 provides a demonstration of the division of three regions.

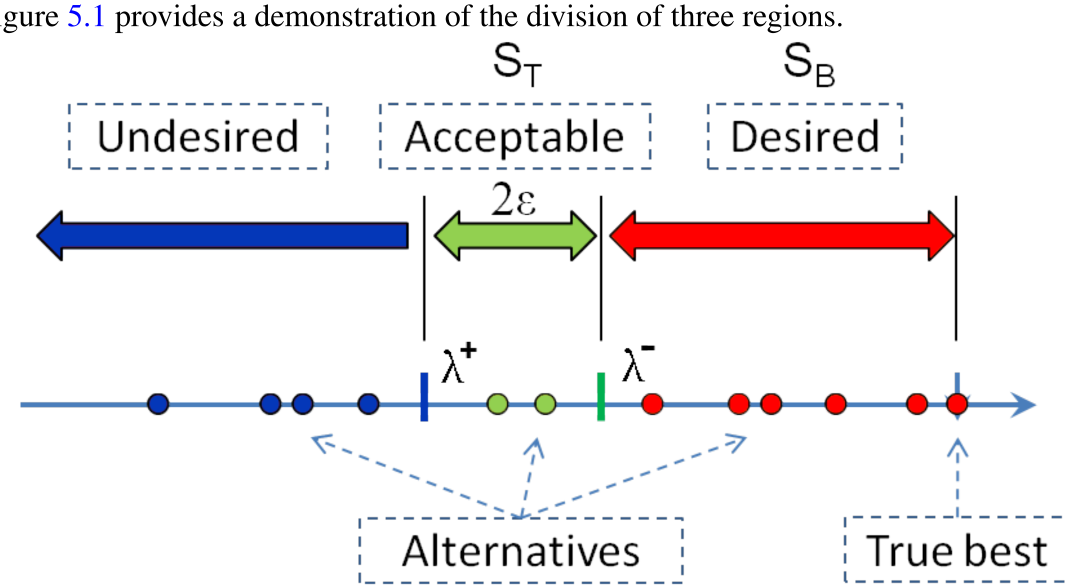


Figure 5.1: Three regions (Desired, Acceptable and Undesired)

For convenience, we choose the parameters λ and ε so that $\lambda = (\lambda^- + \lambda^+)/2$, and define $\varepsilon = (\lambda^+ - \lambda^-)/2$. Essentially, λ is a target value that behaves as a cutoff point between desired and undesired systems, and ε is the level of precision to the specification of λ, that is, ε specifies how much we are willing to be off from λ. In other words, ε defines an indifference-zone around λ and plays a similar role as δ dened in Kim and Nelson (2001) [109] and Kim and Nelson (2006) [114].

5.4 BEST-SUBSET SELECTION (BSS) PROCEDURE

In this section, we present a procedure that eliminates all undesired systems and to return a resultant set that contains all the desired systems plus some or none of the acceptable systems. The complete proof of the statistical validity of the procedure can be found in Section 5.5.

Setup: Select the overall desired probability of correct selection (PCS) (confident level) $P^* = 1-\alpha$ $(0 < P^* < 1)$, boundary parameter $\lambda > 0$ and common first-stage sample size $n_0 \geq 2$. Choose a small value for the indifference-zone parameter ε $(\varepsilon > 0)$, which indicates the half-width of the transition region. Calculate η as described below:

$$g(\eta) \equiv \sum_{\ell=1}^{c} (-1)^{\ell+1} \left(1 - \frac{1}{2}\mathscr{I}(\ell = c)\right)\left(1 + \frac{2\eta(2c-\ell)\ell}{c}\right)^{-(n_0-1)/2} = \frac{2\alpha}{k(k-1)}$$

where $\mathscr{I}$ is the indicator function. In the special case that $c = 1$, we have the closed-form solution

$$\eta = \frac{1}{2}\left[\left(\frac{4\alpha}{k(k-1)}\right)^{-2/(n_0-1)} - 1\right]$$

Initialization: Let $I = \{1, 2, \ldots, k\}$ be the set of systems in contention, N be the set of best systems, and $h^2 = 2c\eta(n_0 - 1)$.

Obtain n_0 outputs X_{ij} $(j = 1, 2, \ldots, n_0)$ from each system i $(i = 1, 2, \ldots, k)$ and let $\bar{X}_i(n_0) = (\sum_{j=1}^{n_0} X_{ij})/n_0$ denote the sample mean of the first n_0 outputs from system i.

For all $i \neq \ell$ calculate the estimated sample variance of the pair difference between system i and ℓ.

$$S_{i\ell}^2 = \frac{1}{n_0-1}\sum_{j=1}^{n_0}\left(X_{ij}-X_{\ell j}-[\bar{X}_i(n_0)-\bar{X}_\ell(n_0)]\right)^2$$

Set the observation counter $r = n_0$ and go to **Screening**.

Screening: Define $\bar{Y}_{i\ell}(r) = \bar{X}_i(r) - \bar{X}_\ell(r)$. For each system combination (i,ℓ) (where $i \in I, \ell \in I, i \neq \ell$), if $\bar{Y}_{i\ell}(r) - \lambda \geq +R_{i\ell}(r) \quad (i \in I, \ell \in I)$, then eliminate system ℓ from set I, where

$$R_{i\ell}(r) = \max\left\{0, \frac{\varepsilon}{2cr}\left(\frac{h^2 S_{i\ell}^2}{\varepsilon^2} - r\right)\right\}$$

Stopping Rule: If $\forall i \in I$ and $\forall \ell \in I$, $\bar{Y}_{i\ell}(r) - \lambda \leq -R_{i\ell}(r)$, then stop and return I as the best subset N. Otherwise, take one additional output $X_{i,r+1}$ from each system $i \in I$, set $r = r+1$ and go to **Screening**.

From the experimental result shown in Section 5.6, we can observe that the procedure returns subset N containing all desired systems, plus some acceptable ones, and without any undesired systems, with very high probability of correct selection (PCS).

5.5 STATISTICAL VALIDITY PROOF

The basic idea of the fully sequential best subset selection procedure is to approximate the sum of differences between two systems as a Brownian motion process and use a triangular continuation region to determine the stopping time of the selection process. To prove the validity of the procedure, we need the following lemmas from Fabian (1974) [166] and Jennison et al (1980) [167]:

Lemma 1 (Fabian, 1974). Consider a standard Brownian motion process with drift $\mathscr{W}(t,\Delta)$ with $\Delta > 0$ and $t \geq 0$. Let $H(t) = a - \gamma t$ for some $a > 0$ and $\gamma \geq 0$. Let $H(t)$ denote the interval $(-h(t), h(t))$ (so that $h(t) = \emptyset$ when $-h(t) \geq h(t)$), and let $T = \min\{n : \mathscr{W}(t,\Delta) \notin H(t)\}$ be the first time $\mathscr{W}(t,\Delta)$ does not fall in the triangular continuation region defined by $(t, H(t))$. Finally, let $\mathscr{E}$ be the event $\{\mathscr{W}(T,\Delta) \leq -h(T)$ and $H(T) \neq \emptyset$, or $\mathscr{W}(T,\Delta) \leq 0$ and $H(T) = \emptyset\}$. If $\gamma = \Delta/(2c)$ for any positive integer c, then

$$\Pr\{\mathscr{E}\} = \sum_{l=1}^{c}(-1)^{\ell+1}\left(1-\frac{1}{2}\mathscr{I}(\ell=c)\right)\exp\{-2a\gamma(2c-\ell)\ell\}$$

Remark 1: In our proof that the fully sequential procedure provides the stated correct selection guarantee, the event $\mathscr{E}$ will correspond to an incorrect selection (could be incorrectly eliminating a "good alternative" from the "best-subset" or incorrectly retaining a "bad system" in the final set).

Lemma 2 (Jennison et al, 1980). Suppose that a continuation region $H(t)$ is $(-h(t),h(t))$ given by a non-negative function $h(t), t \geq 0$. Consider two processes: a continuous process $\{\mathscr{W}(T,\Delta), t \geq 0\}$ with $\Delta > 0$, and a discrete process obtained by observing $\mathscr{W}(t,\Delta)$ at a random, increasing sequence of times $\{t_i : i = 1,2,\ldots\}$ taking values in a given countable set. Let $T_C = \inf\{t > 0 : \mathscr{W}(t,\Delta) \notin H(t)\}$ and $T_D = \inf\{t_i : \mathscr{W}(t_i,\Delta) \notin H(t_i)\}$, and assume that $T_D < \infty$ almost surely. Note that $T_D \geq T_C$. The error probabilities are

$$\Pr\{\mathscr{E}_C\} \equiv \Pr\{\mathscr{W}(T_C,\Delta) \leq -h(T_C)\} = \Pr\{\mathscr{W}(T_C,\Delta) < 0\}$$

$$\Pr\{\mathscr{E}_D\} \equiv \Pr\{\mathscr{W}(T_D,\Delta) \leq -h(T_D)\} = \Pr\{\mathscr{W}(T_D,\Delta) < 0\}$$

Consider an outcome $\{(b(t);t \geq 0),\{t_i\}\}$, where $b(t)$ is the path of a Brownian motion. Assume that the conditional distribution of $\{t_i\}$ given $\mathscr{W}(t,\Delta) = b(t), \forall t \geq 0$, is the same as the conditional distribution of $\{t_i\}$ given $\mathscr{W}(t,\Delta) = -b(t), t \geq 0$. Under these conditions, $\Pr\{\mathscr{E}_D\} \leq \Pr\{\mathscr{E}_C\}$.

Remark2: Lemma 1 gives the probability of an incorrect selection about the sign of the drift Δ for a continuous $\mathscr{W}(t,\Delta)$. When the observations are IID normally distributed with mean Δ and variance one, the distributions of the partial sums of the observations match that of $\mathscr{W}(t,\Delta)$ at each integer point. Lemma 2 states that under very general conditions, the probability of an incorrect selection does not increase when the Brownian motion process is observed at discrete times rather than continuously. Therefore, procedures designed for $\mathscr{W}(t,\Delta)$ provide an upper bound on the probability of an incorrect selection for a corresponding discrete process.

THEOREM 1: Assume that $X_{ij},\ j = 1,2,\ldots,k$ are normally distributed, the best subset selection procedure guarantees

$$\Pr\{\mathbf{CS}\} = \Pr\{S_B \subseteq I \subseteq (S_B \cup S_T)\} \geq 1-\alpha$$

PROOF:

We begin by considering the cases of only two systems (i and ℓ) existing, define $y_{i\ell} \equiv x_i - x_\ell$.

Let $0 < \beta < 1$ and select η such that $g(\eta) = \beta$. For a general output process $\mathbf{G}_{i\ell} = \{G_{i\ell,j}, j = 1, 2, \ldots\}$, let

$$T_{\mathbf{G}_{i\ell}} = \min\left\{ r : r \geq n_0 \text{ and } -R(r) < \sum_{j=1}^{r} G_{i\ell,j} < +R(r) \text{ is violated} \right\}$$

Therefore, $T_{\mathbf{G}_{i\ell}}$ represents the stage at which $\sum_{j=1}^{r} G_{i\ell,j}$ exists the triangular region defined by $R(r)$ for the first time after n_0. Let $\mathrm{ICD}_{i\ell}$ denote the event that a wrong decision is made, which could be:

- $\mathrm{ICE}_{i\ell}$ – incorrectly eliminate one system during screen when both should be be retained, that is, $|\bar{x}_i - \bar{x}_\ell| \leq (\delta - \varepsilon)$

- $\mathrm{ICK}_{i\ell}$ – incorrectly keep both systems in the final set when one of them should be eliminated, that is, $|\bar{x}_i - \bar{x}_\ell| \geq (\delta + \varepsilon)$

We consider three cases.

First consider the case where the difference between system i and ℓ is smaller than δ significantly, that is, $y_{i\ell} \leq \delta - \varepsilon$. Under such a scenario, the correct decision is to keep both system i and ℓ, and the incorrect decision is eliminating system ℓ, that is

$$
\begin{aligned}
& \Pr\{\mathrm{ICD}_{i\ell} \mid y_{i\ell} \le \delta - \varepsilon\} \\
= \;& \Pr\{\text{system } \ell \text{ is eliminated}\} \\
= \;& \Pr\left\{\bar{Y}_{i\ell}(T_{\mathbf{Y}_{i\ell}}) - \delta \ge +R_{i\ell}(T_{\mathbf{Y}_{i\ell}})\right\} \\
= \;& \Pr\left\{\sum_{j=1}^{T_{\mathbf{Y}_{i\ell}}} (Y_{i\ell,j} - \delta) \ge +R_{i\ell}(T_{\mathbf{Y}_{i\ell}}) \cdot T_{\mathbf{Y}_{i\ell}}\right\} \\
= \;& \Pr\left\{\sum_{j=1}^{T_{\mathbf{Y}_{i\ell}}} (Y_{i\ell,j} - \delta) \ge \max\left\{0, \frac{\varepsilon}{2c}\left(\frac{h^2 S_{i\ell}^2}{\varepsilon^2} - T_{\mathbf{Y}_{i\ell}}\right)\right\}\right\} \\
= \;& \Pr\left\{\sum_{j=1}^{T_{\mathbf{Y}_{i\ell}}} (\delta - Y_{i\ell,j}) \le \min\left\{0, -\frac{h^2 S_{i\ell}^2}{2c\varepsilon} + \frac{\varepsilon\, T_{\mathbf{Y}_{i\ell}}}{2c}\right\}\right\} \\
= \;& \mathrm{E}\left[\Pr\left\{\sum_{j=1}^{T_{\mathbf{Y}_{i\ell}}} (\delta - Y_{i\ell,j}) \le \min\left\{0, -\frac{h^2 S_{i\ell}^2}{2c\varepsilon} + \frac{\varepsilon\, T_{\mathbf{Y}_{i\ell}}}{2c}\right\} \middle| S_{i\ell}^2\right\}\right]
\end{aligned}
$$

Now, we define $Z_{i\ell,j} \equiv (\delta - Y_{i\ell,j}) - (\delta - \varepsilon - y_{i\ell})$, where $y_{i\ell} = \mathrm{E}[Y_{i\ell,j}]$. Notice that $0 \le \delta - \varepsilon - y_{i\ell}$, and hence $Z_{i\ell,j} \le (\delta - Y_{i\ell,j})$ due to the assumption that the difference between system i and ℓ is smaller than δ significantly. This implies that $\sum_{j=1}^{r} Z_{i\ell,j}$ is more likely to exit a given continuation region through a lower boundary than $\sum_{j=1}^{r}(\delta - Y_{i\ell,j})$. Therefore,

$$
\Pr\{\mathrm{ICD}_{i\ell} \mid y_{i\ell} \le \delta - \varepsilon\} \le \mathrm{E}\left[\Pr\left\{\sum_{j=1}^{T_{\mathbf{Z}_{i\ell}}} \frac{Z_{i\ell,j}}{\sigma_{i\ell}} \le \min\left\{0, -\frac{h^2 S_{i\ell}^2}{2c\varepsilon\sigma_{i\ell}} + \frac{\varepsilon\, T_{\mathbf{Z}_{i\ell}}}{2c\sigma_{i\ell}}\right\} \middle| S_{i\ell}^2\right\}\right]
$$

where $\sigma_{i\ell}^2 = \mathrm{Var}(Y_{i\ell}) = \mathrm{Var}(Z_{i\ell})$. Notice that $Z_{i\ell,j}/\sigma_{i\ell}$ are IID $\mathrm{N}(\Delta, 1)$ with $\Delta = \varepsilon/\sigma_{i\ell}$, and $h^2 = 2c\eta(n_0 - 1)$. Let

$$
a = \frac{h^2 S_{i\ell}^2}{2c\varepsilon\sigma_{i\ell}} = \frac{\eta(n_0 - 1) S_{i\ell}^2}{\varepsilon\sigma_{i\ell}} > 0
$$

and $\gamma = \varepsilon/(2c\sigma_{i\ell}) = \Delta/(2c)$. The sum of $Z_{i\ell,j},\ j = 1, 2, \ldots, n_0$, is independent of $S_{i\ell}^2$, the sample variance of $Y_{i\ell,j},\ j = 1, 2, \ldots, n_0$, and the observations we take after n_0 do no depend on $S_{i\ell}^2$ as we assume that the $Y_{i\ell,j}$ are IID; that is, the infinite sample path after n_0 does not depend on $S_{i\ell}^2$. Also, notice that the distribution of $\sum_{j=1}^{r} Z_{i\ell,j}/\sigma_{i\ell}$ is identical to that of $\mathscr{W}(t, \Delta)$ for $t = r \in \{n_0, n_0 + 1, \ldots\}$. Then, by Lemma 1 and 2,

$$
\begin{aligned}
& \Pr\{\mathrm{ICD}_{i\ell} \mid y_{i\ell} \leq \delta - \varepsilon\} \leq \mathrm{E}\left[\Pr\{\mathscr{W}(t,\Delta) < 0\} \middle| S_{i\ell}^2\right] \\
= \; & \mathrm{E}\left[\sum_{\ell=1}^{c} (-1)^{\ell+1}\left(1 - \frac{1}{2}\mathscr{I}(\ell = c)\right) \times \exp\{-2a\gamma(2c-\ell)\ell\}\right] \\
= \; & \mathrm{E}\left[\sum_{\ell=1}^{c} (-1)^{\ell+1}\left(1 - \frac{1}{2}\mathscr{I}(\ell = c)\right) \times \exp\left\{-2 \cdot \frac{\eta(n_0-1)S_{i\ell}^2}{\varepsilon\sigma_{i\ell}} \cdot \frac{\varepsilon}{(2c\sigma_{i\ell})} \cdot (2c-\ell)\ell\right\}\right] \\
= \; & \mathrm{E}\left[\sum_{\ell=1}^{c} (-1)^{\ell+1}\left(1 - \frac{1}{2}\mathscr{I}(\ell = c)\right) \times \exp\left\{-\frac{\eta(n_0-1)S_{i\ell}^2}{c\sigma_{i\ell}^2}(2c-\ell)\ell\right\}\right]
\end{aligned}
$$

Now, consider the second case that the difference between system i and ℓ are significantly larger than δ ($y_{i\ell} \geq \delta + \varepsilon$). Then the correct decision is to eliminate system ℓ since it is impossible for system ℓ to belong to the "nearly-best subset". On the opposite, the incorrect decision is to keep both system i and ℓ, that is,

$$
\begin{aligned}
& \Pr\{\mathrm{ICD}_{i\ell} \mid y_{i\ell} \geq \delta + \varepsilon\} \\
= \; & \Pr\{\text{system } \ell \text{ is kept}\} \\
= \; & \Pr\left\{\bar{Y}_{i\ell}(T_{\mathbf{Y}_{i\ell}}) - \delta \leq -R_{i\ell}(T_{\mathbf{Y}_{i\ell}})\right\} \\
= \; & \Pr\left\{\sum_{j=1}^{T_{\mathbf{Y}_{i\ell}}} (Y_{i\ell,j} - \delta) \leq -R_{i\ell}(T_{\mathbf{Y}_{i\ell}}) \cdot T_{\mathbf{Y}_{i\ell}}\right\} \\
= \; & \Pr\left\{\sum_{j=1}^{T_{\mathbf{Y}_{i\ell}}} (Y_{i\ell,j} - \delta) \leq \min\left\{0, \frac{\varepsilon}{2c}\left(-\frac{h^2 S_{i\ell}^2}{\varepsilon^2} + T_{\mathbf{Y}_{i\ell}}\right)\right\}\right\} \\
= \; & \Pr\left\{\sum_{j=1}^{T_{\mathbf{Y}_{i\ell}}} (Y_{i\ell,j} - \delta) \leq \min\left\{0, -\frac{h^2 S_{i\ell}^2}{2c\varepsilon} + \frac{\varepsilon T_{\mathbf{Y}_{i\ell}}}{2c}\right\}\right\}
\end{aligned}
$$

This time, we define $Z'_{i\ell,j} \equiv (Y_{i\ell,j} - \delta) - (y_{i\ell} - \delta - \varepsilon)$. By applying a similar argument to that used for the case $y_{i\ell} \leq \delta - \varepsilon$, we know $Z'_{i\ell,j} \leq Y_{i\ell,j} - \delta$, and

$$
\begin{aligned}
& \Pr\{\mathrm{ICD}_{i\ell} \mid y_{i\ell} \geq \delta+\varepsilon\} \\
\leq\ & \mathrm{E}\left[\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1-\frac{1}{2}\mathscr{I}(\ell=c)\right)\times\exp\left\{-\frac{\eta(n_0-1)S_{i\ell}^2}{c\sigma_{i\ell}^2}(2c-\ell)\ell\right\}\right]
\end{aligned}
$$

So the conditional probability of ICD for both cases is bounded above by the same quantity.

The third case occurs when the difference between system i and ℓ is within the transition region, that is, $\delta-\varepsilon \leq y_{i\ell} \leq \delta+\varepsilon$. In this case, it does not matter whether system i is selected into the nearly-best subset or not, that is, $\Pr\{\mathrm{ICD}_{i\ell} \mid \delta-\varepsilon \leq y_{i\ell} \leq \delta+\varepsilon\} = 0$.

Therefore, in all cases,

$$
\begin{aligned}
\Pr\{\mathrm{ICD}_{i\ell}\} &\leq \mathrm{E}\left[\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1-\frac{1}{2}\mathscr{I}(\ell=c)\right)\times\exp\left\{-\frac{\eta(n_0-1)S_{i\ell}^2}{c\sigma_{i\ell}^2}(2c-\ell)\ell\right\}\right] \\
&= \mathrm{E}\left[\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1-\frac{1}{2}\mathscr{I}(\ell=c)\right)\times\exp\left\{-\frac{\eta(2c-\ell)\ell}{c}\frac{(n_0-1)S_{i\ell}^2}{\sigma_{i\ell}^2}\right\}\right]
\end{aligned}
$$

Notice that $(n_0-1)S_{i\ell}^2/\sigma_{i\ell}^2 \sim \chi_{n_0-1}^2$, a Chi-squared distribution with n_0-1 degree of freedom. To evaluate the expectation in the equitation above, from the moment generating function of a χ_ν^2 random variable, we know that $\mathrm{E}\left[\exp\{t\,\chi_\nu^2\}\right] = (1-2t)^{-\nu/2}$ for $t < 1/2$. Thus, the expected value is

$$
\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1-\frac{1}{2}\mathscr{I}(\ell=c)\right)\times\left(1+\frac{2\eta(2c-\ell)\ell}{c}\right)^{-(n_0-1)/2} = \beta
$$

where the equality follows from the way we choose η.

Thus, we have a bound on the probability of an incorrect decision where there are two systems. Now, consider $k \geq 2$ systems, let ICS be the event that incorrect selection was made during the procedure, which consists of two possible cases, that is, ICS = ICE or ICK.

where

- ICE – incorrect elimination (eliminate one or more "good" systems which should be contained in the final selection set).

- ICK – incorrect keeping (keep one or more "bad" systems in the final selection set which should be eliminated in screening phrase).

Notice that ICE and ICK is mutually exclusive for each pair of comparison, but the probabilities of incorrect decision (ICD) share the same bound, that is, $\Pr\{\text{ICD}\} = \Pr\{\text{ICE}\} = \Pr\{\text{ICK}\}$

Therefore,

$$\Pr\{\text{ICS}\} = \Pr\{\text{ICE or ICK}\} = \Pr\{\text{ICD}\}$$

Now set $\beta = 2\alpha/k(k-1)$, and notice that

$$\Pr\{\text{ICS}\} = \Pr\{\text{ICD}\} \leq \sum_{i\in I,\ell\in I,i\neq\ell} \Pr\{\text{ICD}_{i\ell}\} = \sum_{n=1}^{C_k^2} \frac{2\alpha}{k(k-1)} = \frac{k(k-1)}{2}\left(\frac{2\alpha}{k(k-1)}\right) = \alpha$$

where the first inequality follows from the Bonferroni inequality.

Let CS denote the event that a correct nearly-best subset selection is made when the procedure is applied to all k systems. Then

$$\Pr\{\text{CS}\} = 1 - \Pr\{\text{ICS}\} \geq 1 - \alpha$$

□

5.6 NUMERICAL EXPERIMENTS

In this section we illustrate the performance of the best-subset selection procedure by two numerical experiments.

The first experiment is to demonstrate that the best-subset selection procedure can correctly identify the best-subset with required probability. In the setup stage, we choose the confidence level as $P^* = 1 - \alpha = 0.95$ and take $n_0 = 10$ samples from each alternative. The boundary/indifference-zone parameters are chosen as $(\lambda^-, \lambda^+) = (5, 5.5)$ so that the transition region width $2\varepsilon = \lambda^+ - \lambda^- = 0.5$, which means, any system whose mean value is within 5 unit distance of the true best system should be selected for inclusion ("desired"), and it is also acceptable for the final selection set to include any system whose mean differs from the true best by more than 5 but less than 5.5.

For example, suppose the best system's mean is 100, then all systems whose means are greater or equal to 95 should be selected, and any system whose mean is greater than or equal to 94.5 but less than 95 could be selected; all others (mean less than 94.5) should be excluded from the final selection set.

We use a set of normal random variables X_{ij} with different means and variances to represent the competitive systems ($X_{ij} \sim \mathbf{N}\left(x_i, \sigma_i^2\right)$). The number of total alternative systems $k = 3$. The mean and variance values are shown in Table 5.1. Such a configuration is known as the slippage configuration (SC) since systems 2 and 3 are just at the edge of the desired boundary. According to the region definitions (see Section 5.3) and parameter settings for this case, we know that both systems 2 and 3 should be included in the final set and the selected best-subset should contains all alternatives $\{1,2,3\}$, but it is difficult to make the correct selection.

Table 5.1: Mean and variance configurations for the first experiment (SC)

No.	**1**	**2**	**3**
x_i	100	95	95
σ_i^2	1.0	2.0	4.0

Table 5.2: Replication statistics of the first experiment

Mean (Number of Samples)	**Standard Deviation (Number of Samples)**	Pr$\{\mathbf{CS}\}$
2803.42	1672.78	0.9546

We run 10,000 independent replications. The percentage of correct selection (Pr$\{\mathbf{CS}\}$) and the number of samples required in the replications are summarized in Table 5.2. From Table 5.2 we can see that the percentage of correct selection Pr$\{\mathbf{CS}\} = 0.9546$ which satisfies the requirement of $\mathbf{PCS} \geq 0.95$ very well.

Now let us consider a more complex scenario in the second experiment, where we have $k = 16$ competitive systems ($X_{ij} \sim \mathbf{N}\left(x_i, \sigma_i^2\right)$). The mean and variance configurations are shown in Table 5.3. All other parameters are the same as in the first experiment ($P^* = 1 - \alpha = 0.95$, $n_0 = 10$,

$(\lambda^-, \lambda^+) = (5, 5.5)$, $2\varepsilon = \lambda^+ - \lambda^- = 0.5$). Here we only test configurations with a common variance for all configurations ($\sigma_i^2 = 1.0$).

Table 5.3: Mean and variance configurations for the second experiment

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
x_i	100	95.3	95.2	95.1	94.95	94.9	94.85	94.8	94.75	94.7	94.65	94.6	94.55	94.4	94.3	94.2
σ_i^2	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

For this experiment, we run 30 batches where each batch consists of 500 replications (15,000 in total, independent replications). The benefit of using batches is to enable plotting box-plot of correct selection frequency for each alternative system, as shown in the Figure 5.2. The number of samples required in the replications is summarized in Table 5.4.

Table 5.4: Replication statistics of the second experiment

Mean (Number of Samples)	**Standard Deviation (Number of Samples)**	$\Pr\{\mathbf{CS}\}$
13665.73	4650.98	0.99973

From Figure 5.2 we can see that our best-subset procedure works well. With a very high probability, all desired systems ($x_i \geq 95$) were selected, and no undesired systems were kept in the final selection. For the systems within the transition region, it is obvious that the frequency-of-being-selected drops very quickly as the distance to the best system increases, which means those non-competitive systems can be effectively eliminated during the screening procedures. The resultant percentage of correct selection $\Pr\{\mathbf{CS}\} = 0.99973$ is very high, which may suggest certain conservation existing when handling a large number of alternatives. The possible source of conservation may come from the application of the Bonferroni inequality to control the overall incorrect selection probability by combining pairwise-comparison results together (i.e. the $2\alpha/k(k-1)$ term when calculating η), and possible techniques to reduce this conservatism will be left for future work.

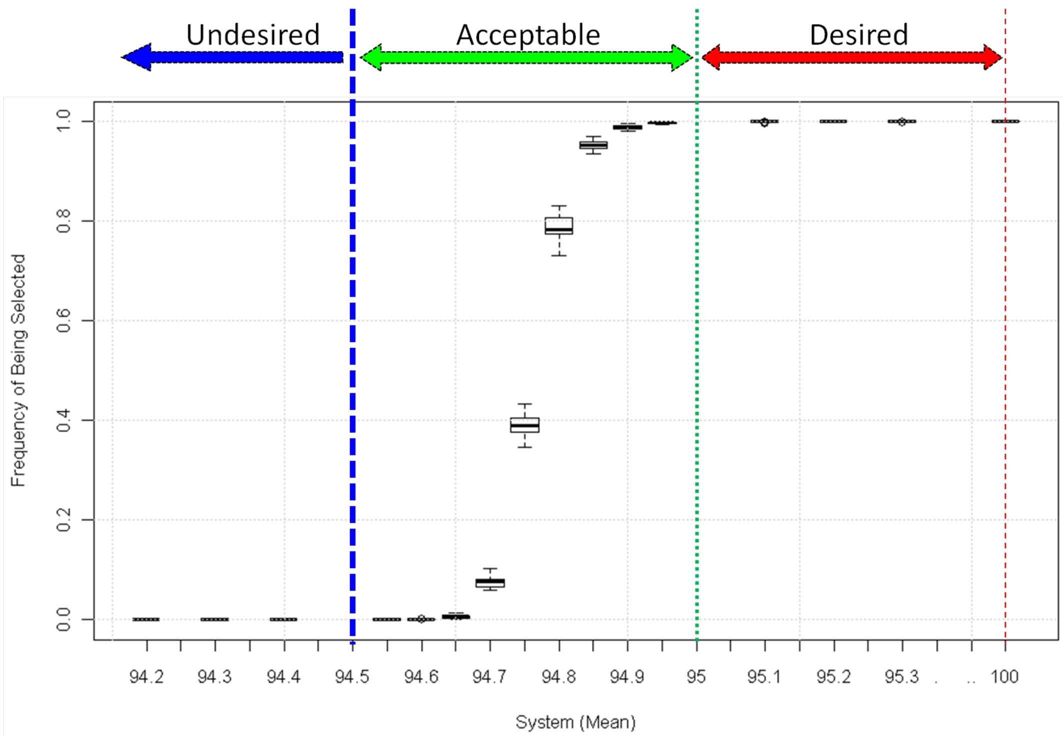


Figure 5.2: Selection frequency of each alternative in the second experiment

5.7 COMPARING BSS WITH MCB

In order to better illustrate the effectiveness of the best-subset selection (BSS) procedure, this section compares BSS with the MCB (Multiple Comparisons with the Best) procedure.

5.7.1 Multiple comparisons with the best (MCB)

MCB procedures are designed to address such a question: how worse are the other alternatives comparing to the unknown best one? The first MCB procedures were developed by Hsu (1984) [128], which is used to provide simultaneous confidence intervals for the difference between the expected performance of each system and the best of the other systems.

For the applications of MCB, Goldsman and Nelson (1990) [131] outlined an MCB procedure for steady-state simulation experiments. They also discussed results on how the batch size can impact the probability of correct selection when using the simulation technique of batch means. Yang and Nelson (1991) [129] and Nelson and Hsu (1993) [130] described modifications to the MCB procedure that incorporate two variance reduction techniques (control variates and CRN) to shorten the length of the confidence intervals for a specified level of confidence. Nelson and Banerjee (2001) [132] present a two-stage MCB procedure that simultaneously achieves several objectives for a given probability of correct selection. For the specific application on the best-subset selection, Nelson and Matejcik (1995) [134] developed two-stage MCB procedures (NM procedure) that provide confidence intervals for the difference between the expected performance of each system and the best of the others.

A typical MCB procedure can be described as follows: Assume that there are k competing systems (treatments), and let $\mu = (\mu_1, \mu_2, \ldots, \mu_k)$ denote the corresponding vector of (unknown) treatment means. Without loss of generality we assume that "the treatment with larger mean is better". Then MCB constructs simultaneous confidence intervals for the parameters $\mu_i - \max_{j \neq i} \mu_j$ for $i = 1, 2, \ldots, k$. These confidence intervals bound the difference between the performance of each system and the best of the others with a pre-specified confidence level. Note that most MCB procedures assume the variances across systems are equal.

To be specific, for k competing systems with means $\mu = (\mu_1, \mu_2, \ldots, \mu_k)$ and a common variance σ^2, the MCB procedure guarantees that

$$P_{\mu,\sigma^2}\left\{\mu_i - \max_{j \neq i} \mu_j \in \left[D_i^-,\ D_i^+\right] \ \ \forall i = 1, 2, \ldots, k\right\} \geq 1 - \alpha$$

where:

$$D_i^- = -(\hat{\mu}_i - \max_{j \neq i} \hat{\mu}_j - d\hat{\sigma}\sqrt{2/r})^-$$

$$D_i^+ = (\hat{\mu}_i - \max_{j \neq i} \hat{\mu}_j + d\hat{\sigma}\sqrt{2/r})^+$$

k – Number of systems;

α – Confidence level;

r – Number of observations for each treatment;

and $d = f(k, \nu, \alpha)$ is a critical point value determined by k, ν, α, where ν is the degree of freedom, $\nu = k(r-1)$. The value of d could be obtained from a table (e.g., Appendix E in [168]) or calculated by a program (e.g., the function *qdunnett()* provided by Chiuzan (2009) [169]).

To better illustrate this, we present a simple MCB example below. The example contains only three competitive systems, which are represented by three normally distributed random variables X_{ij} whose mean and variance values are shown in Table 5.5.

Table 5.5: Means and variances of three competitive systems

No.	**1**	**2**	**3**
x_i	100	95.5	94
σ_i^2	1.0	2.0	4.0

For confidence level $\alpha = 0.05$, we constructed the simultaneous MCB confidence intervals using the MCB procedure. The results are shown in Table 5.6.

From the simultaneous confidence intervals provided by MCB, two things can be inferred: which alternative is the best one (having the largest mean) and how much is the difference between

Table 5.6: MCB Results ($k = 3$, $\alpha = 0.05$)

r	d	$\hat{\sigma}$		**Estimated means & MCB C.I.**		
				estimate	lower	upper
10	2.3334	2.78	1	100.2	1.60	7.40
			2	95.3	-7.80	-2.00
			3	95.7	-7.40	-1.60
				estimate	lower	upper
20	2.2681	1.57	1	100.5	3.24	5.48
			2	96.1	-5.48	-3.24
			3	94.3	-7.24	-5.00
				estimate	lower	upper
50	2.2335	2.80	1	100.0	3.16	5.65
			2	95.5	-5.83	-3.33
			3	95.6	-5.65	-3.16
				estimate	lower	upper
100	2.2121	2.83	1	100.0	3.33	5.10
			2	95.8	-5.10	-3.33
			3	94.1	-6.84	-5.06
				estimate	lower	upper
200	2.2121	2.68	1	100.0	3.82	5.01
			2	95.6	-5.01	-3.82
			3	94.1	-6.44	-5.26
				estimate	lower	upper
500	2.2121	2.69	1	100.0	4.08	4.83
			2	95.6	-4.83	-4.08
			3	93.9	-6.52	-5.77

the performance of each system and the best of the others. For example, for $r = 10$ observations for each alternative, the estimated mean values are $\hat{x}_1$ = 100.2, $\hat{x}_2$ = 95.3, $\hat{x}_3$ = 95.7. And from the confidence intervals for $\mu_i - \max_{j \neq i} \mu_j$, we can infer that system 1 is the best since both its lower bound and upper bound are great than 0, which means that system 1 is between 1.60 and 7.40 better than the best of other systems at the 95% confidence level. In addition, systems 2 and 3 could not be the best, since both their lower and upper bounds are negative. Furthermore, system 2 is worse than the true best by between 2.00 and 7.80, and system 3 is worse than the true best by between 1.60 and 7.40 at the 95% confidence level.

However, the original MCB procedure is not suitable for the best-subset selection problem for two reasons. First, usually the MCB procedure is employed to find the "best" candidates whose confidence intervals overlap with zero (which means their performance is close enough to the unknown best), but neither does it allow one to indicate an indifference-zone beforehand, nor to select the alternatives within a range determined by the indifference-zone and the unknown best. Second, although the simultaneous confidence intervals made by MCB do provide information to infer the "best subset", it is not able to indicate how many replications are needed to select the best-subset with pre-specified confidence level, which is often desired by experimenters in order to control the simulation in an adaptive and efficient manner.

5.7.2 MCB-based method for best-subset selection

As a practical approach for the best-subset selection, the method should be able to:

1. allow experimenters to indicate indifference-zone parameters so that all "close-enough-to-the-best" systems can be selected by the method; and
2. provide the number of replications needed for statistically guaranteed selection result.

The original MCB procedure can not satisfy these two requirements directly, but it does construct simultaneous confidence intervals which provide information about the difference between each system and the (unknown) best of others. In addition, from Table 5.6 we can observe that the widths of the confidence intervals continuously shrink as we increase the number of samples. Based on this, we propose a two-stage MCB-based method for best-subset selection, as described below.

Setup: For k competitive systems, select the overall desired probability of correct selection (PCS) (confident level) $P^* = 1 - \alpha$ $(0 < P^* < 1)$, boundary/indifference-zone parameters (λ^-, λ^+), and a common initial sample size n_0. Calculate the critical point value $d_0 = f(k, \nu_0, \beta)$ using the function *qdunnett()* provided by Chiuzan (2009) [169], where $\nu_0 = k(n_0 - 1)$, $\beta = \alpha/(k-1)$;

Initialization Stage (Stage-0):

1. Take n_0 observations from each of k systems, estimate their means $\hat{\mu}_{i,0}$ and pooled variances $\hat{\sigma}_0^2$; Construct the simultaneous MCB confidence intervals $\left[D_{i,0}^-,\, D_{i,0}^+\right]$ using $\hat{\mu}_{i,0}$, d_0, $\hat{\sigma}_0$, and $r = n_0$
2. Calculate the sample size needed for each system
$$n = \max_{i \in I} \left\{ 2 \Big/ \left(\sqrt{\frac{2}{n_0}} - \frac{\Delta_i}{d_0 \hat{\sigma}_0} \right)^2 \right\}$$
where
$$\Delta_i = \min \left\{ (D_{i,0}^+ + \lambda^-), (-\lambda^+ - D_{i,0}^-) \right\}$$
$$I = \left\{ i : D_{i,0}^+ > (-\lambda^-) \text{ and } D_{i,0}^- < (-\lambda^+) \right\}$$
3. Confirm n by re-calculating $d_1 = f(k, \nu_1, \beta)$, where $\nu_1 = k(n-1)$.
If $d_1 \neq d_0$ then let $n_0 = n$, $d_0 = d_1$ and go back to step 1, otherwise continue to **Screening Stage (Stage-1)**.

Screening Stage (Stage-1): Take n samples from each of the k systems, and construct the simultaneous MCB confidence intervals $\left[D_{i,1}^-,\, D_{i,1}^+\right]$ based on the kn samples. Eliminate any system whose $D_{i,1}^+ < -\lambda^-$, and return the remaining alternatives as the best-subset.

Remark: The basic idea of MCB-based method is to eliminate any system that is significantly worse than the best, where a system is considered as "significantly worse" if its MCB confidence interval is completely outside the boundary defined by the indifferent-zone parameters $(-\lambda^+,\, -\lambda^-)$ (e.g., systems 1 and 2 shown in Figure 5.3). The confidence intervals built in the step 1 of **Initialization Stage** provide basic information about how poor each alternative is when compared with the best. Since the size of initial samples (n_0) is usually a small number, the widths of the resultant confidence intervals are relatively wide. Figure 5.3 displays five possible distributions of those confidence intervals.

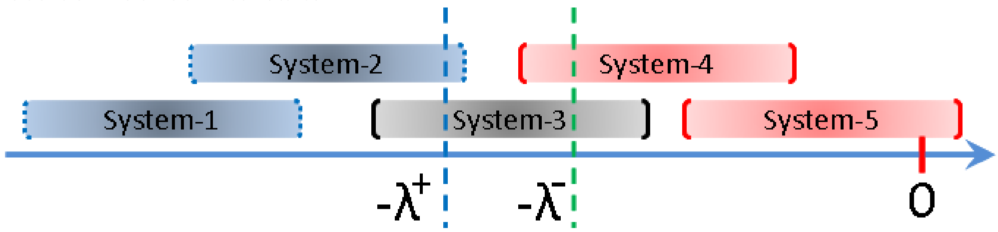


Figure 5.3: Five possible distributions of confidence intervals

From Figure 5.3 we can see that it is safe to eliminate systems 1 and 2 because the upper bounds of their MCB confidence intervals are both less than $-\lambda^-$ (which means that their performance measures could only be located in the undesired or acceptable region). For a similar reason, systems 4 and 5 should be selected into the final best-subset since their lower bounds are both greater than $-\lambda^+$ (which implies that their performance measures are located in the desired or acceptable region).

However, it is difficult to determine whether system 3 should be eliminated or be kept since its MCB confidence interval envelops the range $[-\lambda^+, -\lambda^-]$. In order to make a definitive decision, we have to shrink the MCB confidence interval width for system 3 by additional sampling until one of its boundaries reaches the $-\lambda^-$ or the $-\lambda^+$ boundary. For example, if the upper bound first reaches the $-\lambda^-$ line while the lower bound is still less than $-\lambda^+$ as a result of shrinkage, then system 3 should be eliminated since it will degenerate to system 2 if the shrinkage continues. On the other hand, if the lower bound first touches the $-\lambda^+$ line while the upper bound is greater than $-\lambda^-$ during the shrinkage, system 3 should be kept since it will degenerate to system 4 if the shrinkage continues. With this logic, steps 2 and 3 of the **Initialization Stage** calculate and validate the necessary simple size n that guarantees the widths of MCB confidence intervals of all "system-3-like" alternatives (contained in set I) are narrow (shrunk) enough for experimenters to make judgments easily. Finally, the **Screening Stage** takes extra samples using the calculated sample size and selects the best-subset by eliminating the "significantly worse" alternatives.

It should be noted that since we want to guarantee the overall probability of correct selection for the whole subset, the error probability (α) of MCB is adjusted to $\beta = \alpha/(k-1)$, which follows from the Bonferroni inequality as there are $(k-1)$ total pairs compared.

The effectiveness of the method was tested by an experiment, in which the MCB-based method was employed to select a best-subset from $k = 3$ competitive systems (whose configurations are shown in Table 5.5). In the setup stage, we choose the overall PCS $P^* = 1 - \alpha = 0.95$ ($\alpha = 0.05$), boundary/indifference-zone parameters (λ^-, λ^+) = (5, 5.5), and initial sample size n_0 = 10. Then the initial critical point value was calculated as $d_0 = f(k, \nu_0, \beta)$ = 2.6458, where $\nu_0 = k(n_0 - 1)$ = 27, $\beta = \alpha/(k-1)$ = 0.025.

We ran 5000 independent replications, and compared the best-subset selection results using MCB-based method with the results from the BSS ($k = 3$, $\alpha = 0.05$, $n_0 = 10$, $(\lambda^-,\ \lambda^+) = (5,\ 5.5)$). The comparisons are summarized in Table 5.7.

Table 5.7: Comparison between BSS and MCB (5000 replications for each)

Method	**Number of Samples**		Pr{**CS**}
	Mean	**Standard Deviation**	
MCB-based	3702	57962	87.96%
BSS	1201	652	99.98%

From Table 5.7, we can see that on average the BSS procedure needs only 1201 samples to select the best-subset, and the percentage of correct selection (Pr{**CS**}) is 99.98%. In contrast, the Pr{**CS**} of MCB-based method is only 87.96% with 3702 samples (on average), which is more than three times that for the BSS. The comparison results also imply that MCB-based selection method is not as robust as BSS for its lower Pr{**CS**} and the larger standard deviation of sample size. Therefore, compared with the MCB-based method, BSS procedure is more suitable for selecting the best-subset from a finite number of alternatives.

5.8 ROBUSTNESS ANALYSIS

In this section, we analyze the robustness of the Best-subset Selection procedure with respect to the input parameters (width of indifference-zone, initial samples size), changes of variances, and different distributions of competitive systems.

Throughout, we assume that each competitive system generates samples that are independent and identically distributed according to a normal distribution, where each alternative may have a different mean and/or variance. The upper bound on the overall error probability is set to $\alpha = 0.05$ so that $P^* = 1 - \alpha = 0.95$. For each experiment, 12,000 independent replications are implemented, where each replication comprises one implementation of the BSS.

5.8.1 Input parameters

The experiments performed in this section are used to investigate the impacts due to different input parameters, specifically on two factors: width of the indifference-zone and the initial samples size. The competitive system configurations used in experiments are shown in Table 5.8.

Table 5.8: Competitive system configurations (input parameters)

System i	1	2	3	4	5	6	7	8	9	10
Mean	100	95.3	95.1	94.9	94.8	94.7	94.6	94.4	94.2	94.0
S.D.	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Region	Desired			Acceptable				Undesired		

5.8.1.1 Impact of indifference-zone width The first set of experiments is used to check the impact of varying the width of the indifference-zone ($2\varepsilon = \lambda^+ - \lambda^-$), where (λ^-, λ^+) are the indifference-zone boundaries. The number of samples collected in the initialization stage is $n_0 = 10$, and the experimental results are shown in Table 5.9 and Figure 5.4.

From Table 5.9 and Figure 5.4, we see that narrower indifference-zones correspond to larger sample sizes (in terms of mean and standard deviation) and higher percentage of correct selection ($\Pr\{\mathbf{CS}\}$), which follows intuition. The reason that additional samples are required is because

Table 5.9: Results of different indifference-zone widths

(λ^-, λ^+)	**2ε**	$\Pr\{$**CS**$\}$	**Number of samples**		**Reason of Incorrect Selection**
			Mean	**S.D.**	
(5, 5.5)	0.50	0.9997	5233	1992	4 miss desired
(5, 5.25)	0.25	1.0000	13874	5601	N/A
(5, 5.1)	0.10	1.0000	36843	15547	N/A
(5, 5.05)	0.05	1.0000	66534	25705	N/A

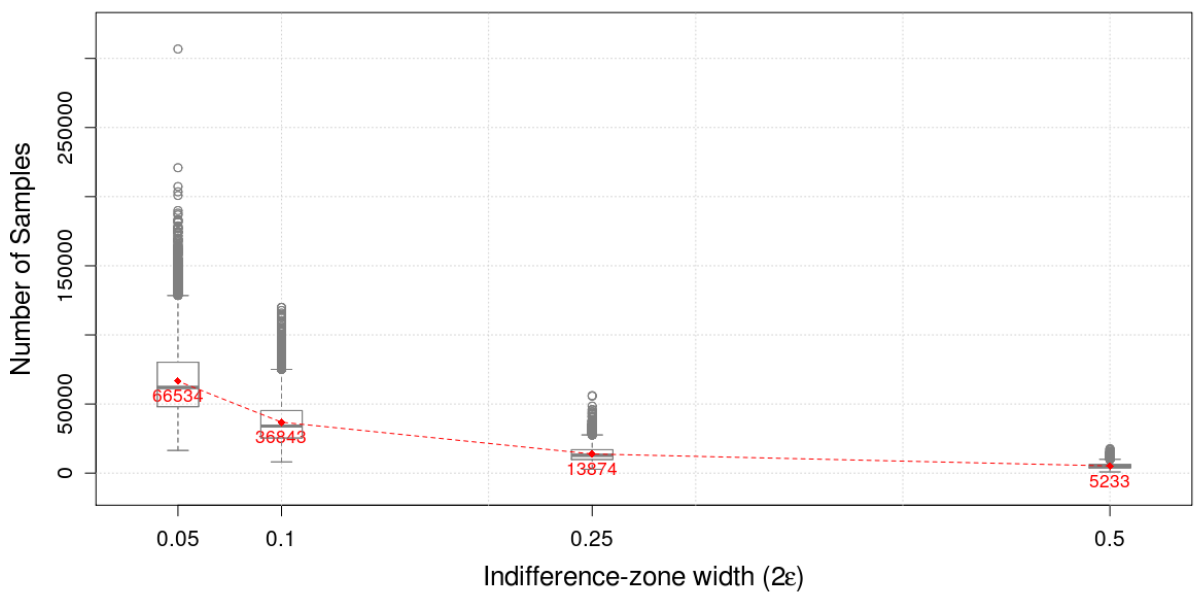


Figure 5.4: Sample sizes with different indifference-zone widths

with the shrinkage of the indifference-zone, it becomes more difficult to confirm that all remaining alternatives are really "desired" or "acceptable". The confidence intervals of the paired-differences among those alternatives must be small enough to provide statistical evidence that all remaining systems are within the region bounded by the λ^+ boundary.

5.8.1.2 Impact of different initial samples size The second set of experiments is to check the impact of different initial sample sizes. The indifference-zone boundaries are chosen to be $(\lambda^-, \lambda^+) = (5, 5.5)$, and the experimental results are shown in Table 5.10 and Figure 5.5.

Table 5.10: Results of different initial sample sizes

$\mathbf{n_0}$	Pr{**CS**}	**Number of samples**		**Reason of Incorrect Selection**
		Mean	**S.D.**	
5	0.9972	18386	9809	34 miss desired
10	0.9997	5233	1992	4 miss desired
20	0.9998	3185	920	2 miss desired, and 1 contains undesired
30	0.9998	2733	690	2 miss desired
50	1.0000	2448	528	N/A
100	1.0000	2322	403	N/A
150	0.9999	2408	353	1 misses desired
200	1.0000	2575	313	N/A
250	1.0000	2804	248	N/A
300	1.0000	3123	155	N/A
400	1.0000	4002	16	N/A

The results shown in Table 5.10 and Figure 5.5 suggest that a certain number of initial samples are important for efficient sampling. If the initial sample size is too small (e.g., n_0 = 5), the total number of samples may be quite large due to the poor estimates of the variance of the difference between alternative pairs. On the other hand, too many initial samples are not conductive to the efficiency of the procedure. Since it is not necessary to have so many samples allocated for variance estimation in the initialization stage, which may suppress the capability of BSS to eliminate the non-competitive alternatives in the screening stages, and thus lead to an increase of total samples,

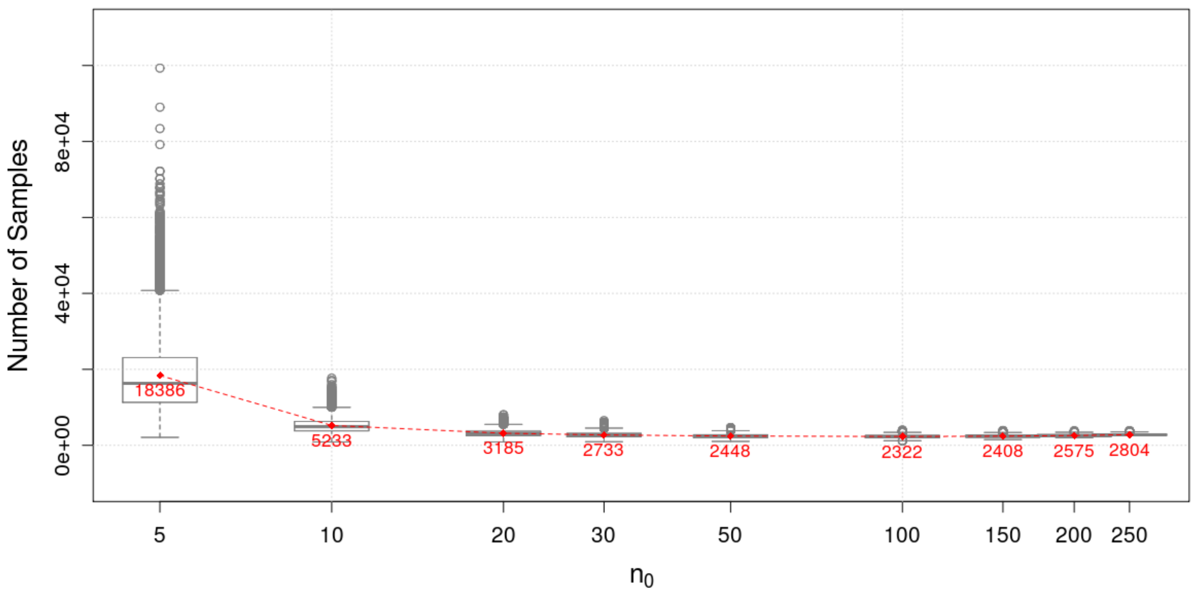


Figure 5.5: Sample sizes with different initial sample sizes

as shown by the cases when $n_0 > 150$. The empirically proper initial samples size should be chosen from the range of $[10, 50]$.

5.8.2 Changing variances

The experiments in this section are designed to investigate the impacts due to the change of variances across alternative systems. Three different scenarios are studied and compared: common variance, increasing variances and decreasing variances respectively. The competitive system configurations are shown in Table 5.11, where the variances of alternatives are represented by their standard deviations (S.D.).

In this example, the indifference-zone boundaries are $(\lambda^-, \lambda^+) = (5, 5.5)$, the initial sample size is $n_0 = 10$, and the experimental results are shown in Table 5.12 and Figure 5.6.

From the results it can be observed that the increasing variances case needs more samples and has a larger standard deviation in terms of sample size compared to the common variance case. In contrast, the decreasing variances case takes less samples and the standard deviation of sample size is also smaller. The reason is because it is easier (with less samples) for the decreasing variance

Table 5.11: Competitive system configurations (changing variances)

System i	1	2	3	4	5	6	7	8	9	10
Mean	100	95.3	95.1	94.9	94.8	94.7	94.6	94.4	94.2	94.0
Common S.D.	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Increasing S.D. ($1.2^{(i-1)}$)	1.0	1.2	1.44	1.728	2.074	2.488	2.986	3.583	4.300	5.160
Decreasing S.D. ($0.8^{(i-1)}$)	1.0	0.8	0.64	0.512	0.410	0.328	0.262	0.210	0.168	0.134
Region	Desired			Acceptable				Undesired		

Table 5.12: Results of different changing variances

Variance	Pr{**CS**}	**Number of samples**		**Reason of Incorrect Selection**
		Mean	**S.D.**	
Common	0.9997	5233	1992	4 miss desired
Increasing	0.9995	23270	6590	6 miss desired
Decreasing	0.9992	2741	1298	8 miss desired, and 2 contain undesired

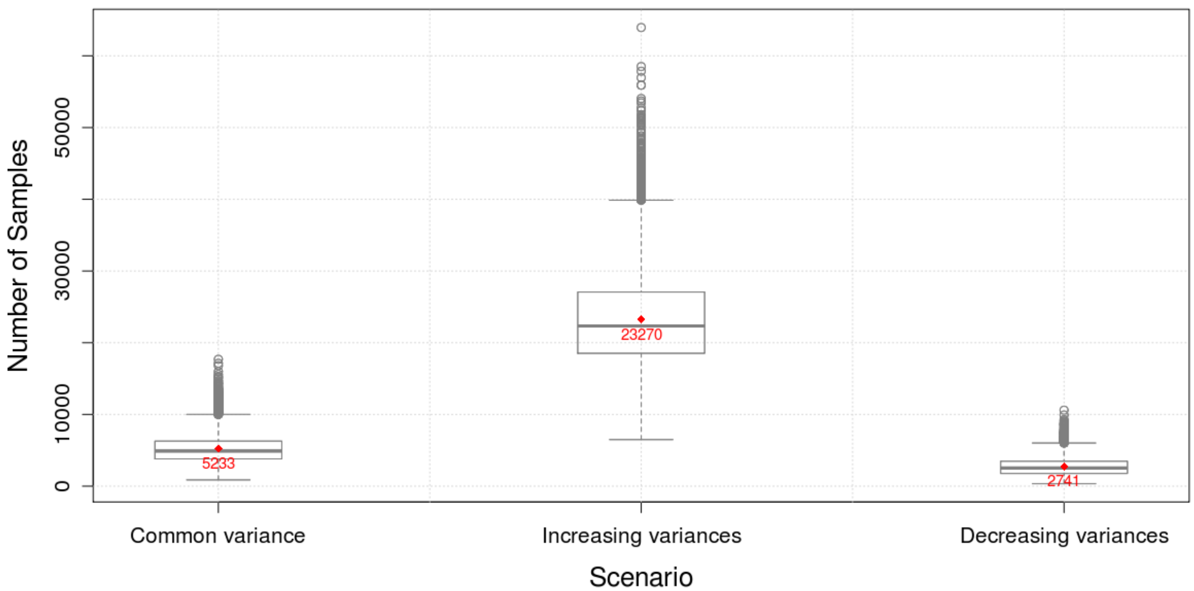


Figure 5.6: Sample sizes with different changing variances

case to obtain a precise variance estimation in the initialization stage, which helps in eliminating non-competitive systems in the screening stages. In addition, smaller variances also make it easy to confirm that the BSS procedure should terminate after all remaining systems are within the λ^+-distance to the best.

5.8.3 Slippage configurations (SCs)

The experiments implemented in this section are to study the behaviors of BSS under the slippage configurations, which is referred to the scenarios where most alternatives locate at (or very close to) the boundary (λ^- or λ^+) except for the one best. The SC's are difficult configurations because all of the other systems are equally close to the best so that it is very difficult to eliminate non-competitive alternatives. The system configurations are shown in Table 5.13.

The BSS parameter settings are similar: $(\lambda^-, \lambda^+) = (5, 5.5)$, $n_0 = 10$, and the experimental results are shown in Table 5.14 and Figure 5.7.

From Table 5.13, we can see that the correction selections of SC-1 and SC-2 should contains all alternatives since all of them are all in the desired region. For SC-3, SC-4, SC-5 and SC-6, all

Table 5.13: Slippage configurations

System i		1	2	3	4	5	6	7	8	9	10
S.D.		1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Mean	**SC-1**	100	95.05	95.05	95.05	95.05	95.05	95.05	95.05	95.05	95.05
	SC-2	100	95	95	95	95	95	95	95	95	95
	SC-3	100	94.95	94.95	94.95	94.95	94.95	94.95	94.95	94.95	94.95
	SC-4	100	94.75	94.75	94.75	94.75	94.75	94.75	94.75	94.75	94.75
	SC-5	100	94.55	94.55	94.55	94.55	94.55	94.55	94.55	94.55	94.55
	SC-6	100	94.5	94.5	94.5	94.5	94.5	94.5	94.5	94.5	94.5
	SC-7	100	94.45	94.45	94.45	94.45	94.45	94.45	94.45	94.45	94.45

Table 5.14: BSS results of slippage configurations

Scenario	Pr{**CS**}	**Number of samples**		**Reason of Incorrect Selection**
		Mean	**S.D.**	
SC-1	0.9929	4324	1385	85 miss desired
SC-2	0.9811	4908	1607	227 miss desired
SC-3	1.0000	5666	1890	N/A
SC-4	1.0000	8273	3058	N/A
SC-5	1.0000	3445	1108	N/A
SC-6	1.0000	3000	948	N/A
SC-7	1.0000	2640	821	N/A

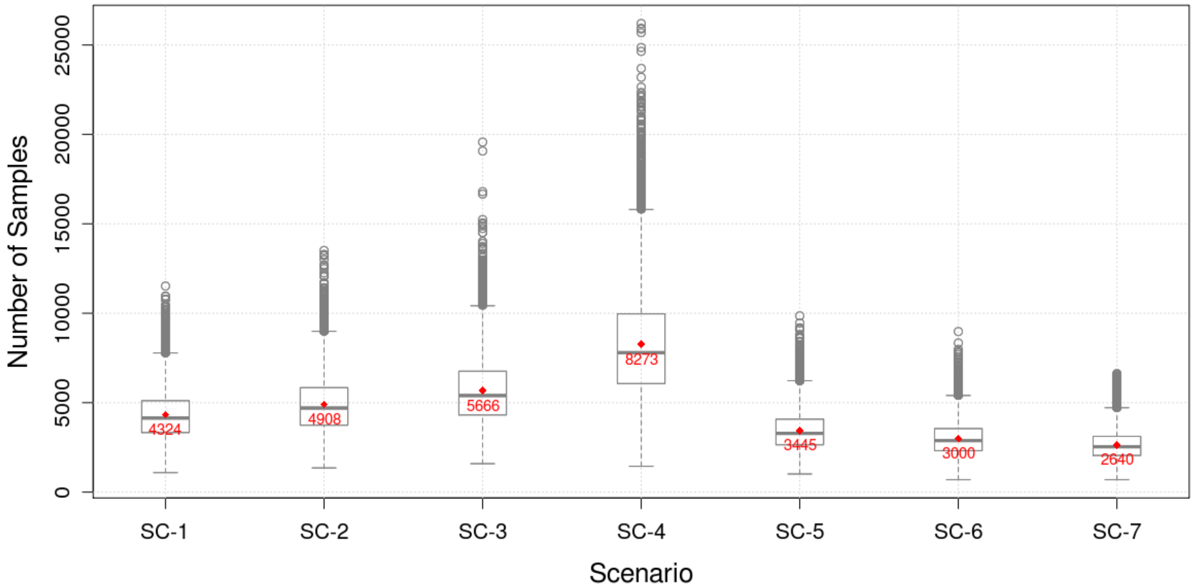


Figure 5.7: Sample sizes under different slippage configurations

other alternatives (System 2 to 10) are in the acceptable region except for System 1 (which is in the desired region since it has the maximum mean value of 100), therefore correct selections of these configuration refer to any selection containing System 1, no matter others are included or not. Since SC-7 has only System 1 in the desired region while all others are in the undesire region, its correct selection should only Systme 1.

The high Pr{**CS**} values in Table 5.14 suggest that the BSS works well. Besides that, the results also show that the most difficult configuration is SC-4, which requires the most samples and has the largest standard deviation of sample size. That is because most alternatives locate exactly at the center of the acceptable region and it become difficult for the algorithm to make a decision to either eliminate them or keep them.

5.8.4 Different distributions of competitive alternatives in each region

The last set of experiments investigates the impacts due to different distributions of competitive alternatives in each region (desired, acceptable, undesired). Three different distributions are studied and the system configurations are shown in Table 5.15.

Table 5.15: System configurations – Different distribution of alternatives in each region

	System i	1	2	3	4	5	6	7	8	9	10
	S.D.	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Mean	**(D:A:U=6:2:2)**	100	99	98	97	96	95.5	94.8	94.6	94.4	94
	(D:A:U=2:6:2)	100	97	94.9	94.8	94.76	94.74	94.7	94.6	94.4	94
	(D:A:U=2:2:6)	100	97	94.8	94.6	94.45	94.4	94.3	94.2	94.1	94

The (D:A:U) in the first column of Table 5.15 indicates the number of alternatives in each region (Desired, Acceptable, Undesired). So for (D:A:U=6:2:2), six systems (1 to 6) are in the desired region; Two systems (7 and 8) in the accpetable region and two systems (9 and 10) are in the undesired region. (D:A:U=2:6:2) and (D:A:U=2:2:6) can be explained in the same way.

The BSS parameters are still chosen as to be $(\lambda^-, \lambda^+) = (5, 5.5)$, $n_0 = 10$, and the experimental results are shown in Table 5.16 and Figure 5.8.

Table 5.16: BSS Results – Different distributions in each region

D:A:U	Pr{**CS**}	**Number of samples**		**Reason of Incorrect Selection**
		Mean	**S.D.**	
6:2:2	1.0000	5056	2253	N/A
2:6:2	1.0000	6357	2354	N/A
2:2:6	1.0000	3329	1210	N/A

From the results we can see that BSS needs more samples when there are more alternatives in the desired or acceptable regions than in the undesired region. Since it is relatively easy to eliminate a non-competitive system as long as it is found to be worse for λ^+or more than any other system (may not be the best). However, in order to stop the screening, the algorithm must accumulate enough evidences that all remaining alternatives are within λ^+ distance for each other, which is very difficult to confirm and therefore consumes a lot of samples.

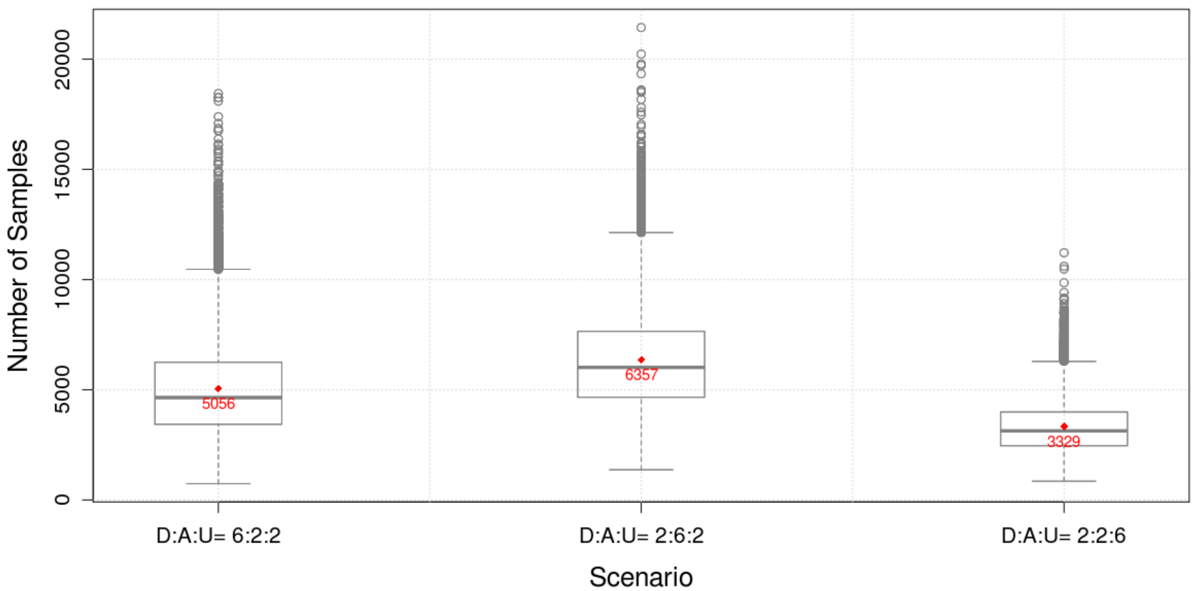


Figure 5.8: Sample sizes of different alternative distributions in each region

5.9 SUMMARY

In this chapter, we considered a new ranking-and-selection (R&S) problem which requires choosing the best subset of a set of alternative systems. We extended the existing KN procedure to a new best-subset selection (BSS) procedure to solve this problem and demonstrated its ability to select all systems that are close enough to the best system so that the decision maker is indifferent to the difference (as demonstrated by numeric experiments). From the comparison between BSS and MCB-based method, we argued that the BSS procedure is more suitable for selecting the best-subset from a finite number of alternatives. Besides that, the robustness of BSS is also analyzed by a series of experiments.

As this work was motivated by the problem of policy selection via an emergency response simulation model, we will apply this method to the model in the next chapter to examine the effects of a range of alternative emergency response policies. We believe that this procedure can provide results that are more useful to policy makers than the current selection of the best system or subset selection procedures.

We propose as future work continuing research into how to improve the efficiency of the procedure, in particular in the case where there is a large set of alternatives, while maintaining the desired probability of correct selection.

6.0 SELECT BEST EVACUATION POLICIES USING BSS

6.1 OVERVIEW

In this chapter we describe how we applied the Best-subset Selection (BSS) procedure (developed in Chapter-5) to the agent-based simulation model for a mass casualty incident (MCI) response (developed in Chapter-3) to select the best response policies under a specified scenario. The best response policies refer to those that lead to the least mortality. Besides that, we also performed sensitivity analysis to investigate the impact on the final selection due to changing parameters in the following aspects

- Required selection precision
- Degradation models of injured casualty
- Characteristics of casualty (number, injury severity distribution, percentage of specialized)
- Conditions of hospitals (available capacity of beds, admission criteria)

6.2 EXPERIMENT CONFIGURATION

The simulated MCI response system configuration is similar to the case study discussed in Section 3.2, that is, we assume an IED (Improvised Explosive Device) explosion at the Pittsburgh D. L. Lawrence Convention Center in downtown Pittsburgh, PA, United States, which resulted in 150 injured patients requiring timely evacuation. There are two types of casualties: children and adults. Children have to be treated at one of two specialized hospitals: Children's Medical Center or Magee Women's Hospital. For each of the 10 total hospitals, we assume that there are 10 available beds in general wards and 5 beds in ICU at the beginning of simulation. The injury severities are represented by different initial survival probability values, where small values correspond to severe injuries. We also assume that the injury severity of each victim is independently and identically distributed according to a specified exponential distribution. The first arriving EMTs perform on-site triage to determine the type of victims (specialized or general) and estimate the injury status based on the information gathered during the on-site triage. The actual survival probability of each casualty will deteriorate continuously until definitive care is received at a hospital. After a casualty arrives at a hospital, emergency room staff will perform in-hospital triage; Sacco's RPM score is used in the model to indicate the casualty's injury severity (after degradation). The RPM score will be used to decide if a casualty should be admitted or discharged by comparing it to a pre-defined threshold; for those being admitted, another threshold value will be used to decide if they are in critical condition or not. The critical patients will be assigned an ICU bed for treatment or be transferred to other hospital if there is no available ICU beds. In addition, we assume that regionally there are 24 ambulances available to respond to the incident. The ambulances initially start in one of 6 bases that are distributed over the Pittsburgh region. The transportation network used for the simulation is shown in Figure 6.1, which also marks the intersections of major roads (in black points), the locations of hospitals (in red crosses) and ambulance bases (in green circles).

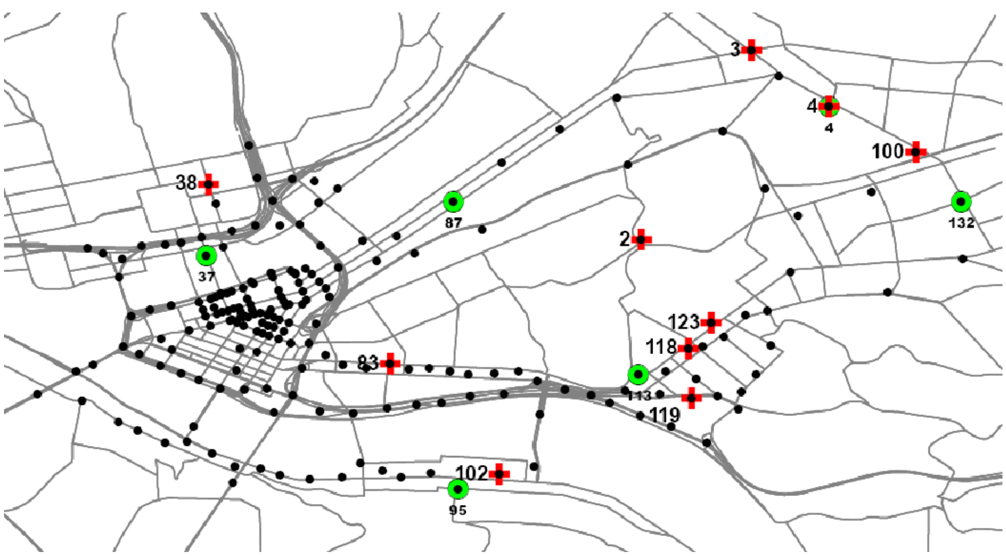


Figure 6.1: Transportation network of Pittsburgh, PA

Twelve different evacuation policies (P-1 to P-12) are proposed and employed to guide the casualty evacuation, which are detailed in Section 4.2 (Table 4.1). These policies are used to respond the incident described above. Table 6.1 summarizes all parameters used in the simulation.

Table 6.1: Simulation parameter settings

Category	**Parameter**	**Value**
Casualty Setting	Num. of Casualties	$n_c = 150$
	Initial Survival Probability Distribution	Expo(Λ), $\Lambda = 0.4$
	Percentage of Specialized Patients (Children)	$P_s = 0.2$ (30 among 150)
	Casualty Degradation Model	Sacco's RPM-based model
Ambulance Setting	Num. of Ambulance Bases	$n_b = 6$
	Num. of Ambulances	$n_a = 24$ (4 at each base)
Hospital Setting	Num. of Hospitals	$n_h = 10$
	Initial Available GW Beds	$n_{GW} = 10$
	Initial Available ICU Beds	$n_{ICU} = 5$
	Surge Capacity Ratio	$r_{sc} = 0$
	Num. of Triage Beds at ED	$n_{ar} = 3$
	Num. of Non-critical Beds at ED	$n_{ncd} = 2$
	Num. of Critical Beds at ED	$n_{cd} = 3$
	Admitting threshold (RPM score)	11
	Critical threshold (RPM score)	4
Time Setting	On-site Triage Time (min)	0.5
	ALS Pickup Time (min)	Gamma(μ = 19.15, sd = 13.98)
	BLS Pickup Time (min)	Gamma(μ = 9.27, sd = 6.43)
	Drop-off Time (min)	Gamma(μ = 23.16, sd = 12.56)
	Arrival Triage Time in hospital (min)	Gamma(μ = 5, sd = 0.5)
	Non-Critical Examination Time (min)	Gamma(μ = 7, sd = 0.5)
	Critical Examination Time (min)	Gamma(μ = 9, sd = 0.5)
Stopping Criteria – All living casualties have reached definitive care.		

The best-subset Selection (BSS) procedure is employed to determine how many times that a policy should be simulated and to select the best-subset of policies that lead to minimal mortality. The parameters are set as

- Number of Systems $k = 12$
- Expected probability of correct selection (PCS) $P^* = 1 - \alpha = 0.95$
- Initial sample size from each system $n_0 = 10$
- Boundary/indifference-zone parameters $(\lambda^-, \lambda^+) = (0.01, 0.05)$

Remark: $(\lambda^-, \lambda^+) = (0.01, 0.05)$ means, any policy whose mortality is equal to or less than (M_{best} + 0.01) must be selected to the best-subset (all policies in the desired region must be kept), where M_{best} is the best (minimal) mortality achieve by certain (unknown) policy(s). And those policies whose mortality is greater than (M_{best} + 0.05) (in the undesired region) must be eliminated and should not appear in the selection set.

6.3 EXPERIMENTAL RESULTS

All experiments (simulation + best-subset selection) are completed using a personal desktop computer with a 2.21 GHz AMD Athlon(tm) 64 CPU and 2.50GB RAM memory. The results of the experiments are shown in Table 6.2 and Figure 6.2.

Table 6.2: Result of BSS procedure (key steps & statistics of mortality)

Key Step	r =10 : I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =14 : I = {1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =17 : I = {1, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =25 : I = {1, 4, 6, 7, 8, 9, 10, 11, 12} r =32 : I = {1, 6, 7, 8, 9, 10, 11, 12} r =83 : I = {6, 7, 8, 9, 10, 11, 12} r =144 : I = {6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.508	0.658	0.585	0.558	0.541	0.491	0.482	0.474	0.479	0.483	0.488	0.491
S.D.	0.0406	0.0437	0.0399	0.0426	0.0381	0.0390	0.0399	0.0418	0.0468	0.0475	0.0425	0.0391
Samples	83	14	17	32	25	144	144	144	144	144	144	144
	1179 (in total)											

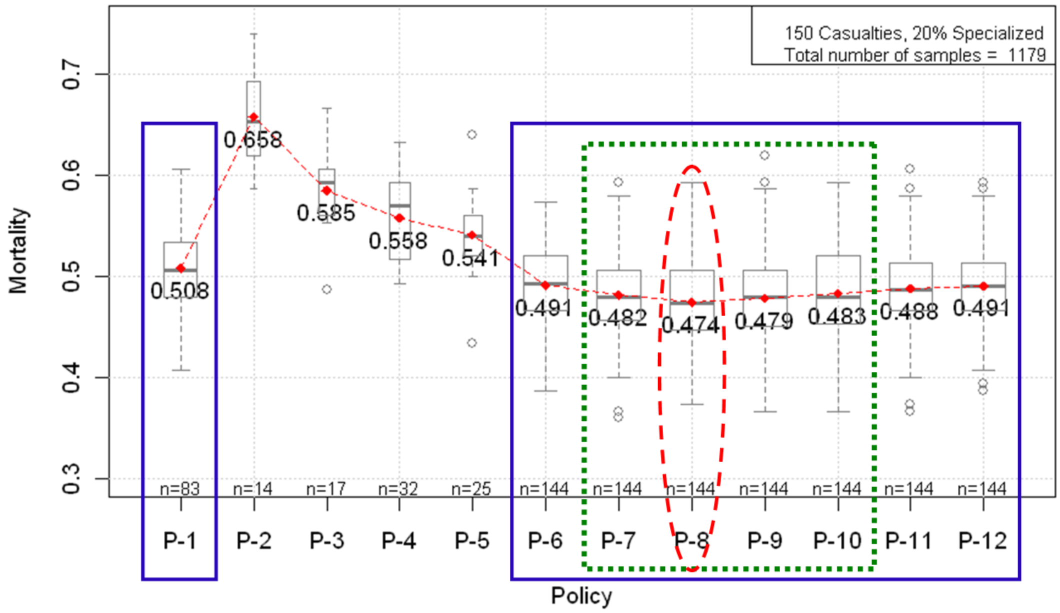


Figure 6.2: Results of BSS

The first row of Table 6.2 lists the "key steps" of BSS procedure, which include the following steps:

- the last step of initialization stage
- steps at which elimination of a candidate policy occurs
- final step that determines the final resulting subset

The last row of Table 6.2 lists the number of replications simulated of each alternative (policy). It can be noted that the numbers are not equal because inferior policies are eliminated in earlier screening stages therefore fewer replications are made. Based on the replications made, the mean and standard deviation of mortality under each policy are calculated and shown in the corresponding rows in Table 6.2.

Figure 6.2 displays the box-plot of the performance of alternative policies, with the width of each box proportional to the number of replications made for each alternative (which is printed at the bottom, above the x-axis). The mean mortality of each policy is also marked (and linked by dash line) on the plot.

By eye-balling the figure, we can observe that the best policy is P-8 (circled by an oval in dash), which has the minimal mean mortality of 0.474 (47.4% of victims died on average). For the policies in a dotted box (P-7 to P-10), their mean values of mortality are with 0.01-distance to P-8 so that they should be included since they are in the desired region. And for those policies out of the box in solid-line (P-2 to P-5), their mean mortality exceed the 0.05-distance to P-8, and thus should excluded from the final selection. Besides that, P-1, P-6, P-11 and P-12 are in the acceptable region, and it does not matter whether they are selected or not. As comparison, the final selection of BSS procedure is I = $\{6, 7, 8, 9, 10, 11, 12\}$, which is consistent with our discussion above.

6.4 SENSITIVITY ANALYSIS

6.4.1 Impact of changing required precision of selection

This experiment is used to check the impact of changing required precision of selection. Assume that we keep all simulation parameters unchanged but make the selection precision requirement stricter by setting $(\lambda^-, \lambda^+) = (0.01, 0.02)$, which means

1. Any policy whose mortality $\leq$ (M_{best} + 0.01) must be selected (desired);
2. Any policy whose mortality $>$ (M_{best} + 0.02) should be excluded from the final selection (Discard any policy whose mortality is greater by 0.02 than the unknown best).

The experiment results are shown in Table 6.3 and Figure 6.3. Following the similar analysis in Section 6.3, we can see that the selection result of BSS (I = $\{7, 8, 9, 10, 11, 12\}$) is plausible since all desired policies that are boxed by dotted line (P-7 to P-11) are selected, where the best one is P-7. Besides that, none of the alternatives in the undesired region (P-1 to P-5, which are out of the solid-line box) was included in the selection set. P-6 and P-12 are in the acceptable region, and it does not matter whether they are selected or not.

Table 6.3: Result of BSS procedure – precision of selection $(\lambda^-, \lambda^+) = (0.01, 0.02)$

Required Precision of Selection $(\lambda^-, \lambda^+) = (0.01, 0.02)$												
Key Step	r =10 : I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =51 : I = {1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =64 : I = {1, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =100 : I = {1, 5, 6, 7, 8, 9, 10, 11, 12} r =111 : I = {1, 6, 7, 8, 9, 10, 11, 12} r =392 : I = {6, 7, 8, 9, 10, 11, 12} r =2817 : I = {7, 8, 9, 10, 11, 12} r =2817 : I = {7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.504	0.660	0.591	0.560	0.532	0.494	0.477	0.479	0.480	0.484	0.487	0.489
S.D.	0.0418	0.0374	0.0412	0.0413	0.0437	0.0411	0.0414	0.0415	0.0419	0.0423	0.0422	0.0419
Samples	392	51	64	100	111	2817	2817	2817	2817	2817	2817	2817
	20437 (in total)											

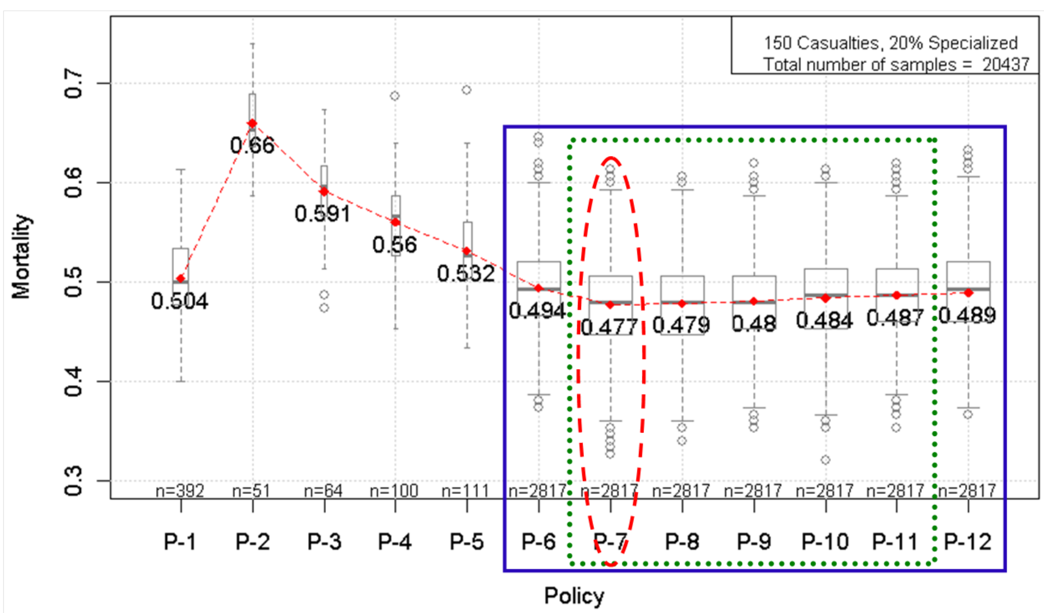


Figure 6.3: Results of BSS – $(\lambda^-, \lambda^+) = (0.01, 0.02)$

Table 6.4 compares the results of $(\lambda^-, \lambda^+) = (0.01, 0.02)$ to the results of $(\lambda^-, \lambda^+) = (0.01, 0.05)$, and it can be found that the strict selection precision setting ($(\lambda^-, \lambda^+) = (0.01, 0.02)$) requires nearly twenty times more samples than the original one ($(\lambda^-, \lambda^+) = (0.01, 0.05)$). The reason is because the width of the indifference-zone (2ε) is smaller for $(\lambda^-, \lambda^+) = (0.01, 0.02)$ than $(\lambda^-, \lambda^+) = (0.01, 0.05)$, which makes it more difficult to confirm that all remaining alternatives are desired or acceptable, and thus causes a sharp increase of the number of replications (a.k.a., total samples), as we have discussed in Section 5.8.1.1.

Table 6.4: Comparion between different precision settings

Parameter	$(\lambda^-, \lambda^+) = (0.01, 0.05)$	$(\lambda^-, \lambda^+) = (0.01, 0.02)$
Indifference-zone Width (2ε)	0.04	0.01
Number of Rounds (r)	144	2817
Total Samples (N)	1179	20437
Final Selection Set (P-)	{6, 7, 8, 9, 10, 11, 12}	{7, 8, 9, 10, 11, 12}

6.4.2 Casualty setting

6.4.2.1 Degradation models This set of experiments is to check whether different degradation models will affect the final selection of policies. Sacco's degradation model (RPM-based) is employed in the original experiment, for comparison, a set of proportional-hazard based degradation models (Formula 3.2 with $g = 1.0$, 1.045, 1.196 and 2.007, the same set of values used in Section 3.2.5.4) is implemented to simulate the injury deterioration for the same set of casualties with the same distribution of initial injury severity. The experimental results are shown and compared in Table 6.5 and Figure 6.4.

Table 6.5: Results of BSS procedure – different degradation models

Proportional-hazard Based Degradation Model ($g = 1.0$)												
Key Step	r =10 : I = $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ r =79 : I = $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.389	0.390	0.387	0.380	0.387	0.389	0.385	0.382	0.385	0.386	0.386	0.387
S.D.	0.0421	0.0420	0.0453	0.0417	0.0368	0.0429	0.0409	0.0404	0.0340	0.0432	0.0461	0.0433
Samples	79	79	79	79	79	79	79	79	79	79	79	79
	948 (in total)											

Proportional-hazard Based Degradation Model ($g = 1.045$)												
Key Step	r =10 : I = $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ r =21 : I = $\{1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ r =33 : I = $\{1, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ r =56 : I = $\{1, 5, 6, 7, 8, 9, 10, 11, 12\}$ r =86 : I = $\{1, 6, 7, 8, 9, 10, 11, 12\}$ r =88 : I = $\{1, 6, 7, 8, 9, 10, 11, 12\}$ (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.507	0.622	0.573	0.547	0.529	0.498	0.500	0.494	0.499	0.499	0.501	0.503
S.D.	0.0405	0.0509	0.0304	0.0426	0.0400	0.0457	0.0420	0.0404	0.0434	0.0374	0.0411	0.0436
Samples	88	33	21	56	86	88	88	88	88	88	88	88
	900 (in total)											

Proportional-hazard Based Degradation Model ($g = 1.196$)												
Key Step	r =10 : I = $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ r =10 : I = $\{1, 5, 6, 7, 8, 9, 10, 11, 12\}$ r =18 : I = $\{1, 6, 7, 8, 9, 10, 11, 12\}$ r =114 : I = $\{6, 7, 8, 9, 10, 11, 12\}$ r =114 : I = $\{6, 7, 8, 9, 10, 11, 12\}$ (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.725	0.900	0.855	0.811	0.771	0.697	0.695	0.692	0.697	0.698	0.697	0.696
S.D.	0.0310	0.0183	0.0174	0.0213	0.0314	0.0359	0.0334	0.0364	0.0373	0.0356	0.0364	0.0342
Samples	114	10	10	10	18	114	114	114	114	114	114	114
	960 (in total)											

(Continued on next page . . .)

(Table 6.5: continued)

Proportional-hazard Based Degradation Model ($g = 2.007$)												
Key Step	r =10 : I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =12 : I = {1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =68 : I = {1, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =68 : I = {1, 4, 5, 6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.901	0.937	0.924	0.896	0.892	0.890	0.894	0.890	0.887	0.889	0.890	0.890
S.D.	0.0245	0.0227	0.0229	0.0254	0.0217	0.0233	0.0273	0.0230	0.0230	0.0255	0.0238	0.0233
Samples	68	12	68	68	68	68	68	68	68	68	68	68
	760 (in total)											

Sacco's degradation model (RPM-based)												
Key Step	r =10 : I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =14 : I = {1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =17 : I = {1, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =25 : I = {1, 4, 6, 7, 8, 9, 10, 11, 12} r =32 : I = {1, 6, 7, 8, 9, 10, 11, 12} r =83 : I = {6, 7, 8, 9, 10, 11, 12} r =144 : I = {6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.508	0.658	0.585	0.558	0.541	0.491	0.482	0.474	0.479	0.483	0.488	0.491
S.D.	0.0406	0.0437	0.0399	0.0426	0.0381	0.0390	0.0399	0.0418	0.0468	0.0475	0.0425	0.0391
Samples	83	14	17	32	25	144	144	144	144	144	144	144
	1179 (in total)											

From Table 6.5 and Figure 6.4, the first observation is the mortality results of $g = 1.0$ are nearly equal, in comparison with other cases where obvious differences in mortality can be observed for different evacuation policies. It is because $g = 1.0$ corresponds to no degradation, so that the only determinant factor of mortality is the initial survival probability of casualty. Whether a casualty survives or not is determined in the very beginning, and has nothing with the evacuation policies. Such a counterintuitive finding suggests in turn that the casualty degradation is an important factor in our simulation.

The mortality results of $g = 1.045$ are very close to the results of Sacco's model, and it also results in a similar selection set with Sacco's model except for including one more acceptable policy

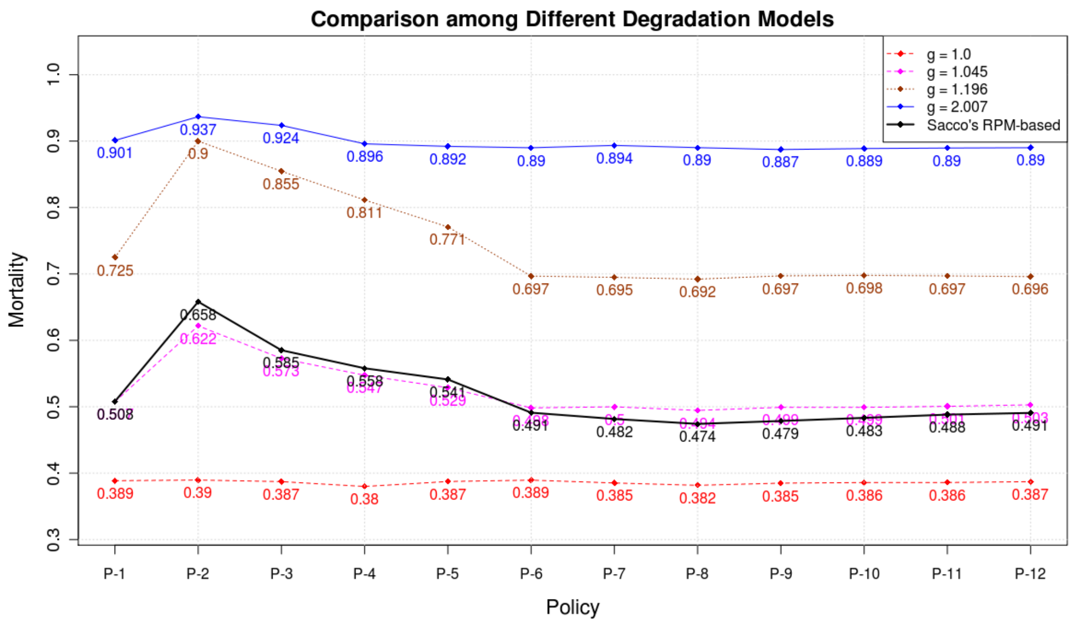


Figure 6.4: Comparison among different degradation models

(P-1) in its selection set. Besides that, it can be found that $g = 1.196$ results in an exactly same policy selection set with Sacco's (although its mortalities are higher). These findings imply that it is possible to replace the Sacco's RPM-based model by a proportional-hazard based model with appropriate base value (g), since the policy selection is not very sensitive to the specific degradation model.

The mortalities of $g = 1.196$ and $g = 2.007$ are much higher than Sacco's degradation model, no matter which policy is specified. This is because that the proportional-hazard based model assumes the same degradation rate for all casualties despite his/her initial injury severity, which makes mildly injured casualties deteriorate much faster than those using Sacco's degradation model, and thus leading to higher mortalities.

By comparing the selection set, the final selection of proportional-hazard based model with $g = 2.007$ contains three more policies (P-1, P-4, and P-5) than Sacco's model, which implies the $g = 2.007$ provides less differentiation for various policies. Besides that, it can be observed the Sacco's model consumes the most number of samples, which is because that P-1, P-6, P-11 and P-12 are in the acceptable region and close to the center ($\sim$0.504), which makes it difficulty to decide whether to eliminate it (e.g., P-1) or to keep it (e.g., P-6, P-11, P-12), as we have discussed in Section 5.8.3.

6.4.2.2 Different number of casualties Table 6.6 lists the results of best-subset selection of policy via simulation for different number of victims. All other parameters are the same as the original experiment configuration in Section 6.2. The mean values of mortality are compared in Figure 6.5.

Table 6.6: Results of BSS procedure – different number of casualties

Casualty Number n_c = 100												
Key Step	r =10 : I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =41 : I = {1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12} r =53 : I = {1, 3, 5, 6, 7, 8, 9, 10, 11, 12} r =76 : I = {1, 5, 6, 7, 8, 9, 10, 11, 12} r =184 : I = {1, 5, 6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.461	0.556	0.512	0.489	0.461	0.440	0.441	0.438	0.436	0.439	0.437	0.440
S.D.	0.0485	0.0605	0.0524	0.0435	0.0489	0.0496	0.0485	0.0489	0.0497	0.0497	0.0501	0.0511
Samples	184	53	76	41	184	184	184	184	184	184	184	184
	1826 (in total)											
Casualty Number n_c = 150												
Key Step	r =10 : I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =14 : I = {1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =17 : I = {1, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =25 : I = {1, 4, 6, 7, 8, 9, 10, 11, 12} r =32 : I = {1, 6, 7, 8, 9, 10, 11, 12} r =83 : I = {6, 7, 8, 9, 10, 11, 12} r =144 : I = {6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.508	0.658	0.585	0.558	0.541	0.491	0.482	0.474	0.479	0.483	0.488	0.491
S.D.	0.0406	0.0437	0.0399	0.0426	0.0381	0.0390	0.0399	0.0418	0.0468	0.0475	0.0425	0.0391
Samples	83	14	17	32	25	144	144	144	144	144	144	144
	1179 (in total)											

(Continued on next page . . .)

Casualty Number $n_c = 200$												
Key Step	r =10 : I = {1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =12 : I = {1, 3, 5, 6, 7, 8, 9, 10, 11, 12} r =14 : I = {1, 5, 6, 7, 8, 9, 10, 11, 12} r =15 : I = {1, 6, 7, 8, 9, 10, 11, 12} r =83 : I = {6, 7, 8, 9, 10, 11, 12} r =117 : I = {6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.567	0.673	0.642	0.608	0.588	0.544	0.535	0.540	0.537	0.535	0.534	0.539
S.D.	0.0376	0.0356	0.0412	0.0208	0.0237	0.0360	0.0346	0.0325	0.0346	0.0347	0.0329	0.0324
Samples	83	10	14	12	15	117	117	117	117	117	117	117
	953 (in total)											
Casualty Number $n_c = 300$												
Key Step	r =10 : I = {1, 4, 6, 7, 8, 9, 10, 11, 12} r =32 : I = {1, 6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.706	0.805	0.774	0.741	0.734	0.699	0.697	0.701	0.697	0.692	0.693	0.699
S.D.	0.0244	0.0264	0.0224	0.0277	0.0240	0.0309	0.0229	0.0253	0.0274	0.0282	0.0277	0.0266
Samples	32	10	10	32	10	32	32	32	32	32	32	32
	318 (in total)											
Casualty Number $n_c = 400$												
Key Step	r =10 : I = {1, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =33 : I = {1, 5, 6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.795	0.854	0.852	0.824	0.807	0.794	0.798	0.798	0.798	0.802	0.802	0.799
S.D.	0.0178	0.0257	0.0188	0.0186	0.0182	0.0192	0.0203	0.0186	0.0174	0.0194	0.0187	0.0169
Samples	33	10	10	33	33	33	33	33	33	33	33	33
	350 (in total)											

From the results, it is easy to find that less casualties correspond to less mortalities, which is because more casualties cause queues to build up resulting in longer waiting times at the hospital

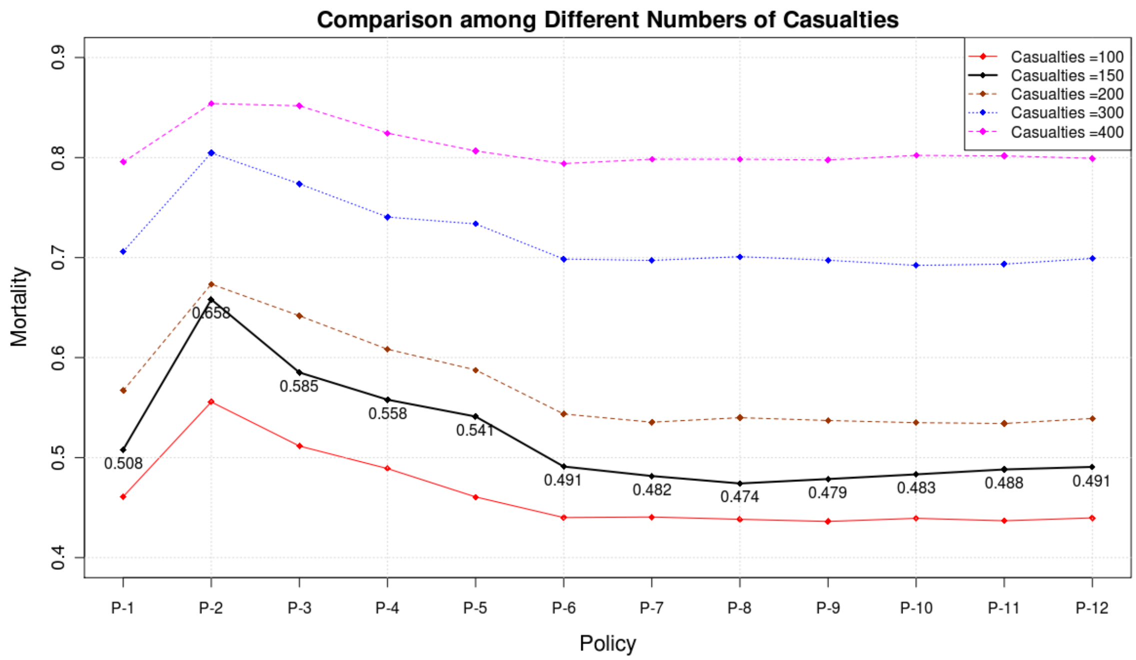


Figure 6.5: Comparison among different numbers of casualites

and exhausting the limited capacity of hospitals so that many severe patients pass away because they do not receive timely care.

Besides that, the distinction of policies is more obvious when the number of casualties is close to the number of available hospital beds. More or less casualties both blur the distinction. If there are an excess number of patients, then many would not receive timely treatment due to the shortage of capacity (beds), no matter which evacuation policy is in use, which leads to high mortalities (means) with less differences among policies. On the contrary, a lower number of casualties reduces the possible waiting and transitions, so that most patients receive timely treatment and avoid further deterioration, which decreases the mortality and also leads to less differentiation among policies as well.

Based on similar analysis, it is not unexpected to observe that the variances in mortality with less casualties are larger than those with more casualties for a given evacuation policy, which results in more replications (samples) being needed in order to select the best policy subset for cases with less casualties (100, 150) than those with more casualties (300, 400).

6.4.2.3 Different distributions of injury severity The injury severity of victims is measured by their initial survival probability, which is assumed to be distributed exponentially with a certain rate Λ. The value of Λ determines the distribution pattern of the injury severity of casualties. In general, a smaller Λ corresponds to a more severe incident since it leads to more of the casualties having lower initial survival probabilities, and a larger Λ corresponds to a milder incident because more victims have higher initial survival probabilities, as illustrated in Section 3.2.5.2 (Figure 3.7). Therefore, the value of Λ can be considered as a measure of the level of incident severity. This set of experiments is used to investigate the impact of different incident severity on the policy selection by changing the value of Λ ($\Lambda = 0.1, 0.4, 0.7, 0.9$), and the experimental results are shown in Table 6.7 and Figure 6.6.

Table 6.7: Results of BSS procedure – different initial injury severity distributions

Distributions of injury severity EXPO(Λ) $\Lambda = 0.1$												
Key Step	r =10 : I = {1, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =78 : I = {1, 5, 6, 7, 8, 9, 10, 11, 12} r =106 : I = {1, 5, 6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.735	0.831	0.822	0.767	0.742	0.719	0.720	0.722	0.721	0.722	0.718	0.718
S.D.	0.0415	0.0211	0.0376	0.0363	0.0412	0.0506	0.0461	0.0510	0.0505	0.0509	0.0480	0.0498
Samples	106	10	10	78	106	106	106	106	106	106	106	106
	1052 (in total)											
Distributions of injury severity EXPO(Λ) $\Lambda = 0.4$												
Key Step	r =10 : I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =14 : I = {1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =17 : I = {1, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =25 : I = {1, 4, 6, 7, 8, 9, 10, 11, 12} r =32 : I = {1, 6, 7, 8, 9, 10, 11, 12} r =83 : I = {6, 7, 8, 9, 10, 11, 12} r =144 : I = {6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.508	0.658	0.585	0.558	0.541	0.491	0.482	0.474	0.479	0.483	0.488	0.491
S.D.	0.0406	0.0437	0.0399	0.0426	0.0381	0.0390	0.0399	0.0418	0.0468	0.0475	0.0425	0.0391

(Continued on next page . . .)

(Table 6.7: continued)

Samples	83	14	17	32	25	144	144	144	144	144	144	144
	1179 (in total)											
Distributions of injury severity EXPO(Λ) $\Lambda = 0.7$												
Key Step	r =10 : I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =11 : I = {1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =36 : I = {1, 3, 4, 6, 7, 8, 9, 10, 11, 12} r =51 : I = {1, 4, 6, 7, 8, 9, 10, 11, 12} r =87 : I = {1, 6, 7, 8, 9, 10, 11, 12} r =271 : I = {1, 6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.420	0.547	0.467	0.456	0.427	0.407	0.403	0.405	0.406	0.408	0.410	0.408
S.D.	0.0392	0.0428	0.0380	0.0442	0.0355	0.0390	0.0401	0.0397	0.0390	0.0388	0.0381	0.0380
Samples	271	11	51	87	36	271	271	271	271	271	271	271
	2353 (in total)											
Distributions of injury severity EXPO(Λ) $\Lambda = 0.9$												
Key Step	r =10 : I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =16 : I = {1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =26 : I = {1, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =187 : I = {1, 4, 5, 6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.402	0.491	0.422	0.417	0.406	0.390	0.394	0.392	0.391	0.393	0.391	0.394
S.D.	0.0396	0.0465	0.0409	0.0398	0.0432	0.0389	0.0382	0.0396	0.0411	0.0395	0.0420	0.0424
Samples	187	26	16	187	187	187	187	187	187	187	187	187
	1912 (in total)											

From these results we can see that for those relative mild incidents (e.g., $\Lambda = 0.7$ or 0.9), the mean values of mortality for different policies are generally lower, which follows our intuition since the initial survival probabilities of the casualties are higher, hence their injury degradations are slower accordingly, so that more patients are likely to survive as long as they can receive treatments within moderate time periods. On the other hand, the mortality of a severe incident (e.g., $\Lambda = 0.1$) is higher because of the lower initial survival probabilities, which cause victims to more likely pass away in a short time.

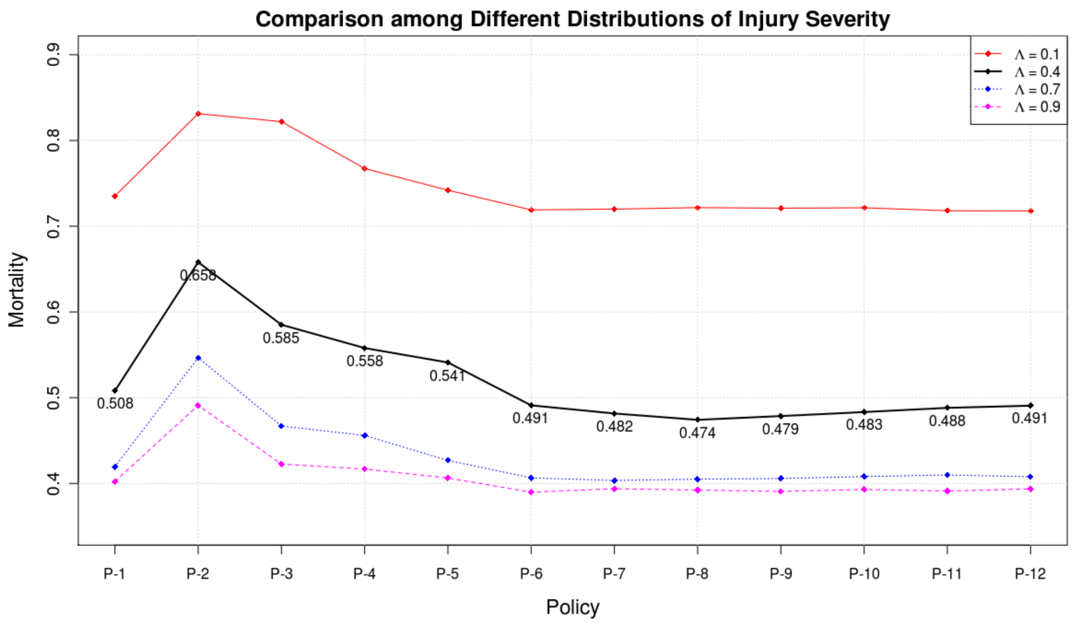


Figure 6.6: Comparison among different initial injury severity distributions

It can be observed when compared with the moderate incident (e.g., $\Lambda = 0.4$) that the differences of the means for mortality under different evacuation policies are less in either a very mild incident (e.g., $\Lambda = 0.9$) or a very severe incident (e.g., $\Lambda = 0.1$). This is because the main difference between different evacuation policies is reflected in the length of time that casualties have to wait before receiving definitive care at certain hospitals. However, based on the analysis above, very high initial survival probabilities could support patients' survival for a quite long period, while very low initial survival probabilities would lead to quick deaths in a very short time. For both cases, mortality is less sensitive to the variation of evacuation time, and thus lessening the difference between different evacuation policies. And it also explains why more policies are selected for ($\Lambda = 0.1$) and ($\Lambda = 0.9$) than ($\Lambda = 0.4$).

People may also notice that the scenario of $\Lambda = 0.7$ consumes much more replications than other scenario, while its sample variances are generally smaller than others. The reason is because the mortality of P-1 is only slightly less than the median of the acceptable region, which makes hard for the algorithm to determine whether P-1 should be kept or eliminated (refer to the discussion in Section 5.8.3), and therefore more replications are required to reach the decision.

6.4.2.4 Different percentage of specialized casualties In this experiment, specialized casualties refer to children since they need to be cared in either of the specialized hospitals: Children's Medical Center or Magee Women's Hospital. The experiments in this section are used to investigate the impact to the policy selection due to the change of percentage of specialized casualties (P_s = 0.0, 0.2, 0.6, 1.0). The experimental results are shown in Table 6.8 and Figure 6.7.

Table 6.8: Results of BSS procedure – different percentages of specialized patients

	P_s = 0.0											
Key Step	r =10 : I = {1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =19 : I = {1, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =28 : I = {1, 5, 6, 7, 8, 9, 10, 11, 12} r =50 : I = {1, 6, 7, 8, 9, 10, 11, 12} r =132 : I = {1, 6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.491	0.652	0.568	0.557	0.515	0.475	0.495	0.475	0.475	0.475	0.475	0.475
S.D.	0.0415	0.0400	0.0364	0.0394	0.0384	0.0435	0.0453	0.0435	0.0435	0.0435	0.0435	0.0435
Samples	132	10	19	28	50	132	132	132	132	132	132	132
	1163 (in total)											
	P_s = 0.2											
Key Step	r =10 : I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =14 : I = {1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =17 : I = {1, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =25 : I = {1, 4, 6, 7, 8, 9, 10, 11, 12} r =32 : I = {1, 6, 7, 8, 9, 10, 11, 12} r =83 : I = {6, 7, 8, 9, 10, 11, 12} r =144 : I = {6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.508	0.658	0.585	0.558	0.541	0.491	0.482	0.474	0.479	0.483	0.488	0.491
S.D.	0.0406	0.0437	0.0399	0.0426	0.0381	0.0390	0.0399	0.0418	0.0468	0.0475	0.0425	0.0391
Samples	83	14	17	32	25	144	144	144	144	144	144	144
	1179 (in total)											
	P_s = 0.6											

(Continued on next page . . .)

(Table 6.8: continued)

Key Step	r =10 : I = {1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =27 : I = {1, 3, 4, 6, 7, 8, 9, 10, 11, 12} r =35 : I = {1, 4, 6, 7, 8, 9, 10, 11, 12} r =157 : I = {1, 4, 6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.574	0.717	0.611	0.583	0.605	0.565	0.559	0.558	0.560	0.559	0.559	0.561
S.D.	0.0446	0.0269	0.0421	0.0420	0.0362	0.0437	0.0457	0.0483	0.0488	0.0494	0.0485	0.0492
Samples	157	10	35	157	27	157	157	157	157	157	157	157
	1485 (in total)											
$P_s = 1.0$												
Key Step	r =10 : I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =89 : I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.714	0.723	0.716	0.713	0.704	0.709	0.709	0.709	0.709	0.709	0.709	0.709
S.D.	0.0372	0.0335	0.0326	0.0326	0.0330	0.0278	0.0278	0.0278	0.0278	0.0278	0.0278	0.0278
Samples	89	89	89	89	89	89	89	89	89	89	89	89
	1068 (in total)											

From the results, it can be found that the scenario with specialized percentage $P_s = 1.0$ (which means all casualties are children) has the highest mean values of mortality with almost no difference among different policies, and the variances of mortality for all policies are generally smaller than other scenarios, which leads to all alternatives being selected into the final set with minimum replications. The reason is because there is not enough total capacities (beds) in the two specialized hospitals, no matter which evacuation policy is used, the only possible result is the two specialized hospitals are overwhelmed soon by excess patients, so that the means of mortality are roughly equally high and the variances are relatively small despite the policy choice. As the percentage of specialized patients decreases, there are more general (adult) patients who can be cared for by general hospitals, which reduces the load on the specialized hospitals and thus decreases mortalities accordingly.

The most replications is required by the scenario with percentage $P_s = 0.6$, which is because its variances of mortality for the selected policies are generally larger than other scenarios, so it

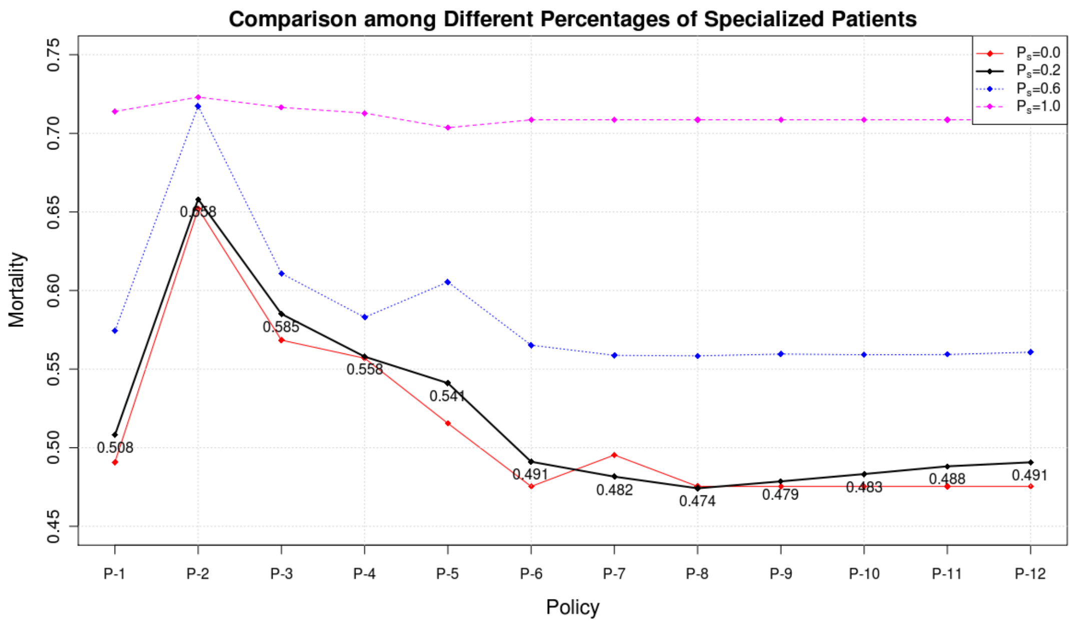


Figure 6.7: Comparison among different percentages of specialized patients

needs additional replications to confirm that the remaining alternatives are the best while only a few replications are required to eliminate the inferior policies.

A small bump can be observed at P-5 for the scenario of P_s = 0.6. The reason for the bump is analyzed as follows. The decision criteria of P-5 are the available ED capacity and the distance from scene to hospital, which aims to provide immediate treatment to patients upon their arrival. However, a high percentage of specialized patients implies less general (adult) patients that need to be treated. Consequently, adult patients more quickly pass through the triage at ED in the nearby hospitals with almost no waiting, and releasing their capacity. However, such a policy also causes the general patients to be continuously sent to a few nearby hospitals and exhausts their available ICU and GW beds in a short time, leading to the small bump of mortality.

Another interesting phenomenon deserving mention is the small peak that appeared at P-7 for the scenario with percentage P_s = 0.0. As we already know, P_s = 0.0 means there are no children casualties at all, but P-7 still reserves the capacity of specialized hospitals for the non-existing children patients, which wastes the capacity and leads to the small peak of mortality at P-7 for P_s = 0.0.

6.4.3 Hospital setting

6.4.3.1 Different number of Beds The original number of available beds is (ICU, GW) = (5, 10). This set of experiments changes the available number of beds in ICU and General Wards to (3, 6) (60% of the original) and (10, 20) (200% of the original) respectively, and checks their impacts on the final selection. All other parameters are the same as the original setting. The experimental results are shown in Table 6.9 and Figure 6.8.

Table 6.9: Results of BSS procedure – different numbers of beds in a hospital

Number of Available Beds (ICU, GW) = (3, 6)												
Key Step	r =10 : I = {1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =16 : I = {1, 3, 5, 6, 7, 8, 9, 10, 11, 12} r =22 : I = {1, 5, 6, 7, 8, 9, 10, 11, 12} r =54 : I = {1, 6, 7, 8, 9, 10, 11, 12} r =175 : I = {1, 6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.559	0.681	0.620	0.617	0.582	0.541	0.543	0.545	0.540	0.541	0.542	0.542
S.D.	0.0451	0.0387	0.0372	0.0321	0.0448	0.0432	0.0434	0.0425	0.0421	0.0421	0.0419	0.0434
Samples	175	10	22	16	54	175	175	175	175	175	175	175
	1502 (in total)											

Number of Available Beds (ICU, GW) = (5, 10)												
Key Step	r =10 : I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =14 : I = {1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =17 : I = {1, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =25 : I = {1, 4, 6, 7, 8, 9, 10, 11, 12} r =32 : I = {1, 6, 7, 8, 9, 10, 11, 12} r =83 : I = {6, 7, 8, 9, 10, 11, 12} r =144 : I = {6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.508	0.658	0.585	0.558	0.541	0.491	0.482	0.474	0.479	0.483	0.488	0.491
S.D.	0.0406	0.0437	0.0399	0.0426	0.0381	0.0390	0.0399	0.0418	0.0468	0.0475	0.0425	0.0391
Samples	83	14	17	32	25	144	144	144	144	144	144	144
	1179 (in total)											

(Continued on next page . . .)

(Table 6.9: continued)

Number of Available Beds (ICU, GW) = (10, 20)												
Key Step	r =10 : I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =19 : I = {1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =73 : I = {1, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =113 : I = {1, 4, 5, 6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.468	0.545	0.488	0.476	0.456	0.453	0.455	0.455	0.460	0.454	0.454	0.452
S.D.	0.0462	0.0383	0.0432	0.0430	0.0420	0.0429	0.0462	0.0471	0.0437	0.0399	0.0440	0.0400
Samples	113	19	73	113	113	113	113	113	113	113	113	113
	1222 (in total)											

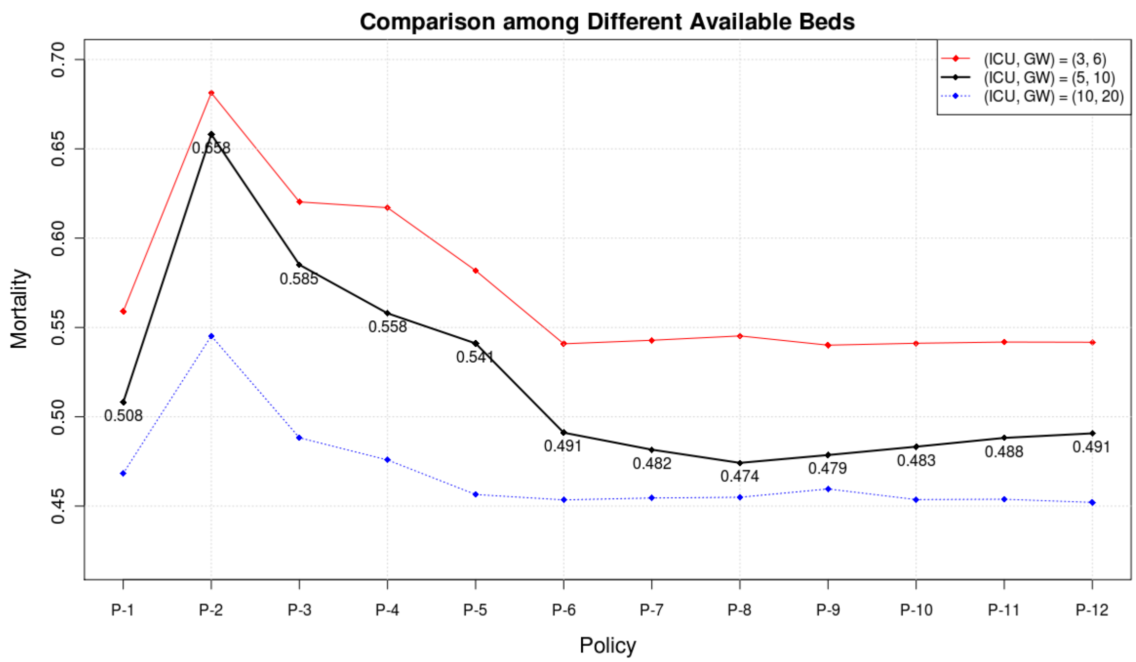


Figure 6.8: Comparison among different numbers of beds in a hospital

It can be seen that for the scenarios of (10, 20) (more beds) and (3, 6) (less beds), either of them makes the policies less distinguishable than the original configuration (5, 10). The reason is similar to what has been discussed in Section 6.4.2.2 (different number of casualties). Less beds make it easy to exhaust the capacity of hospitals and lead to higher mortality with less distinction among policies, while more beds bring excess capacity which helps reduce possible waiting and transition and lower the mean and variance of mortality.

In terms of number of replications, the less-beds scenario (3, 6) requires more replications to determine whether or not to keep the P-1 since its mean mortality is in the "desired-side" of acceptable region but close to the median of acceptable region, as we have analyzed in Section 6.4.2.3 (different distributions of injury severity, the scenario of $\Lambda = 0.7$)

6.4.3.2 Different admission criteria The admission criteria consists of two thresholds measured in RPM score (S_c, S_a). S_a is used to decide whether a casualty should be admitted or discharged, and is estimated during the in-hospital triage upon casualty's arrival. S_c is used to decide whether the admitted casualty is in critical condition or not, and determines the specific type of bed (i.e., ICU or GW bed) that the patient will received. For example, the admission threshold used in the original experiment is (S_c, S_a) = (4, 11), which means that any casualty whose RPM score (after degradation) is less than or equal to 11 will be admitted and be assigned to a bed. Furthermore, if the degraded RPM score is less than or equal to 4 (which indicate a severe injury), a ICU bed will be assigned otherwise a GW bed will be assigned to the patient.

For better understanding the impact of the function of each threshold value, each time we only change one parameter. For example, the original admission thresholds are (S_c, S_a) = (4, 11), which means that any patient whose degraded RPM score is less than or equal to S_a=11 will be admitted as an in-patient, but only those whose RPM $\leq$ 4 (S_c) can be assigned to an ICU bed. Now if we raise the threshold for critical patients, that is, let (S_c, S_a) = (3, 11), then the patients with RPM $\leq$ 11 still can be admitted, but only those whose RPM $\leq$ 3 will receive an ICU bed.

In a similar way we investigate the impact of raising/lowering the admitting/critical threshold, the simulation results are shown in Table 6.10 and Figure 6.9.

Table 6.10: Results of BSS procedure – different admission thresholds

Admission Threshold (S_c, S_a) = (3, 11)												
Key Step	r =10 : I = {1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12} r =11 : I = {1, 3, 5, 6, 7, 8, 9, 10, 11, 12} r =18 : I = {1, 5, 6, 7, 8, 9, 10, 11, 12} r =26 : I = {1, 6, 7, 8, 9, 10, 11, 12} r =150 : I = {6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.510	0.635	0.589	0.602	0.526	0.482	0.483	0.484	0.480	0.477	0.477	0.478
S.D.	0.0440	0.0285	0.0403	0.0399	0.0382	0.0428	0.0416	0.0402	0.0402	0.0414	0.0431	0.0440
Samples	150	11	18	10	26	150	150	150	150	150	150	150
	1265 (in total)											

Admission Threshold (S_c, S_a) = (4, 10)												
Key Step	r =10 : I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =11 : I = {1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =27 : I = {1, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =80 : I = {1, 5, 6, 7, 8, 9, 10, 11, 12} r =102 : I = {1, 5, 6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.517	0.592	0.564	0.538	0.516	0.500	0.511	0.505	0.502	0.503	0.499	0.503
S.D.	0.0431	0.0407	0.0472	0.0422	0.0481	0.0468	0.0456	0.0428	0.0427	0.0431	0.0431	0.0465
Samples	102	11	27	80	102	102	102	102	102	102	102	102
	1036 (in total)											

Admission Threshold (S_c, S_a) = (4, 11)												
Key Step	r =10 : I = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =14 : I = {1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =17 : I = {1, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =25 : I = {1, 4, 6, 7, 8, 9, 10, 11, 12} r =32 : I = {1, 6, 7, 8, 9, 10, 11, 12} r =83 : I = {6, 7, 8, 9, 10, 11, 12} r =144 : I = {6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.508	0.658	0.585	0.558	0.541	0.491	0.482	0.474	0.479	0.483	0.488	0.491
S.D.	0.0406	0.0437	0.0399	0.0426	0.0381	0.0390	0.0399	0.0418	0.0468	0.0475	0.0425	0.0391
Samples	83	14	17	32	25	144	144	144	144	144	144	144
	1179 (in total)											

(Continued on next page ...)

Admission Threshold $(S_c, S_a) = (4, 12)$												
Key Step	r =10 : I = {1, 3, 5, 6, 7, 8, 9, 10, 11, 12} r =16 : I = {1, 5, 6, 7, 8, 9, 10, 11, 12} r =31 : I = {1, 6, 7, 8, 9, 10, 11, 12} r =204 : I = {1, 6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.520	0.672	0.607	0.580	0.544	0.500	0.498	0.496	0.498	0.497	0.501	0.501
S.D.	0.0407	0.0270	0.0273	0.0401	0.0356	0.0390	0.0427	0.0375	0.0378	0.0376	0.0374	0.0390
Samples	204	10	16	10	31	204	204	204	204	204	204	204
	1699 (in total)											

Admission Threshold $(S_c, S_a) = (5, 11)$												
Key Step	r =10 : I = {1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =23 : I = {1, 4, 5, 6, 7, 8, 9, 10, 11, 12} r =27 : I = {1, 5, 6, 7, 8, 9, 10, 11, 12} r =37 : I = {1, 6, 7, 8, 9, 10, 11, 12} r =47 : I = {6, 7, 8, 9, 10, 11, 12} r =75 : I = {6, 7, 8, 9, 10, 11, 12} (FINAL)											
Policy	P-1	P-2	P-3	P-4	P-5	P-6	P-7	P-8	P-9	P-10	P-11	P-12
Means	0.509	0.623	0.561	0.547	0.522	0.474	0.476	0.477	0.473	0.473	0.472	0.470
S.D.	0.0456	0.0422	0.0425	0.0368	0.0337	0.0455	0.0455	0.0457	0.0443	0.0405	0.0438	0.0426
Samples	47	10	23	27	37	75	75	75	75	75	75	75
	669 (in total)											

From the results, it can be concluded that critical threshold (S_c) has less impact than admitting threshold (S_a). With the same critical threshold (S_c = 4), the admitting threshold of (S_a = 11) achieves less mortality for the most policies (only except for P-2, P-3 and P-5). The reason is the lower threshold value (S_a = 10) prevents many moderate injured patients from being admitted so that they receive no treatments but only deteriorate as is, which increases the mortality. In contrast the higher threshold value (S_a = 12) allows any patients being admitted into hospitals (since the maximal possible RPM = 12), which increases the work loads of the hospitals and may waste the capacity since it admits those mild patients unnecessarily, while leaving those indeed severe patients untreated for a longer time (during the waiting and transition), and thus increasing the mortality as well.

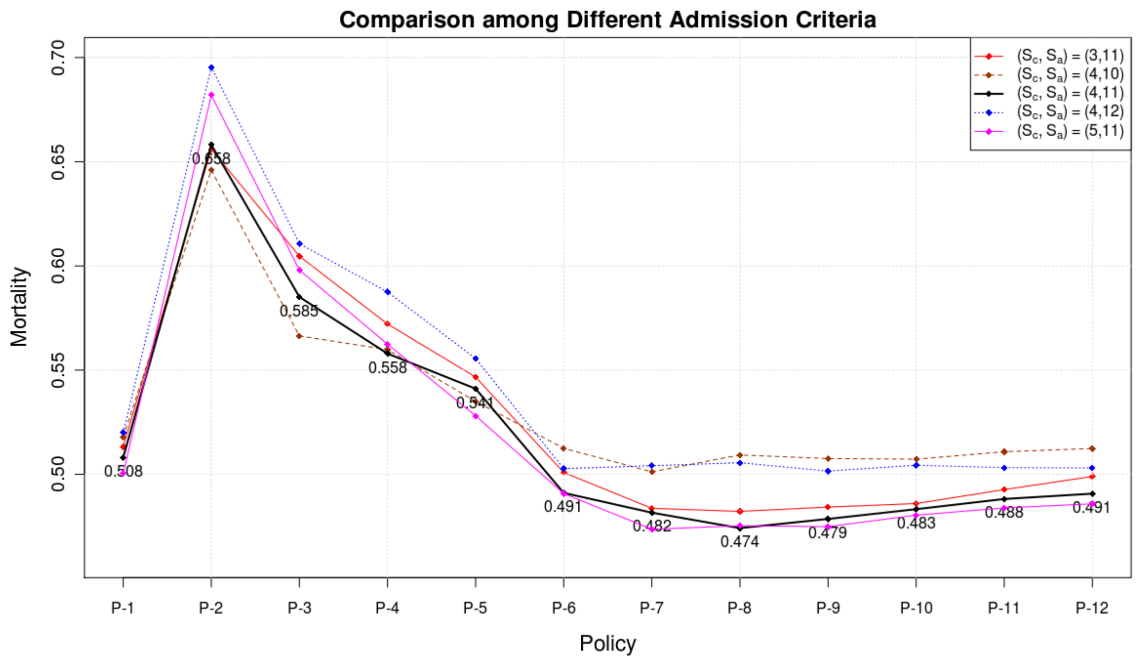


Figure 6.9: Comparison among different admission thresholds used by hospitals

Besides that, the higher admitting threshold ($S_a = 12$) makes the mortality of P-1 less distinct to the policies of P-6 $\sim$ P-12, which explains why P-1 was included in the final select set. For the same reason, the lower admitting threshold ($S_a = 10$) kept both P-1 and P-5 in the final selection.

6.5 SUMMARY

This chapter combines the best-subset selection procedure with the agent-based simulation model developed in previous chapters to specifically solve the problem of the best-subset selection of evacuation policy to respond a mass casualty incident. Computational results of a case study demonstrate the effectiveness of the methodology. In addition, a series of sensitivity analysis experiments were performed to study the impacts of different factors on the policy selection results, as well as the number of replications needed, which provides us more insights into the response process and helps identify the impact factors that are the most important to the response.

7.0 CONCLUSIONS AND FUTURE RESEARCH

The major contribution of this dissertation is an extension to ranking-and-selection (R&S) methods for use in a new problem type, the best-subset selection problem. This enables decision makers to use stochastic simulation models to find a best-subset of policies that optimize a pre-defined performance measure from a finite number of alternatives, from which they can make a final decision based on tradeoffs between performance and other criteria such as resource availability, physical and human resources constraints, or other policy reasons. In addition, this dissertation details the modeling of a complex system and policy alternatives to be implemented in a mass casualty incident (MCI) response. This chapter summarizes the major developments and results reached through the research. It also outlines several possible directions to extend current research work, with respect to the simulation modeling and output analysis correspondingly.

7.1 SUMMARY AND CONCLUSIONS

The first part of this dissertation presents the existing difficulties in selecting good control policies for a complex system. This problem could be considered as a specific case of so-called "wicked problem" [170]. A wicked problem is difficult to address because of the complex interdependencies in the system. Using traditional analysis by decomposition approaches, the effort to solve one aspect of a wicked problem may reveal or create other problems. In order to tackle such a problem, the first requirement is to model the complex system in a comprehensive and effective way, so that sufficient detail is captured in order to best identify underlying complex interrelationships.

As a specific example of a complex system, a mass casualty incident (MCI) response system containing multiple participants was selected to test our methodology by selecting the best-subset of evacuation policies that lead to minimal mortality. After a thorough literature review, we proposed that such a complex system could be appropriately described by a complex adaptive system (CAS), since the system operates according to the actions and interactions of the system members who are equipped with localized decision-making capabilities, rather than operates in an integrated way with a single decision maker making system-wide decisions. For CAS modeling, agent-based simulation (ABS) is an effective tool since it allows researchers to reproduce and investigate processes observed in reality and avoids imposing overly simplified assumptions or constructing too many conditional probability mass functions in modeling decision uncertainty in CAS. Using ABS, complex operational processes can be analyzed by a dynamic system with artificial agents representing real world objects. This dynamic system evolves in iteration steps over time, in which the artificial agents communicate with each other within certain contexts, so that ABS can properly capture individual agents' actions and interactions that determine the full system behavior.

In this dissertation, an ABS model was developed to investigate the performance of a MCI response system under different evacuation policies, while various artificial agents were created to represent different participants in the system, such as injured casualties, on-site EMTs, ambulances, incident command, and hospitals. A divide-and-conquer strategy was employed to build the model. The MCI response system was first decomposed into three interrelated functional sub-systems, then each sub-system was built individually, and finally all three sub-systems were integrated together to form the whole response system. Such a development procedure follows the "bottom-to-up" principle and has been demonstrated to work well in modeling such a complex system.

Chapter-3 details the characteristics of the ABS model. Some highlights include the GIS integration and the definitions of three new generic agent types (Indicator, Performer, and Commander). The new generic agent types extend the agent definition architecture and provide a set of prototype templates so that concrete agents with specific functional roles can be derived from them. Such an implementation achieves better code reusability and makes it easier to add new agent instances into the model.

The ABS model enables researchers to study how different factors affect overall mortality – the system performance measure. Among these factors, one of the most important is victim degradation. In this dissertation, two well-known degradation models, a proportional-hazard based model and Sacco's RPM based model, were implemented and compared with a goal to find whether or not different casualty degradation models would impact the final policy selection. The experimental results show that casualty degradation has significant impact on the overall mortality. However, the results also suggest it is possible to exchange these two models without dramatically affecting the policy selection results.

Specifically, although the estimated mortality estimates are sensitive to different degradation models, the best-subset selection policy is relatively insensitive to the variation in degradation models – similar selection results can be obtained by either Sacco's RPM-based model or proportional-hazard based models with a range of parameters. This outcome is significant since we are trying to select the best sets of policies in the face of uncertain assumptions about patient deterioration.

Although the simulation model provides a good test bed for policy evaluation and comparison, the model itself cannot help decision makers to select the best policy (or set of policies). Such a problem could be addressed by comparing the outputs from different alternatives strategically. There are two reasons motivating us to research statistical analysis techniques. One is because of the randomness inherent in the system, multiple replications are required to run the simulation model for each policy and proper statistical analysis techniques are needed to analyze the results in order to reach a statistically confident conclusion. The other reason is because ABS is a computationally intensive procedure, which may consume lots of computational time and other resources unnecessarily without appropriate simulation control techniques.

Ranking-and-selection (R&S) procedures are especially designed to fulfill these requirements. Chapter-4 illustrates how the R&S procedures can be used in conjunction with the ABS model to select the best evacuation policy. Two credited R&S procedures (the Rinott and the KN procedures) are implemented, and their efficiencies are compared to each other.

Although existing R&S procedures work well in selecting a single best alternative, they are deficient in selecting a best-subset that contains all alternatives that are "close enough" to the best with a pre-specified statistical confidence level, which is a format that may be the most useful to decision makers who are unwilling to accept a single "answer" generated by a computer algorithm. To address this problem, Chapter-5 develops a new best-subset selection (BSS) procedure. As a fully sequential procedure, BSS continuously compares results from different alternatives in each simulation round, eliminates those inferior alternatives during the screening stages, and stops the simulation timely when enough evidence has been accumulated to reach a reliable conclusion. The performance comparison results between the BBS and a MCB-based procedure show that the BSS exceeds the MCB-based procedure in both efficiency and precision. In other words, the BSS achieves a better probability of correct selection with fewer samples in comparing with the MCB-based procedure.

It should be noted that the BSS procedure is different from Simon's "satisficing" strategy [171], which attempts to meet criteria for adequacy, rather than to identify an optimal solution. Although their effects look similar (the decision maker does not seek the true optimal, but only requires that the final selection is "good enough"), the objective of satisficing is to find acceptable (or feasible) alternatives by comparing alternatives with a pre-specified standard. In contrast, the BSS provides a guarantee that all alternatives in the selection set are close to the optimal, which is a stronger requirement, but certainly would meet the satisficing criteria.

The development of the BSS procedure is the major theoretical contribution of this research. It provides an extension of R&S for decision makers to obtain a relatively small subset of best alternatives in an effective and efficient way, so that they can choose their final decision among the selected alternatives based on other criteria not in the model, such as social or political feasibility. The BSS procedure also provides an effective control mechanism to run simulated scenarios for a number of replications as necessary.

Chapter-7 applies the BSS procedure to the MCI ABS model to address the subset selection problem for best evacuation policies. The experimental results confirm that our methodology works well by selecting the evacuation policies leading to minimal mortality in an efficiently manner. In the sensitivity analysis, we conducted extensive computational experiments to test our method under different system configurations.

An interesting phenomenon that can be observed from the experimental results is that the random dispatching policy (P-1) achieves quite low estimated mortality. Compared with P-1, the estimated mortality under P-2, P-3, P-4 and P-5 is generally higher, and the mortality under P-6 to P-12 is lower. These policies differ on the information used in decision making. For examples, P-1 collects no information but just dispatches casualties randomly (and thus evenly) to the various hospitals. Starting from P-2, additional information is added into the decision criteria set. The decision criteria of P-2 include lengths of arrival waiting queues in hospital emergency departments and distances from hospitals to the incident site. P-3, P-4 and P-5 adds other waiting queues, the number of patients currently en-route to the hospital and the available capacity in emergency departments into the decision criteria set, respectively. However, this additional information still results in lower performance than dispatching patients to hospitals randomly under P-1. P-6 to P-12 adds the number of available beds in each hospital to the information used and these policies are superior to P-1.

From the analysis we can see that too little information has no benefit but may have harmful impact on decision making instead (e.g., P-2 to P-5 versus P-1). In order to reduce mortality, the number of available beds at each hospital is an important piece of information to consider in decision making (e.g., P-6 to P-12 versus P-1). In addition, the result that P-1 achieves quite low mortality suggests that it is wise to utilize all available facility resources as much as possible in a MCI response, since it helps balance the load among hospitals and avoids unnecessary waiting and transferring for patients. Such a finding provides a useful guidance for emergency managers, especially when it is difficult to collect information under certain extreme conditions (e.g., after a severe earthquake). In summary, these experiments provide researchers with an important insight into the MCI response problem, and illustrate the importance of different impact factors on mortality. Hence, this provides a good foundation for the subsequent research on policy optimization by identifying the important factors that should be given priority.

As a concluding remark, this research brings together the fields of MCI response system analysis, agent-based simulation, and statistical output comparison and simulation control techniques. The combination of these fields itself is an intellectual contribution. In addition, there are several derived practical and theoretical contributions within each field. Practical implementation of a large-scale model using these techniques establishes a methodology template suitable for reuse to simulate other complex systems. Theoretical development of new best-subset selection procedure, combined with the practical aspects of the implementation, enables managers in finding the best alternatives from competing systems in an effective and efficient manner. In a unified sense, this research enables enhanced use of OvS in analysis of complex systems.

7.2 LIMITATIONS DISCUSSION AND FUTURE RESEARCH DIRECTIONS

There are several directions that our research can be extended, which can be classified into two categories, which are simulation modeling and output analysis respectively.

7.2.1 Possible extensions on simulation modeling

One advantage of simulation is it provides a convenient way for researchers to check the impacts of different factors on the system performance, where a factor may belong to either one of the two categories: quantitative measures of the situation (environment), and action rules responding to specific situations. For this research, the examples of the first category are the total number of injured casualties, the percentage of specialized patients, etc. An instance of the second category is the admission criteria used in hospitals. So the first possible extension is to build a decision support tool based on the ABS model. The tool could be utilized to evaluate the impact of the factors that are considered to have non-negligible impacts on the outcome, prioritize the factors according to their importance, and identify the key factors that really matter to the system. Such a tool could help decision makers focus their attentions on a few important factors and avoid wasting time and effort on issues that are not significant. In other words, the tool can effectively reduce the dimensions of decision space for decision makers in policy making or optimization, so that it becomes easier for them to gain situational awareness and to find optimal solutions in an efficient way.

In our research context, a policy is essentially a decision tree, which consists of a series of rules that prescribe in detail what kind of actions should be taken when certain type of situations are encountered. In the current implementation, the policies or decision trees are pre-defined and all action rules are hard-coded in the program, which makes it difficult to test a new policy or a variant of an existing policy after certain modifications. Thus, the second possible extension of the simulation model is to develop a set of flexible schemes to store the policy trees (or rule sets). A potential direction could be the development of a "rule database" (rule-base), which stores pairs of conditions and consequences to represent action rules. By adding/modifying/activating/deactivating certain records, it is easy to change the policy in use. And the decisions of agents can be obtained by logically inquiring the rule-base iteratively via an interface between the rule-base and the simulator, which might be a foundation to create intelligent agents in future. The intelligent agents can autonomously learn from previous experiences and choose the best rule to execute, which would provide another effective way to improve or optimize existing policies.

Compared with other models appearing in the literature, our MCI response simulation model is more comprehensive since it covers almost all aspects in a MCI response system, from up-stream (the incident scene) to down-stream (hospitals). However, due to the limitations on time and resources, our simulation model only implemented a very simple hospital model, which simulates the major processes in the emergency department and simplifies other departments (ICU, General Wards, etc). It could be argued that such a simplified hospital model may not be adequate to reflect the complex structure and various processes in hospitals, which may significantly affect simulation outcomes. Therefore it may be fruitful to develop a more complete hospital model to replace the current simple one, and the complete hospital model could simulate the various medical facilities in a more precise way so to obtain more accurate simulation results.

An important characteristic of the current simulation model is the GIS integration. It provides an interface to import processed GIS data in constructing the urban transportation network, which facilitates practitioners considering the effects of different traffic to the casualty evacuations. In addition, it also enables displaying the ongoing evacuation status on a GIS view dynamically along with the simulation running, which gives a more direct illustration to the practitioners about the evacuation process and can help them identify potential problems. Currently, all GIS data are read from a static database, which is valid for coordination-related information, such as hospital locations, road connections, etc. However, static data are inadequate in describing the traffic status. For example, currently we only use static values of average speed on each route segment to depict a specific traffic pattern that is unchangeable during the simulation. For a more practical simulation, a useful extension is to integrate the real-time traffic data from the GPS system into the simulation to better reflect the actual situation. Such an extension may also help in converting the simulation model to a real-time decision support tool for enabling the incident managers in finding an optimal strategy for large-scale evacuations under disasters.

Besides these, with respect to the incident response simulation model, other possible research directions include multi-scene response planning via simulation, post-hospital transfer and operations, etc. For modeling such a complex system, a large number of open questions are left for the successor researchers to complete.

7.2.2 Possible directions on output analysis

In the current research, we are only considering one performance measure – mortality – as the single criterion in choosing the best response policies. But in practice, it is very common to observe situations that intend to use multiple performance measures instead of a single measure as the decision criteria. For example, in MCI response, the morbidity of injured patients could become another measure to evaluate the response performance. One of the advantages of the BSS procedure is to enable decision makers to select the "good enough" alternatives based on the most important criteria first, then apply other criteria to make their final decision from the selected subset. For instance using our case, researchers can use the BSS to select the best policy subset leading to the minimal mortality (the first criterion), then choose the policies with the lowest morbidities from the subset (the second criterion).

For multiple criteria selection problem, one possible direction is to extend the BSS procedure to select a subset consisting of the policies having Pareto-optimality instead of only being outstanding in one performance measure. And the other possible direction is to filter out policies based on certain screening criteria, and then apply the BSS to the remaining alternatives based on critical criteria. In other words, the first step is to identify alternatives that meet a performance standard on criteria-1(2,3, ...), then select the best subset based on criteria-C. Note that usually criteria-C is the most critical one, and the screening criteria-1(2,3, ...) are of the form "must be better than X(Y,Z, ...) in performance measure-1(2,3, ...)".

In terms of R&S analysis technique research, an important extension is to improve the efficiency of the BSS procedure. As we may have observed in Chapter-5, although the BSS procedure offers improved performance over the MCB-based procedure in selecting a best-subset, it is still conservative in the sense that it samples more than strictly necessary and over-delivers on the target of probability of correct selection. The possible source of conservation may come from the application of the Bonferroni inequality to control the overall incorrect selection probability, which is usually unnecessary except for so-called slippage configurations. A potential direction to address this problem is to employ Bayesian-based approaches in conjunction with the Frequentist-based BSS procedure to improve its efficiency. Besides that, another possible direction to improve the efficiency is to develop certain parallel computing techniques to distribute the computation work onto different computer nodes, which could also help in obtaining the selection result in an efficient manner.

APPENDIX

ACRONYMS

ABS	Agent-based Simulation
ALS	Advanced Life Support
BLS	Basic Life Support
BSS	Best-subset Selection
CAS	Complex Adaptive System
CRN	Common Random Numbers
DES	Discrete-event Simulation
DHS	Department of Homeland Security
ED	Emergency Department (in hospital)
EMS	Emergency Medical Services
EMT	Emergency Medical Technician
GIS	Geographical Information System
GPS	Global Positioning System
GW	General Wards (in hospital)
ICU	Intensive Care Unit (in hospital)

IED	Improvised Explosive Device
IZS	Indifference-zone Selection
KN	A fully-sequential R&S procedure developed by Kim and Nelson (2001)
MCA	All pairwise Multiple Comparisons
MCB	Multiple Comparisons with the Best
MCC	Multiple Comparisons with a Control
MCI	Mass Casualty Incident
MCP	Multiple Comparison Procedures
NIMS	National Incident Management System
NRF	National Response Framework
OCBA	Optimal Computing Budget Allocation
OR	Operating Room (in hospital)
OvS	Optimization via Simulation
PCS	Probability of Correct Selection
R&S	Ranking and Selection
Repast	Recursive Porous Agent Simulation Toolkit
RPM	The sum of coded values for Respiratory rate, Pulse rate, and best Motor response, which is used to score victim severity and to predict survivability.
RSM	Response Surface Methods
SC	Slippage Configuration
SS	Subset Selection

BIBLIOGRAPHY

[1] F. Nilsson and V. Darley, "On complex adaptive systems and agent-based modelling for improving decision-making in manufacturing and logistics settings: Experiences from a packaging company," *International Journal of Operations & Production Management*, vol. 26, no. 12, pp. 1351–1373, 2006.

[2] E. Bonabeau, "Agent-based modeling: Methods and techniques for simulating human systems," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. Suppl 3, pp. 7280–7287, 2002.

[3] W. J. Sacco, D. M. Navin, R. K. Waddell, K. E. Fiedler, W. B. Long, and R. F. Buckman, "A new resource-constrained triage method applied to victims of penetrating injury," *The Journal of Trauma: Injury, Infection, and Critical Care*, vol. 63, no. 2, pp. 316–325, 2007.

[4] M. W. Chaffee and M. M. McNeill, "A model of nursing as a complex adaptive system," *Nursing Outlook*, vol. 55, no. 5, pp. 232–241.e3, 2007.

[5] N. R. Jennings, "On agent-based software engineering," *Artificial Intelligence*, vol. 117, no. 2, pp. 277–296, 2000.

[6] L. V. Snyder and Z.-J. M. Shen, "Managing disruptions to supply chains," *The Bridge*, vol. 36, pp. 39–45, 2006.

[7] J. Clausen, A. Larsen, J. Larsen, and N. J. Rezanova, "Disruption management in the airline industry–concepts, models and methods," *Computers & Operations Research*, vol. 37, no. 5, pp. 809–821, 2010.

[8] G. Yu, M. Arguello, G. Song, S. M. McCowan, and A. White, "A new era for crew recovery at continental airlines," *Interfaces*, vol. 33, no. 1, pp. 5–22, 2003.

[9] R. McGarvey, R. Tripp, R. Rue, T. Lang, J. Sollinger, W. Conner, and L. Luangkesorn, *Global Combat Support Basing: Robust Prepositioning Strategies for Air Force War Reserve Materiel*. RAND Corporation, 2010.

[10] M. Schrage, *Serious play: how the world's best companies simulate to innovate*. Harvard Business Press, 2000.

[11] P. D. Wright, M. J. Liberatore, and R. L. Nydick, “A survey of operations research models and applications in homeland security,” *INTERFACES*, vol. 36, no. 6, pp. 514–529, 2006.

[12] R. C. Larson, M. D. Metzger, and M. F. Cahn, “Responding to emergencies: Lessons learned and the need for analysis,” *INTERFACES*, vol. 36, no. 6, pp. 486–501, 2006.

[13] G. Bimonte, *Predictability of SOC Systems. Technological Extreme Events*, pp. 221–234. New Economic Windows, Springer Milan, 2010.

[14] M. J. North and C. M. Macal, *Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation*. Oxford University Press, USA, 2007.

[15] S. Wu, *AGENT-BASED DISCRETE EVENT SIMULATION MODELING AND EVOLUTIONARY REAL-TIME DECISION MAKING FOR LARGE-SCALE SYSTEMS*. PhD thesis, University of Pittsburgh, 2009.

[16] B. Y. Lee, R. R. Bailey, K. J. Smith, R. R. Muder, E. S. Strotmeyer, G. J. Lewis, P. J. Ufberg, Y. Song, and L. H. Harrison, “Universal methicillin-resistant staphylococcus aureus (mrsa) surveillance for adults at hospital admission: an economic model and analysis,” *Infection Control and Hospital Epidemiology: The Official Journal of the Society of Hospital Epidemiologists of America*, vol. 31, no. 6, pp. 598–606, 2010.

[17] B. Y. Lee, S. T. Brown, P. C. Cooley, R. K. Zimmerman, W. D. Wheaton, S. M. Zimmer, J. J. Grefenstette, T.-M. Assi, T. J. Furphy, D. K. Wagener, and D. S. Burke, “A computer simulation of employee vaccination to mitigate an influenza epidemic,” *American Journal of Preventive Medicine*, vol. 38, no. 3, pp. 247–257, 2010.

[18] B. Y. Lee, S. T. Brown, G. W. Korch, P. C. Cooley, R. K. Zimmerman, W. D. Wheaton, S. M. Zimmer, J. J. Grefenstette, R. R. Bailey, T.-M. Assi, and D. S. Burke, “A computer simulation of vaccine prioritization, allocation, and rationing during the 2009 h1n1 influenza pandemic,” *Vaccine*, vol. 28, no. 31, pp. 4875–4879, 2010.

[19] B. Zimmerman, Y. Wang, D. Nawn, B. Kuhlman, L. Luangkesorn, K. Sochats, C. Balaban, and B. Bidanda, “Dynamic model generation for agent-based emergency response simulation.,” in *ESRI International User Conference (2010)*, 2010.

[20] I. P. McCarthy, C. Tsinopoulos, P. Allen, and C. Rose-Anderssen, “New product development as a complex adaptive system of decisions,” *Journal of Product Innovation Management*, vol. 23, no. 5, pp. 437–456, 2006.

[21] E. Ahmed, A. S. Elgazzar, and A. S. Hegazi, “An overview of complex adaptive systems,” *nlin/0506059*, 2005.

[22] S. A. Levin, “Ecosystems and the biosphere as complex adaptive systems,” *Ecosystems*, vol. 1, no. 5, pp. 431–436, 1998.

[23] L. Fleming and O. Sorenson, “Technology as a complex adaptive system: evidence from patent data,” *Research Policy*, vol. 30, no. 7, pp. 1019–1039, 2001.

[24] H. Akkermans, “Emergent supply networks: system dynamics simulation of adaptive supply agents,” in *System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on*, p. 11 pp., 2001.

[25] T. Y. Choi, K. J. Dooley, and M. Rungtusanatham, “Supply networks and complex adaptive systems: control versus emergence,” *Journal of Operations Management*, vol. 19, no. 3, pp. 351–366, 2001.

[26] G. Li, P. Ji, L. Sun, and W. Lee, “Modeling and simulation of supply network evolution based on complex adaptive system and fitness landscape,” *Computers & Industrial Engineering*, vol. 56, no. 3, pp. 839–853, 2009.

[27] A. Nair, R. Narasimhan, and T. Y. Choi, “Supply networks as a complex adaptive system: Toward simulation-based theory building on evolutionary decision making,” *Decision Sciences*, vol. 40, no. 4, pp. 783–815, 2009.

[28] S. Pathak and D. Dilts, “Simulation of supply chain networks using complex adaptive system theory,” in *Engineering Management Conference, 2002. IEMC ’02. 2002 IEEE International*, vol. 2, pp. 655–660 vol.2, 2002.

[29] M. J. Mauboussin, “Revisiting market efficiency: The stock market as a complex adaptive system,” *Journal of Applied Corporate Finance*, vol. 14, no. 4, pp. 47–55, 2002.

[30] M. Boisot and J. Child, “Organizations as adaptive systems in complex environments: The case of china,” *Organization Science*, vol. 10, no. 3, pp. 237–252, 1999.

[31] “Complex adaptive system - wikipedia, the free encyclopedia wikipedia.” http://en.wikipedia.org/wiki/Complex_adaptive_system. Accessed Feb 21, 2011.

[32] W. B. Rouse, “Health care as a complex adaptive system: Implications for design and management,” *The Bridge*, pp. 17–25, 2008.

[33] J. H. Holland, “Complex adaptive systems,” *Daedalus*, vol. 121, no. 1, pp. 17–30, 1992.

[34] D. Chu, R. Strand, and R. Fjelland, “Theories of complexity,” *Complexity*, vol. 8, no. 3, pp. 19–30, 2003.

[35] A. Nossiter, “New orleans population is reduced nearly 60%,” *The New York Times*, 2006.

[36] T. WASHINGTON, “Four years after katrina, new orleans still struggling to recover from the storm,” *democracynow.org*, 2009.

[37] C. Cooper and R. Block, *Disaster: Hurricane Katrina and the Failure of Homeland Security*. Times Books, 1st ed., 2006.

[38] Department of Homeland Security (DHS), *National Preparedness Guidelines*. 2007.

[39] Department of Homeland Security (DHS), *National Incident Management System*. 2004.

[40] Department of Homeland Security (DHS), *National Response Framework*. 2007.

[41] Office of Emergency Management, City & County of Boulder, Colorado, "Emergency operations plan (boulderoem.com)." http://boulderoem.com/plan.html. [Accessed: Feb 21, 2011].

[42] C. Toregas, R. Swain, C. ReVelle, and L. Bergman, "The location of emergency service facilities," *OPERATIONS RESEARCH*, vol. 19, no. 6, pp. 1363–1373, 1971.

[43] J. R. Weaver and R. L. Church, "A median location model with nonclosest facility service," *TRANSPORTATION SCIENCE*, vol. 19, no. 1, pp. 58–74, 1985.

[44] V. Marianov and C. Revelle, "The queuing probabilistic location set covering problem and some extensions," *Socio-Economic Planning Sciences*, vol. 28, pp. 167–178, 1994.

[45] D. A. Schilling, C. Revelle, J. Cohon, and D. J. Elzinga, "Some models for fire protection locational decisions," *European Journal of Operational Research*, vol. 5, no. 1, pp. 1–7, 1980.

[46] C. Revelle, J. Schweitzer, and S. Snyder, "The maximal conditional covering problem," *Location Science*, vol. 5, p. 61, 1997.

[47] M. A. Badri, A. K. Mortagy, and C. A. Alsayed, "A multi-objective model for locating fire stations," *European Journal of Operational Research*, vol. 110, no. 2, pp. 243–260, 1998.

[48] R. Larson, "A hypercube queuing model for facility location and redistricting in urban emergency services," *Computers & Operations Research*, vol. 1, pp. 67–95, 1974.

[49] F. Mendonca and R. Morabito, "Analysing emergency medical service ambulance deployment on a brazilian highway using the hypercube model," *The Journal of the Operational Research Society*, vol. 52, no. 3, pp. 261–270, 2001.

[50] H. K. Rajagopalan, C. Saydam, and J. Xiao, "A multiperiod set covering location model for dynamic redeployment of ambulances," *Computers & Operations Research*, vol. 35, no. 3, pp. 814–826, 2008.

[51] R. A. Takeda, J. A. Widmer, and R. Morabito, "Analysis of ambulance decentralization in an urban emergency medical service using the hypercube queueing model," *Computers & Operations Research*, vol. 34, no. 3, pp. 727–741, 2007.

[52] A. Haghani, Q. Tian, and H. Hu, "Simulation model for real-time emergency vehicle dispatching and routing," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1882, pp. 176–183, 2004.

[53] G. Barbarosoğlu and Y. Arda, "A two-stage stochastic programming framework for transportation planning in disaster response," *The Journal of the Operational Research Society*, vol. 55, no. 1, pp. 43–53, 2004.

[54] W. Yi and L. Özdamar, "A dynamic logistics coordination model for evacuation and support in disaster response activities," *European Journal of Operational Research*, vol. 179, no. 3, pp. 1177–1193, 2007.

[55] S. I. Chien and V. V. Korikanthimath, "Analysis and modeling of simultaneous and staged emergency evacuations," *Journal of Transportation Engineering*, vol. 133, no. 3, pp. 190–197, 2007.

[56] J. Goldberg, R. Dietrich, J. M. Chen, M. Mitwasi, T. Valenzuela, and E. Criss, "A simulation model for evaluating a set of emergency vehicle base locations: Development, validation, and usage," *Socio-Economic Planning Sciences*, vol. 24, no. 2, pp. 125–141, 1990.

[57] L. J. Shuman, H. Wolfe, and M. J. Gunter, "Ruralsim: the design and implementation of a rural ems simulator," *Journal of the Society for Health Systems*, vol. 3, no. 3, pp. 54–71, 1992.

[58] A. Hirshberg, M. Stein, and R. Walden, "Surgical resource utilization in urban terrorist bombing: a computer simulation," *The Journal of Trauma*, vol. 47, no. 3, pp. 545–550, 1999.

[59] S. Su and C.-L. Shih, "Modeling an emergency medical services system using computer simulation," *International Journal of Medical Informatics*, vol. 72, no. 1-3, pp. 57–72, 2003.

[60] G. R. Hung, S. R. Whitehouse, C. ONeill, A. P. Gray, and N. Kissoon, "Computer modeling of patient flow in a pediatric emergency department using discrete event simulation," *Pediatric Emergency Care*, vol. 23, no. 1, pp. 5–10, 2007.

[61] V. Boginski, I. Mun, Y. Wu, K. Mason, and C. Zhang, "Simulation and analysis of hospital operations and resource utilization using rfid data," in *RFID, 2007. IEEE International Conference on*, pp. 199–204, 2007.

[62] A. Kolker, "Process modeling of emergency department patient flow: Effect of patient length of stay on ed diversion," *J. Med. Syst.*, vol. 32, no. 5, pp. 389–401, 2008.

[63] D. K. Carley, D. D. Fridsma, D. E. Casman, D. J. Chang, D. B. Kaminsky, D. Nave, and A. Yahja, "Biowar: Scalable multi-agent social and epidemiological simulation of bioterrorism events," *IN PROCEEDINGS OF NORTH AMERICAN ASSOCIATION FOR COMPUTATIONAL SOCIAL AND ORGANIZATIONAL SCIENCE (NAACSOS) CONFERENCE 2004*, 2003.

[64] G. Narzisi, J. Mincer, S. Smith, and B. Mishra, *Resilience in the Face of Disaster: Accounting for Varying Disaster Magnitudes, Resource Topologies, and (Sub)Population Distributions in the PLAN C Emergency Planning Tool*, pp. 433–446. 2007.

[65] D. Massaguer, V. Balasubramanian, S. Mehrotra, and N. Venkatasubramanian, "Multi-agent simulation of disaster response," *IN ATDM WORKSHOP IN AAMAS 2006*, 2006.

[66] K. M. Khalil, M. Abdel-Aziz, T. T. Nazmy, and A.-B. M. Salem, "Multi-agent crisis response systems - design requirements and analysis of current systems," tech. rep., http://arxiv.org/abs/0903.2543, 2009.

[67] X. Chen and F. B. Zhan, "Agent-based modelling and simulation of urban evacuation: relative effectiveness of simultaneous and staged evacuation strategies," *Journal of the Operational Research Society*, vol. 59, no. 1, pp. 25–33, 2008.

[68] T. Schoenharl, Z. Zhai, R. McCune, A. Pawling, and G. Madey, "Design and implementation of an agent-based simulation for emergency response and crisis management," in *Proceedings of the 2009 Spring Simulation Multiconference*, SpringSim '09, pp. 18:1–18:12, Society for Computer Simulation International, 2009.

[69] M. J. North, N. T. Collier, and J. R. Vos, "Experiences creating three implementations of the repast agent modeling toolkit," *ACM Trans. Model. Comput. Simul.*, vol. 16, no. 1, pp. 1–25, 2006.

[70] B. Y. Lee, S. T. Brown, P. Cooley, M. A. Potter, W. D. Wheaton, R. E. Voorhees, S. Stebbins, J. J. Grefenstette, S. M. Zimmer, R. Zimmerman, T.-M. Assi, R. R. Bailey, D. K. Wagener, and D. S. Burke, "Simulating school closure strategies to mitigate an influenza epidemic," *Journal of public health management and practice : JPHMP*, vol. 16, no. 3, pp. 252–261, 2010.

[71] S. Zhu, J. Abraham, S. Paul, M. Reddy, J. Yen, M. Pfaff, and C. DeFlitch, *R-CAST-MED: Applying Intelligent Agents to Support Emergency Medical Decision-Making Teams*, pp. 24–33. 2007.

[72] A. Daknou, H. Zgaya, S. Hammadi, and H. Hubert, "Toward a multi-agent model for the care of patients at the emergency department," in *Proceedings of the 10th WSEAS international conference on Mathematical methods, computational techniques and intelligent systems*, pp. 264–269, World Scientific and Engineering Academy and Society (WSEAS), 2008.

[73] J. R. Clymer, "Simulation-based engineering of complex adaptive systems," *SIMULATION*, vol. 72, no. 4, pp. 250–260, 1999.

[74] A. Wildberger, "Complex adaptive systems: concepts and power industry applications," *Control Systems Magazine, IEEE*, vol. 17, no. 6, pp. 77–88, 1997.

[75] M. Luck, P. McBurney, and C. Preist, *Agent technology: enabling next generation computing (a roadmap for agent based computing)*. AgentLink / University of Southampton, 2003.

[76] C. M. Macal and M. J. North, "Tutorial on agent-based modelling and simulation," *Journal of Simulation*, vol. 4, no. 3, pp. 151–162, 2010.

[77] T.-T. Tu, *The Development of a Hybrid Simulation Modelling Approach Based on Agents and Discrete-Event Modelling*. PhD thesis, UNIVERSITY OF SOUTHAMPTON, 2008.

[78] C. M. Macal and M. J. North, "Tutorial on agent-based modeling and simulation," in *Proceedings of the 37th conference on Winter simulation*, WSC 2005, pp. 2–15, Winter Simulation Conference, 2005.

[79] M. Luck, P. McBurney, O. Shehory, and S. Willmott, *Agent technology: computing as interaction (a roadmap for agent based computing)*. AgentLink, 2005.

[80] C. Linard, N. Ponçon, D. Fontenille, and E. F. Lambin, "A multi-agent simulation to assess the risk of malaria re-emergence in southern france," *Ecological Modelling*, vol. 220, no. 2, pp. 160–174, 2009.

[81] B. Dubiel and O. Tsimhoni, "Integrating agent based modeling into a discrete event simulation," in *Proceedings of the 37th conference on Winter simulation*, pp. 1029–1037, Winter Simulation Conference, 2005.

[82] Y. J. Son, R. A. Wysk, and A. T. Jones, "Simulation-based shop floor control: formal model, model generation and control interface," *IIE Transactions*, vol. 35, no. 1, pp. 29–48, 2003.

[83] D. Brown, S. Page, R. Riolo, and W. Rand, "Agent-based and analytical modeling to evaluate the effectiveness of greenbelts," *Environmental Modelling & Software*, vol. 19, no. 12, pp. 1097–1109, 2004.

[84] J. Chu, C. Wang, J. Chen, and H. Wang, "Agent-based residential water use behavior simulation and policy implications: A case-study in beijing city," *Water Resources Management*, vol. 23, no. 15, pp. 3267–3295, 2009.

[85] P. Sridhar, S. Sheikh-Bahaei, S. Xia, and M. Jamshidi, "Multi-agent simulation using discrete event and soft-computing methodologies," in *IEEE International Conference on Systems, Man and Cybernetics*, 2003.

[86] S. Lee, A. Pritchett, and D. Goldsman, "Hybrid agent-based simulation for analyzing the national airspace system," in *Proceedings of the 33nd conference on Winter simulation*, pp. 1029–1036, IEEE Computer Society, 2001.

[87] N. V. Patel, T. Eldabi, and T. M. Khan, "Theory of deferred action: Agent-based simulation model for designing complex adaptive systems," *Journal of Enterprise Information Management*, vol. 23, no. 4, pp. 521–537, 2010.

[88] A. R. Pritchett, S. Lee, D. Huang, and D. Goldsman, "Hybrid-system simulation for national airspace system safety analysis," in *Proceedings of the 32nd conference on Winter simulation*, pp. 1132–1142, Society for Computer Simulation International, 2000.

[89] A. R. Pritchett, D. Goldsman, C. Alexopoulos, R. M. Fujimoto, and M. L. Loper, "Agent-based simulations using human performance models for national airspace system risk assessment," Tech. Rep. School of Aerospace Engineering ; Project no. E-16-X84, Georgia Institute of Technology, 2009.

[90] C. M. Macal and M. J. North, "Tutorial on agent-based modeling and simulation part 2: how to model with agents," in *Proceedings of the 38th conference on Winter simulation*, pp. 73–83, Winter Simulation Conference, 2006.

[91] S. F. Railsback, S. L. Lytinen, and S. K. Jackson, "Agent-based simulation platforms: Review and development recommendations," *SIMULATION*, vol. 82, no. 9, pp. 609–623, 2006.

[92] R. Allan, "Survey of agent based modelling and simulation tools," Tech. Rep. oai:epubs.cclrc.ac.uk:work/50398, CSE CSE-HEC STFC, 2009.

[93] K. C. Benson, *Adaptive Control of Large-Scale Simulations*. PhD thesis, Georgia Tech, 2004.

[94] B. Calvez and G. Hutzler, *Automatic Tuning of Agent-Based Models Using Genetic Algorithms*, vol. 3891, pp. 41–57. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.

[95] C. Baydar, "A hybrid parallel simulated annealing algorithm to optimize store performance," in *Workshop on Evolutionary Computing for Optimisation in Industry at the Genetic and Evolutionary Computation Conference (GECCO-2002)*, (New York), July 2002.

[96] T. Laine, *Agent-based model selection framework for complex adaptive systems*. PhD thesis, Indiana University, 2006.

[97] B. Klemens, "Finding optimal agent-based models," *Brookings Center on Social and Economic Dynamics Working Paper*, vol. 49, 2007.

[98] K. Naitoh and T. Terano, "Agent-based modeling of corporate behaviors with evolutionary computation," in *2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation*, vol. 3, pp. 1334–1339, IEEE, July 2003.

[99] G. Dueck and T. Scheuer, "Threshold accepting: A general purpose optimization algorithm appearing superior to simulated annealing," *Journal of Computational Physics*, vol. 90, no. 1, pp. 161–175, 1990.

[100] P. Winker and D. Maringer, *The Threshold Accepting Optimisation Algorithm in Economics and Statistics*, vol. 9, pp. 107–125. Berlin, Heidelberg: Springer Berlin Heidelberg.

[101] M. Dorigo and C. Blum, "Ant colony optimization theory: A survey," *Theoretical Computer Science*, vol. 344, pp. 243–278, Nov. 2005.

[102] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE Computational Intelligence Magazine*, vol. 1, pp. 28–39, Nov. 2006.

[103] S.-H. Kim and B. L. Nelson, "Recent advances in ranking and selection," in *Proceedings of the 39th conference on Winter simulation: 40 years! The best is yet to come*, pp. 162–172, IEEE Press, 2007.

[104] D. Goldsman, B. L. Nelson, and B. Schmeiser, "Methods for selecting the best system," in *Proceedings of the 1991 Winter Simulation Conference, Edited by B. L. Nelson, W. D. Kelton, and G. M. Clark*, pp. 177–186, NJ:IEEE, 1991.

[105] C.-H. Chen, J. Lin, E. Yücesan, and S. E. Chick, "Simulation budget allocation for further enhancing the efficiency of ordinal optimization," *Discrete Event Dynamic Systems*, vol. 10, no. 3, pp. 251–270, 2000.

[106] S. E. Chick and K. Inoue, "New two-stage and sequential procedures for selecting the best simulated system," *Operations Research*, vol. 49, no. 5, pp. 732–743, 2001.

[107] L. J. Hong and B. L. Nelson, "Selecting the best system when systems are revealed sequentially," *IIE Transactions*, vol. 39, no. 7, pp. 723–734, 2007.

[108] Y. Rinott, "On two-stage selection procedures and related probability-inequalities," *Communications in Statistics - Theory and Methods*, vol. 7, no. 8, pp. 799–811, 1978.

[109] S.-H. Kim and B. L. Nelson, "A fully sequential procedure for indifference-zone selection in simulation," *ACM Transactions on Modeling and Computer Simulation*, vol. 11, no. 3, pp. 251–273, 2001.

[110] J. R. Swisher and S. H. Jacobson, "A survey of ranking, selection, and multiple comparison procedures for discrete-event simulation," in *Proceedings of the 31st conference on Winter simulation: Simulation—a bridge to the future - Volume 1*, pp. 492–501, ACM, 1999.

[111] R. E. Bechhofer, "A single-sample multiple decision procedure for ranking means of normal populations with known variances," *The Annals of Mathematical Statistics*, vol. 25, no. 1, pp. 16–39, 1954.

[112] E. J. Dudewicz and S. R. Dalal, "Allocation of observations in ranking and selection with unequal variances," *Sankhyā: The Indian Journal of Statistics, Series B*, vol. 37, no. 1, pp. 28–78, 1975.

[113] B. L. Nelson, J. Swann, D. Goldsman, and W. Song, "Simple procedures for selecting the best simulated system when the number of alternatives is large," *Operations Research*, vol. 49, no. 6, pp. 950–963, 2001.

[114] S.-H. Kim and B. L. Nelson, "On the asymptotic validity of fully sequential selection procedures for steady-state simulation," *Operations Research*, vol. 54, no. 3, pp. 475–488, 2006.

[115] L. J. Hong and B. L. Nelson, "The tradeoff between sampling and switching: New sequential procedures for indifference-zone selection," *IIE Transactions*, vol. 37, no. 7, pp. 623–634, 2005.

[116] T. Osogami, "Finding probably best systems quickly via simulations," *ACM Transactions on Modeling and Computer Simulation*, vol. 19, no. 3, pp. 1–19, 2009.

[117] S. C. Tsai and B. L. Nelson, "Fully sequential selection procedures with control variates," *IIE Transactions*, vol. 42, no. 1, pp. 71–82, 2010.

[118] S. S. Gupta, "On some multiple decision (selection and ranking) rules," *Technometrics*, vol. 7, no. 2, pp. 225–245, 1965.

[119] D. W. Sullivan and J. R. Wilson, "Restricted subset selection procedures for simulation," *Operations Research*, vol. 37, no. 1, pp. 52–71, 1989.

[120] L. W. Koenig and A. M. Law, "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 - Simulation and Computation*, vol. 14, no. 3, pp. 719–734, 1985.

[121] E. J. Chen, "Subset selection procedures," *Journal of Simulation*, vol. 3, no. 4, pp. 202–210, 2009.

[122] J. R. Swisher, S. H. Jacobson, and E. Yücesan, "Discrete-event simulation optimization using ranking, selection, and multiple comparison procedures: A survey," *ACM Trans. Model. Comput. Simul.*, vol. 13, no. 2, pp. 134–154, 2003.

[123] M. Fu, "Optimization via simulation: A review," *Annals of Operations Research*, vol. 53, no. 1, pp. 199–247, 1994.

[124] Y. Benjamini and H. Braun, "John w. tukey's contributions to multiple comparisons," *The Annals of Statistics*, vol. 30, no. 6, pp. 1576–1594, 2002.

[125] C. W. Dunnett, "A multiple comparison procedure for comparing several treatments with a control," *Journal of the American Statistical Association*, vol. 50, no. 272, pp. 1096–1121, 1955.

[126] E. Bofinger and G. J. Lewis, "Two-stage procedures for multiple comparisons with a control.," *American Journal of Mathematical and Management Sciences*, vol. 12, pp. 253–275, 1992.

[127] H. Damerdji and M. K. Nakayama, "Two-stage procedures for multiple comparisons with a control in steady-state simulations," in *Proceedings of the 28th conference on Winter simulation*, pp. 372–375, IEEE Computer Society, 1996.

[128] J. C. Hsu, "Constrained simultaneous confidence intervals for multiple comparisons with the best," *The Annals of Statistics*, vol. 12, no. 3, pp. 1136–1144, 1984.

[129] W.-N. Yang and B. L. Nelson, "Using common random numbers and control variates in multiple-comparison procedures," *Operations Research*, vol. 39, no. 4, pp. 583–591, 1991.

[130] B. L. Nelson and J. C. Hsu, "Control-variate models of common random numbers for multiple comparisons with the best," *Management Science*, vol. 39, no. 8, pp. 989–1001, 1993.

[131] L. Goldsman and B. L. Nelson, "Batch-size effects on simulation optimization using multiple comparisons with the best," in *Proceedings of the 22nd conference on Winter simulation*, pp. 288–293, IEEE Press, 1990.

[132] B. L. Nelson and S. Banerjee, "Selecting a good system: procedures and inference," *IIE - Transactions*, vol. 33, no. 3, p. 149, 2001.

[133] S. S. Gupta and J. C. Hsu, "A computer package for ranking, selection, and multiple comparisons with the best," in *Proceedings of the 16th conference on Winter simulation*, pp. 250–257, IEEE Press, 1984.

[134] B. L. Nelson and F. J. Matejcik, "Using common random numbers for indifference-zone selection and multiple comparisons in simulation," *Management Science*, vol. 41, no. 12, pp. 1935–1945, 1995.

[135] J. R. Swisher and S. H. Jacobson, "Evaluating the design of a family practice healthcare clinic using discrete-event simulation," *Health Care Management Science*, vol. 5, no. 2, pp. 75–88, 2002.

[136] L. Hong and B. Nelson, "A brief introduction to optimization via simulation," in *Winter Simulation Conference (WSC), Proceedings of the 2009*, 2009.

[137] R. H. Myers, A. I. Khuri, and W. H. Carter, "Response surface methodology: 1966-1988," *Technometrics*, vol. 31, no. 2, pp. 137–157, 1989.

[138] S.-H. Kim, "Comparison with a standard via fully sequential procedures," *ACM Transactions on Modeling and Computer Simulation*, vol. 15, no. 2, pp. 155–174, 2005.

[139] S. Andradóttir, D. Goldsman, and S.-H. Kim, "Finding the best in the presence of a stochastic constraint," in *Proceedings of the 2005 Winter Simulation Conference, Edited by M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines*, pp. 732–738, NJ:IEEE, 2005.

[140] S. Andradóttir and S.-H. Kim, "Fully sequential procedures for comparing constrained systems via simulation," *Naval Research Logistics*, vol. 57, no. 5, pp. 403–421, 2010.

[141] D. Batur and S.-H. Kim, "Procedures for feasibility detection in the presence of multiple constraints," in *Proceedings of the 2005 Winter Simulation Conference, Edited by M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines*, pp. 692–698, NJ:IEEE, 2005.

[142] D. Batur and S.-H. Kim, "Finding feasible systems in the presence of constraints on multiple performance measures," *ACM Transactions on Modeling and Computer Simulation*, vol. 20, no. 3, pp. 1–26, 2010.

[143] W. J. Sacco, D. M. Navin, K. E. Fiedler, R. K. Waddell, and W. B. Long, "Precise formulation and evidence-based application of resource-constrained triage," *Academic Emergency Medicine*, vol. 12, no. 8, pp. 759–770, 2005.

[144] M. Benson, K. L. Koenig, and C. H. Schultz, "Disaster triage: Start, then save-a new method of dynamic triage for victims of a catastrophic earthquake," *Prehospital and Disaster Medicine*, vol. 11, no. 2, pp. 117–124, 1996.

[145] N. A. Bostick, I. Subbarao, F. M. Burkle, E. B. Hsu, J. H. Armstrong, and J. J. James, "Disaster triage systems for large-scale catastrophic events," *Disaster Medicine and Public Health Preparedness*, vol. 2, no. Supplement_1, pp. S35–39, 2008.

[146] E. B. Lerner, R. B. Schwartz, P. L. Coule, E. S. Weinstein, D. C. Cone, R. C. Hunt, S. M. Sasser, J. M. Liu, N. G. Nudell, I. S. Wedmore, J. Hammond, E. M. Bulger, J. P. Salomone, T. L. Sanddal, G. C. Lord, D. Markenson, and R. E. OConnor, "Mass casualty triage: An evaluation of the data and development of a proposed national guideline," *DISASTER MEDICINE AND PUBLIC HEALTH PREPAREDNESS*, vol. 2, no. Supplement_1, pp. S25–34, 2008.

[147] New York Centers for Terrorism Preparedness and Planning (NYCTP), "Mass casualty/trauma event protocol," 2006.

[148] C. H. Schultz, K. L. Koenig, and E. K. Noji, "A medical disaster response to reduce immediate mortality after an earthquake," *The New England Journal of Medicine*, vol. 334, no. 7, pp. 438–444, 1996.

[149] National Fire Protection Association (NFPA), "Nfpa 1561: Standard on emergency services incident management system," 2008.

[150] M. North, E. Tatara, N. Collier, and J. Ozik, "Visual agent-based model development with repast simphony," in *Proceedings of the Agent 2007 Conference on Complex Interaction and Social Emergence, Argonne National Laboratory, Argonne, IL USA*, 2007.

[151] NetLogo, "Netlogo home page." http://ccl.northwestern.edu/netlogo/. [Accessed: Apr 10, 2009].

[152] S. Luke, C. Cioffi-Revilla, L. Panait, and K. Sullivan, "Mason: A new multi-agent simulation toolkit," in *Proceedings of the 2004 SwarmFest Workshop.*, 2004.

[153] N. Minar, R. Burkhart, C. Langton, and M. Askenazi, "The swarm simulation system: A toolkit for building multi-agent simulations," tech. rep., Santa Fe Institute, 1996.

[154] "Self-study guide for repastj - java-based repast (tesfatsion)." http://www.econ.iastate.edu/tesfatsi/repastsg.htm#CASIntro. [Accessed: Apr 10, 2009].

[155] A. T. Crooks, "The repast simulation/modelling system for geospatial simulation," Tech. Rep. Paper 123 - Sept 07, UCL CENTRE FOR ADVANCED SPATIAL ANALYSIS, 2007.

[156] N. Collier, "Repast: An extensible framework for agent simulation,"

[157] D. Machin, Y. B. Cheung, and M. Parmar, *Survival Analysis: A Practical Approach*. Wiley, 2 ed., 2006.

[158] D. M. Navin, W. J. Sacco, and G. McGill, "Application of a new resource-constrained triage method to military-age victims," *Military Medicine*, vol. 174, no. 12, pp. 1247–1255, 2009.

[159] S. I. Gass, "Decision-aiding models: Validation, assessment, and related issues for policy analysis," *Operations Research*, vol. 31, no. 4, pp. 603–631, 1983.

[160] R. E. Bechhofer, T. J. Santner, and D. M. Goldsman, *Design and analysis of experiments for statistical selection, screening, and multiple comparisons*. Wiley, 1995.

[161] H. Finner and G. Giani, "Closed subset selection procedures for selecting good populations," *Journal of Statistical Planning and Inference*, vol. 38, no. 2, pp. 179–199, 1994.

[162] H. Finner and K. Strassburger, "On δ-equivalence with the best in k-sample models," *Journal of the American Statistical Association*, vol. 101, no. 474, pp. 737–746, 2006.

[163] S. S. Gupta and D.-Y. Huang, "Subset selection procedures for the means and variances of normal populations: Unequal sample sizes case," *Sankhyā: The Indian Journal of Statistics, Series B*, vol. 38, no. 2, pp. 112–128, 1976.

[164] K. Lam, "A new procedure for selecting good populations," *Biometrika*, vol. 73, no. 1, pp. 201–206, 1986.

[165] E. J. Chen, "Restricted subset selection," in *Proceedings of the 40th Conference on Winter Simulation*, pp. 281–289, IEEE, 2008.

[166] V. Fabian, "Note on anderson's sequential procedures with triangular boundary," *The Annals of Statistics*, vol. 2, no. 1, pp. 170–176, 1974.

[167] C. Jennison, I. Johnstone, and B. Turnbull, "Asymptoti-cally optimal procedures for sequential adaptive selection of the best of several normal means," Tech. Rep. No. 463, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, 1980.

[168] J. C. Hsu, *Multiple comparisons: theory and methods*. Chapman and Hall/CRC, 1 ed., 1996.

[169] C. C. Chiuzan, *Multiple Comparisons with a Control in R*. PhD thesis, University of South Carolina, 2009.

[170] "Wicked problem - wikipedia, the free encyclopedia." http://en.wikipedia.org/wiki/Wicked_problem. Accessed Nov 27, 2011.

[171] "Satisficing - wikipedia, the free encyclopedia." http://en.wikipedia.org/wiki/Satisficing. Accessed Nov 27, 2011.