PRIORITIZATION AND DISTRIBUTION OF CASUALTIES IN DISASTER RESPONSE MANAGEMENT

Zheqi Zhang

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research.

Chapel Hill
2019

Approved by:

Nilay Tanık Argon

Serhan Ziya

Vidyadhar G. Kulkarni

Shu Lu

Quoc Tran-Dinh

ABSTRACT

Zheqi Zhang: Prioritization and Distribution of Casualties in Disaster Response Management
(Under the direction of Nilay Tanık Argon and Serhan Ziya)

This dissertation focused on two different problems that typically arise in the aftermath of disasters. In the first part of the dissertation, we study the problem of how casualties should be prioritized and distributed to different medical facilities in the aftermath of mass casualty incidents (MCIs) with the objective of maximizing the expected total number of survivors. Assuming that casualties have been triaged into two classes differentiated by their severity levels and medical needs, the decision-maker needs to prioritize and distribute casualties using a limited number of ambulances to multiple medical facilities with different capacities. By explicitly taking into consideration the capacity and service time at each medical facility, we formulate this sequential decision-making problem as a Markov decision process (MDP). Based on this MDP formulation, we propose heuristic policies that prescribe decisions on prioritization and distribution of casualties. We then employ discrete-event simulations to demonstrate the benefits of using the proposed heuristics against some benchmark policies under several realistic mass casualty incident scenarios such as terrorist attacks, major traffic accidents, and earthquakes.

In the second part of the dissertation, we study the resource allocation problem in urban search and rescue operations that follow natural disasters. Specifically, we consider a scenario in which some individuals are trapped at various locations within a geographical area and there is a limited time window during which these individuals can be rescued. We model the problem as an MDP. Then, we characterize the optimal policy under the assumption that individuals belong to only one of two locations. We propose heuristics for the general version of the problem. Finally, the proposed heuristics are examined with a simulation.

ACKNOWLEDGEMENTS

There are many people who I feel grateful for their contribution to my time in graduate school. Specifically, I would like to express my sincerest appreciation to my advisors, other committee members, my family, my school mates. Without them, this thesis would not have been possible. First, I am indebted to my dissertation advisors, Professor Nilay Tanık Argon and Professor Serhan Ziya. They have offered me guidance and support in both academic and personal growth. I am deeply proud to have been their student. Besides my advisor, I would like to thank the rest of my dissertation committee members: Professor Vidyadhar G. Kulkarni, Prof. Shu Lu, and Prof. Quoc Tran-Dinh for their time and advice. I am especially thankful to Prof. Quoc Tran-Dinh for his direction in the early stage of my graduate studies. I am also grateful for the financial support for my research projects from the National Science Foundation (grant number CMMI1635574). Last but not least, I would like to express my deepest gratitude to my family and friends. Without their warm love, continued patience, and endless support, I would never be able to make it through graduate school.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

The well-known scarcity principle does not only apply in economics. It can also be observed in the aftermath of disasters where the emergency rescue and medical resources and personnel are limited compared with the massive rescue and medical requests. Unlike the situation in economics, the allocation of scarce resources in disaster response management is fatal, thus should not be determined by the invisible hand of market. In this dissertation, we consider certain decisions that are given by emergency responders and coordinators on the allocation of various scarce service resources in catastrophic incidents. Specifically, we study two emergency response situations and related decisions: (i) prioritization and distribution of casualties to medical facilities in the aftermath of mass casualty incidents (MCIs) such as terrorist shootings and earthquakes, and (ii) prioritization of people who need to be rescued from their homes in the aftermath of a flood.

MCIs are events in which emergency medical service resources, such as personnel and equipment, are overwhelmed by the quantity and severity of casualties [36]. In the aftermath of MCIs, a surge of casualties demanding immediate medical treatment overwhelms the emergency medical system. It is impossible to provide timely treatment to all casualties involved. Hence, the objective of emergency responses to such events is generally stated as "doing the best for the most people" [54]. In this dissertation, we focus on the casualty prioritization and distribution problem in response to MCIs. Finding the optimal solution to the problem is not a trivial task. It requires a comprehensive consideration of casualties' acuity and their medical needs, the distance to and congestion at medical facilities, and the capacity and capability of medical facilities. Our goal is to develop dynamic decision rules that prescribe *priorities* among casualty classes for transportation from the scene and the *routing* of ambulances carrying casualties to medical facilities.

We used a Markov Decision Process (MDP) formulation to model the casualty prioritization and distribution problem with two distinct casualty classes. When a mass casualty event occurs, a surge of casualties appears at possibly multiple incident locations and overwhelms available transportation and medical resources. Typically, a limited number of ambulances are available for transporting casualties from incident locations to multiple medical facilities. The service capacity of each medical facility is constrained by its number of equipment and medical personnel. Solving the MDP problem for instances of realistic size is not possible. Therefore, we propose easy-to-implement index type heuristic policies that can be applied to all kinds of mass casualty events. More details on the casualty prioritization and distribution problem in response to an MCI are presented in Chapter 2.

In Chapter 3, we examine the urban search and rescue operations that follow floodings such as the one that happened in Texas in 2017 following Hurricane Harvey. The urban search and rescue operations are critical in saving people's lives. However, managing operations is difficult due to hazardous weather conditions, large numbers of rescue requests, and limited resources. To the best of our knowledge, no standard guideline exists for coordinating the urban search and rescue operations at least in the U.S. and rescue operation decisions are made in an "ad hoc" fashion based on previous experience.

We developed an MDP model for the urban search and rescue operations assuming there are two classes of requests with different service rates, rewards, holding costs, and class-wise deadlines. We fully characterized the optimal policy. Then, we designed heuristics to find near-optimal solutions for the general model that assumes more than two classes of requests and multiple servers. More details on this problem of allocating urban search and rescue resources in flooding disasters are presented in Chapter 3.

CHAPTER 2

DISTRIBUTION AND PRIORITIZATION OF PATIENTS IN THE AFTERMATH OF MASS CASUALTY INCIDENTS

2.1 Introduction

Mass casualty incidents (MCIs) are events in which emergency medical service resources, such as personnel and equipment, are overwhelmed by the quantity and severity of casualties [36]. Mass casualty incidents can be categorized into two: natural disasters (such as flooding, earthquakes, and hurricanes) and man-made incidents (such as traffic accidents, nuclear plant meltdowns, and acts of terrorism). In recent years, mass casualty incidents have occurred more frequently and affected more people. Particularly, terrorist attacks, such as mass shootings, vehicle ramming attacks, and release of chemical or biological weapons are overgrowing threats according to the Global Terrorism Database [40]. Figure 2.1 illustrates the trend of total number of fatalities from the terrorist attacks in recent decades.

It is common belief that more lives could be saved after such mass casualty events if the affected communities have a scalable and well prepared emergency response system established. According to the World Health Organization (WHO) [54], even though the insufficiency in preparing for such emergency events is well recognized around the world, the problem has not been addressed yet in a comprehensive and systematic way.

In the aftermath of mass casualty incidents, a surge of casualties demands urgent medical services which overwhelms the emergency medical system and makes it impossible to provide timely treatment to all casualties involved. Therefore, the objective of emergency response to such events is generally stated as “doing the best for the most people” [54]. Response to such events typically requires close collaboration between multiple organizations and agencies.

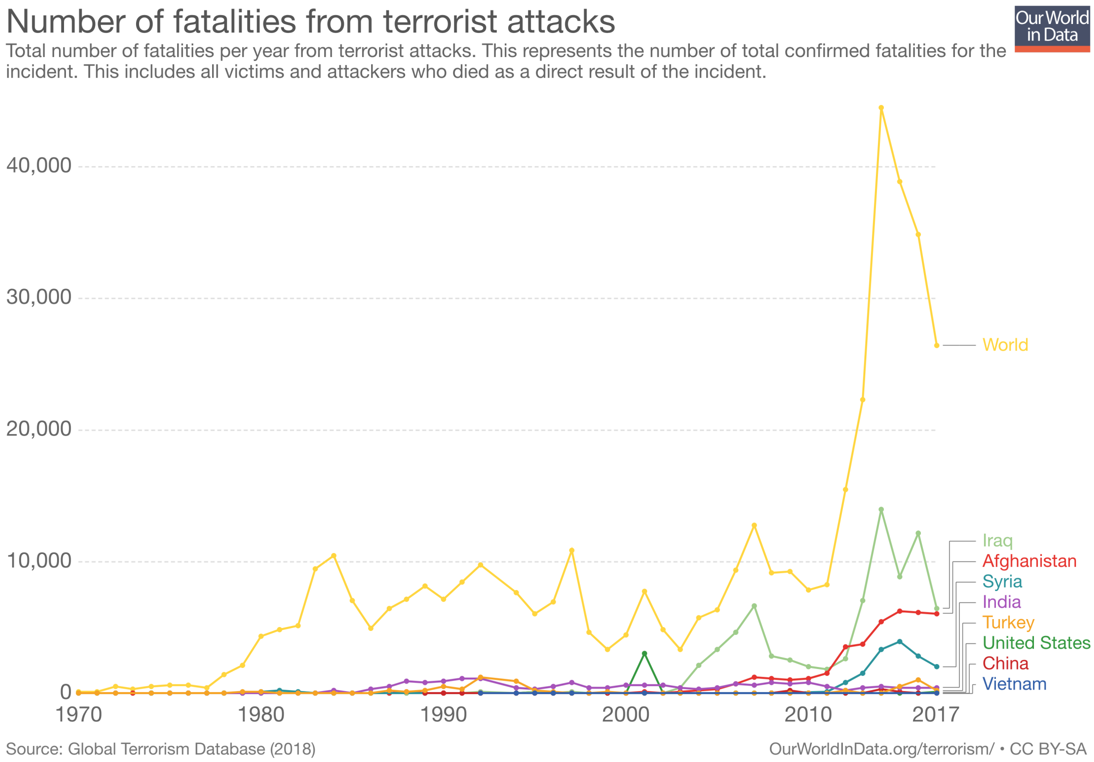

Figure 2.1: Total number of fatalities per year from terrorist attacks [40]

Furthermore, emergency medical resources have to be assigned and adjusted dynamically in real time as the status of the incident changes. In this article, we focus on the casualty distribution problem in response to a mass casualty incident. The problem consists of two critical decisions – the prioritization and transportation of casualties to medical facilities. Due to casualties' urgent needs for medical services and the chaotic scene of a mass casualty incident, those decisions have to be made quickly with incomplete and imprecise information.

The typical flow of events during response to mass casualty incidents consists of three phases: triage, transportation, and treatment. The triage and transportation together are often named the pre-hospital phase and the treatment is also called the hospital phase. Triage refers to the classification of casualties based on the severity of their injuries. Simple Triage and Rapid Transport (START) [70] and its variations are the most widely used mass casualty triage methods in the U.S. according to [53]. First responders arriving at the scene of a mass casualty incident first evaluate casualties using START based on the ability of casualties' walking, respiration, perfusion, and mental status at safe locations. After triage, the decision makers need to determine the distribution of casualties to medical facilities. For "static" triage methods, such as START, priorities for transportation and treatment are determined by casualties' triage classes. Such static prioritization rules are simple enough to be implemented quickly by emergency personnel but their main drawback is that they only consider the severity of casualties while overlooking the number of casualties and the availability of resources. As pointed out in [32, 49, 63] when medical resources are scarce compared with the number of casualties, more lives can be saved by incorporating state-dependent triage rules. The casualty distribution decision is perhaps even more difficult. Sending an excessive number of casualties to the closest medical facility leads to excessive congestion which causes a delay in medical care and imperils the survival of critical casualties [58]. On the other hand, if a casualty were transported to a remote medical facility, the treatment will also be deferred due to longer traveling time, which also undermines other casualties since the number of ambulances is limited compared with the number of casualties in a mass casualty event. A dynamic policy taking into consideration both the number of casualties and the availability of resource *factors* could potentially outperform simple priority and distribution policies [49].

Our goal is to develop dynamic decision rules that prescribe *priorities* among casualty classes for transportation from the scene and the *routing* of ambulances carrying casualties to medical facilities. We start with a stylized mathematical model of the casualty prioritization and distribution problem involving two casualty classes. When a mass casualty event takes place, a surge of casualties appears at possibly multiple incident locations which overwhelm the available transportation and medical resources. A limited number of ambulances are available for transporting casualties from incident locations to multiple medical facilities. Furthermore, at each medical facility, service capacity is constrained by the number of equipment and medical personnel. Therefore, casualties have to first wait for ambulances at event locations and again wait for medical attention at medical facilities. Our stochastic model considers this underlying tandem structure of congestion and captures the randomness in transportation and treatment times at medical facilities. Unfortunately, this stochastic model even with its many assumptions is too complex to study analytically. However, it is still useful as it paves the way for several easy-to-implement heuristics that provide joint prioritization and distribution decisions. The proposed heuristics are of index type and most of the indices can be calculated from one line of mathematical formulas using a calculator. What is more, our heuristics are widely applicable to various types of mass casualty events involving multiple casualty classes, multiple incident locations, and multiple medical facilities. More importantly, we examined the proposed heuristics in a comprehensive simulation study that would give insights into various types of MCIs. In particular we created four types of hypothetical scenarios based on four different kinds of real MCIs: single-location terrorist attacks, multiple-location terrorist attacks, single-location major traffic accidents, and earthquakes. Each type of scenario has unique characteristics such as injury types, scales of events, and geography that all affect the performance of heuristics. Indeed, we find in our simulation study, there is not a heuristic outperforms all other heuristics in all scenarios. Certain events may require more sophisticated heuristics whereas in others, simpler heuristics may provide satisfactory performance.

The remaining of this chapter is organized as follows. We review both medical and operations research literature relevant to this work in Section 2.2. The casualty prioritization and distribution problem is formulated as a MDP in Section 2.3. Based on the insights derived from our MDP model, several heuristic policies are constructed in Section 2.4 when the status of all medical facilities are perfectly known. We extend our heuristics to the situation where the information on the status of medical facilities is incomplete in Section 2.5. We provide the results of our simulations study to test the proposed heuristics in hypothetical but realistic scenarios in Section 2.6. We conclude the chapter by a discussion of the main insights gained from this study in Section 2.7. Additional supporting material is provided in the Appendix.

2.2 Literature Review

We start by reviewing existing mass-casualty triage methods in the medical literature. A committee study identified and compared 9 existing triage systems including 2 pediatric systems in [34, 41]. Most of the MCI triage systems are similar in terms of their use of physiological criteria to classify casualties into 4 or 5 classes. Most broadly accepted triage methods are not evidence-based and no quantitative research had been conducted to evaluate the usability, reliability, and the ability to apply those triage system in MCIs according to [34]. In the review on triage methods sponsored by the Center for Disease Control and Prevention (CDC), the review committee that consists of medical experts concluded there is no evidence to support that one triage system outperforms the others [41]. The reasons are that data collected during the mass casualty incidents are anecdotal [41] and data from daily emergency department activities may not reflect the situation during mass casualty incidents [24]. A national guideline was also proposed in [41] to unify the mass casualty triage process across the U.S. More recently, Bazyar et al. [9] conducted a review of twenty different adult triage systems used in practice that were proposed between 1990 and 2018 across the world. The authors have not declared any triage system as the best and recommended that different countries should have their own triage systems.

Simple Triage and Rapid Transport (START) [70] and its variants are the most widely used mass-casualty triage methods in the U.S. It has been implemented during the 1994 Northridge earthquake in California, U.S.A. and the 9/11 attacks to the New York World Trade Center in 2001. START uses the ability to obey commands, respiratory rate, and radial pulse to classify casualties into four triage classes: The casualties in expectant class (black tag) are unlikely to survive given the severity of injuries and available level of medical care. Immediate casualties (red tag) require urgent medical intervention to survive. Delayed casualties (yellow tag) may have life-threatening injuries but their conditions are not expected to deteriorate rapidly. Minor casualties (green tag) have non life-threatening injuries and they may be able to access medical care on their own, thus often referred to as "walking wounded." Once the casualties are triaged, the priority for transportation and treatment are given in the following order: immediate class first, followed by the delayed class, and finally the minor and the expectant casualties. Within each category, casualties in the worst condition have the priority. SAVE [11] was proposed to compliment START as a secondary triage method. It considers the limitation of on-site medical resources to prioritize casualties within each START class further.

SALT (Sort, Access, Life-saving interventions, Treatment and/or transportation) triage is proposed in [41] as a national triage guideline "using aspects of all identified systems that were supported by the best available evidence and expert opinion." SALT starts with a global sorting of casualties into three groups for individual assessment. Based on motor function, casualties who are able to walk have the least priority, those who cannot move or with obvious life threats are given the highest priority, and remaining casualties who are able to respond with purposeful movement are assigned medium priority. The individual assessment begins with damage control stabilization. Then, casualties are further prioritized into five categories: immediate, delayed, minimal, expectant, and dead based on physiological criterion similar to START.

Sacco Triage Method (STM) developed by Sacco and his colleagues in [52, 63, 64] is a triage method that takes into consideration the severity of causalities together with the availability of resources simultaneously in real time. First responders evaluate each casualty and assign an RPM (Respiration, Pulse and Motor response) score describing the severity of each casualty initially. Deterioration for each RPM score is estimated by the Delphi method and logistic regression is used to compute the survival probability as a function of time for each RPM score assuming minimal medical intervention. A mixed integer program is solved in real time to determine the prioritization for transportation and treatment with the objective of maximizing the expected number of survivors.

Static methods such as START and SALT can be performed quickly by medical personnel without extensive training. The drawbacks for such triage systems are also clear: they only consider the severity of casualties while ignoring the scale of the event and the availability of resources at medical facilities. The disadvantages for state-dependent methods such as STM are pointed out in [18, 34]: detailed data on the triage results and the availability of resources are difficult to collect within a short period of time in the aftermath of mass casualty incidents. Also, software and hardware are necessary for solving a complex mathematical programming problem. As suggested by Brandeau et al. [13] quantitative methods are important tools for planning effective health sector responses to disasters while an appropriate balance between simplicity and complexity is critical. We believe an ideal triage method should have the property of simplicity to implement such as START and SALT but should also take into consideration the characteristic of casualties together with the limitation of resources as in STM.

In the operations research literature, the prioritization and distribution of casualties in the aftermath of mass casualty incidents have drawn much attention recently. Most of the prior work studies these problems separately: some consider the distribution problem only assuming all casualties are identical and others focus on finding the best prioritization while considering a single medical facility or ignoring the capacity at medical facilities. Only a few looked at the prioritization and distribution problems simultaneously like we do in this chapter.

A large number of studies focus solely on the casualty distribution problem (also referred to as the ambulance allocation problem). Gong and Batta [27] studied the ambulance allocation and re-allocation problem in the aftermath of mass casualty incidents. Their model assumed casualties in clusters are discovered and picked up by ambulances. Ambulances are re-allocated at deterministic time epochs with the objective of minimizing the makespan. Jotshi et al. [37] used data fusion and simulation methods to investigate the dispatch and routing of ambulances post-disaster. Their model fuse the information of casualties, roads, ambulances, and hospitals to estimate and maximize the life expectancy of casualties. The Severity Adjusted Victim Evacuation method (also known as 'SAVE' which is different from the 'SAVE' triage method in [11]) proposed by Dean and Nair [19] is a mixed-integer programming formulation for finding the optimal patient distribution after a mass casualty incident. The objective is to maximize the expected number of survivors with the constraints on available ambulances and the capacity at medical facilities. Mills et. al. [49] in their most recent chapter studied the casualty distribution problem in the aftermath of mass casualty incidents with a comprehensive model. The model considered general survival probabilities of casualties at distinct locations, limited transportation resources, dynamic capacity at medical facilities, as well as uncertainties in transportation and treatment times. Two heuristics were proposed using a myopic approach and one-step policy improvement approach. The simulation studies demonstrate the robustness and advantage of both heuristics against benchmark policies with limited number of ambulances.

There is also a vast literature dedicated to solving the casualty prioritization problem in the aftermath of MCIs. Argon et al. [4] and Jacobson et al. [32] modeled the casualty prioritization problem in the aftermath of mass casualty incidents as a scheduling problem for a clearing system. The objective is to maximize the expected number of survivors where each casualty is assumed to have a random deadline. Stochastic comparisons were used to identify conditions under which state-independent polices are optimal and optimal policies were partially characterized when they depend on the system state by means of an MDP formulation. [4, 32] demonstrated the benefits of state dependent policies for prioritization decisions by numerical studies. A fluid approximation to the stochastic problem was employed by Mills et al. [48, 50] to study prioritization of casualties. Casualties were modeled as a fluid that flow continuously from a single casualty location to a single medical facility. Criticalities of casualties are captured by non-increasing survival functions rather than abandonments as in earlier studies [4, 32]. Mills et al. [48] proved that the priority of transportation will switch at most twice under the optimal policy when two distinct quasi-concave functions represent survival probabilities for the immediate and delayed classes in START. An algorithm (ReSTART) was proposed to identify these switching times. Mills [47] further extended the results to multiple classes of casualties and multiple medical facilities. A heuristic policy that integrates the myopic and look-ahead policies was proposed to determine the priorities. Numerical studies showed that the performance of proposed heuristics are close to the optimal solution obtained from solving a mixed integer program as in SAVE [19] and STM models [63, 64].

Mizumoto et al. [51] showed that the combined casualty distribution and prioritization problem (they call it transportation scheduling problem) is NP-hard. They applied top k breath first search (DkBFS) to achieve a near-optimal solution. The authors introduced E-triage or E-tag, which can sense the vital signs of patients to estimate their survival probabilities in real time. Casualties with different time-varying probabilities of survival and multiple medical facilities with fixed capacity are considered. The objectives are to maximize the number of casualties whose probability of survival is greater than a threshold as well as to maximize the average survival probabilities. Jin et al. [35] proposed an emergency logistic model for casualty delivery and medical resource allocation in a mass casualty event. Triaged casualties from multiple locations have different probabilities of survival in each class. Each casualty has the choice of receiving care at the on-site clinic to prolong patients' survival before visiting a general hospital for treatment or visit the general hospital directly. Resources are limited at both on-site clinic and general hospital. Mixed integer programming is used to maximize the number of casualties with the probability of survival above a threshold. Sung and Lee [69] also formulated the ambulance routing problem using as a mixed-integer program. Branch-and-price scheme was applied to find close-to-optimal solutions. The main drawback with either solving a mathematical program as in [35, 69] or searching on a graph model as in [51] is the lack of practicality. These approaches are suitable for making plans prior to the events. Waiting for a computer program to find a solution and then deploying it in the chaotic environment of mass casualty incidents are still deemed unrealistic. In addition, these deterministic methods overlook the stochastic nature of the mass casualty incidents. In a recent paper by Shin and Lee [65], simulation-based Approximate Dynamic Programming (ADP) is used to obtain a near-optimal solution of an MDP formulation for the joint prioritization and distribution problem. They also developed a heuristic policy for a single location multiple hospital scenario. The proposed heuristic is tested in a case study with a limited number of casualties.

Patient distribution problem is also well studied in the context of daily emergency medical services (EMS) and military medical operations. ADP is adopted by Maxwell et al. [45] and Rettke et al. [59] to obtain near optimal solution for the respective MDP models. Rettke et al. [59] focus on the dispatching of military medical evacuation assets, whereas Maxwell et al. [45] are motivated by the ambulance reallocation problem for daily emergencies. Recent reviews on EMS systems could be found in [5, 10, 60].

2.3 Problem Formulation

2.3 Model Assumptions

We model the joint casualty prioritization and distribution problem using a Markov decision process (MDP). We prune away some elements in order to focus on the essence of the problem and have a tractable model. The simplicity of the MDP model allows us to gain insights into the problem and develop heuristic policies. Neglected features are placed back into the problem in the simulation stage to test the performance of the proposed heuristic policies in more realistic settings in Section 2.6.

In our MDP model, mass casualty events take place at $\mathcal{L}$ distinct locations. Let $\mathbb{L}$ be the set of all such locations. At each casualty location, we assume casualties have already been triaged according to START classification [70]. We only consider casualties in the immediate and delayed classes in our model since their survivals are most sensitive to timely medical intervention. While minor casualties are less urgent for medical services, expectant casualties are unlikely to survive given the level of available care and severity. Furthermore, Frykberg [23] pointed out that the survival of critical casualties (immediate and delayed) is the best measure of the success for emergency medical response in mass casualty incidents. Let $\mathbb{C} = \{i, d\}$ denote the set of casualty classes where i denotes the immediate class and d denotes the delayed class.

Multiple medical facilities, such as trauma centers and hospitals, are available to serve casualties. Let $\mathbb{H}$ denote the set of $\mathcal{H}$ medical facilities. Each medical facility may have a different service capability and capacity. Assume medical facility $h \in \mathbb{H}$ has $b_h^{(c)} < \infty$ servers dedicated to class $c \in \mathbb{C}$ casualties. The contingency plan for mass casualty incidents varies from hospital to hospital – arriving casualties could be served at ED, ICU or other departments of the medical facility [21]. In this context, the total number of servers will refer to the maximum number of casualties a medical facility will be able to handle at a time. $b_h^{(c)} = 0$ indicates that medical facility h is not capable of providing care to class c casualties. If a casualty were directed to a medical facility not capable of serving him/her, he/she would have to be transferred to another medical facility. Such a route would be less preferable than transporting the casualty directly to a medical facility with the ability to treat him/her. Therefore, we assume a casualty will only be transported to medical facilities capable of providing care to that casualty in our model.

Upon arriving at medical facility h, a casualty of class c receives treatment immediately from one of $b_h^{(c)}$ servers if there is an idle server. If all servers are busy serving other class c casualties, he/she will join a first-come-first-serve queue. We assume independent and identical exponential service times with rate $\mu_h^{(c)} < \infty$ for each server serving a class c casualty at medical facility h. At service completion, the survival probability of a class c casualty is $r_h^{(c)}$ at medical facility h. The survival probability depends on both the class of casualty and medical facility, which may reflect the severity of casualty and the capability of the medical facility, respectively. Since we assume dedicated servers for immediate and delayed casualties with independent service performance at all medical facilities, we hypothetically separate each medical facility into two independent facilities with the same location in our model, but one treating immediate casualties and the other treating delayed casualties.

Limited transportation resources, such as ambulances, are available for transporting casualties between casualty locations and medical facilities. Let $\mathcal{A} < \infty$ be the total number of ambulances. We assume an ambulance carries at most one casualty at a time. The travel time from casualty location $l \in \mathbb{L}$ to medical facility $h \in \mathbb{H}$ follows an exponential distribution with rate $\lambda_{lh} < \infty$ for any ambulance. For modeling tractability, we ignore the travel time for empty ambulances returning to casualty locations from a medical facility as in Mills et al. [49].

In order to capture different deterioration rates for different classes of casualties, we assume exponential discounting with a discount rate of $\alpha_i > 0$ for the immediate class and $\alpha_d > 0$ for the delayed class. The performance measure is the expected total discounted survival probability, which can be interpreted as the expected number of survivors. We aim to find an optimal policy that determines the priority for transportation as well as the destination facility for each casualty.

In our sequential decision making problem, decisions are made every time the state of the system changes. When an ambulance arrives at a casualty location, we need to make the prioritization and distribution decisions simultaneously. More specifically, we need to select a casualty from either the immediate class or the delayed class and determine a medical facility as the destination for transportation. There is no decision to make when a casualty completes his/her service at a medical facility but a reward representing the expected survival probability for the casualty is collected.

Our model generalizes the model in Mills et al. [49] to two classes of casualties. The presence of a second casualty class induces the casualty prioritization problem, which is neglected in [49] since all casualties are assumed to be identical. The single class problem studied in [49] is already too complicated to obtain the optimal solution. The extra dimension of the problem makes it even more complex to solve analytically. Instead, we focus on developing heuristic policies based on insights gained from the MDP formulation. We then test all heuristic policies in more realistic settings which mimic various types of actual mass casualty incidents including terrorist attacks, major traffic accidents, and natural disasters. Proposed heuristic policies are compared with benchmark policies in multiple simulation studies to identify their advantages and disadvantages.

2.3 MDP Formulation

We first define the state of the system, let $\mathbf{S}^{(c)}(t) = \Big(\mathbf{W}^{(c)}(t), \mathbf{X}^{(c)}(t)\Big)$ be the state of class $c \in \mathbb{C}$ casualties in the system at time $t > 0$, where $\mathbf{W}^{(c)}(t) = \Big(w_1^{(c)}(t), w_2^{(c)}(t), \dots, w_{\mathcal{L}}^{(c)}(t)\Big)$, $w_l^{(c)}(t)$ is the number of class c casualties at location $l \in \mathbb{L}$ at time t, and $\mathbf{X}^{(c)}(t) = \Big(x_1^{(c)}(t), x_2^{(c)}(t), \dots, x_{\mathcal{H}}^{(c)}(t)\Big)$, $x_h^{(c)}(t)$ is the number of class c casualties at medical facility $h \in \mathbb{H}$ at time t. Then, the system state at time $t \geq 0$ can be express as $\mathbf{S}(t) = \Big(\mathbf{S}^{(i)}(t), \mathbf{S}^{(d)}(t)\Big) \in \mathcal{S}$, where $\mathcal{S}$ denotes the state space.

Let $\mathbf{a} = \{a_{lh}^{(c)} : c \in \mathbb{C}, l \in \mathbb{L}, h \in \mathbb{H}\}$ denote the decision matrix at a decision epoch, where $a_{lh}^{(c)}$ is the number of ambulances carrying class c casualties from location l to facility h. At any given state $\mathbf{S}$ (we will drop the time index t for notational simplicity whenever not needed), a admissible action satisfies the following constraints:

$$\sum_{h \in \mathbb{H}} a_{lh}^{(c)} \leq w_l^{(c)}, \forall l \in \mathbb{L}, c \in \mathbb{C}, \tag{2.1}$$

$$\sum_{l \in \mathbb{L}} \sum_{h \in \mathbb{H}} \sum_{c \in \mathbb{C}} a_{lh}^{(c)} \leq \mathcal{A}. \tag{2.2}$$

Constraint (2.1) is the casualty availability constraint at each location for each casualty class and constraint (2.2) is the ambulance availability constraint. Let $\mathbb{A}_{\mathbf{S}}$ denote the set of all admissible actions at state $\mathbf{S}$ that satisfy (2.1) and (2.2).

Let $f^{(c)}(t|\mathbf{S}, \mathbf{a}, \mathbf{S}')$ denote the discounted reward generated when action $\mathbf{a}$ taken at state $\mathbf{S}$ takes the system to state $\mathbf{S}'$ after t units of time. It can be expressed as

$$f^{(c)}(t|\mathbf{S}, \mathbf{a}, \mathbf{S}') = \begin{cases} r_h^{(i)} e^{-\alpha_i t} & \text{if } \mathbf{S}' = \left(\left(\mathbf{W}^{(i)}, \mathbf{X}^{(i)} - \mathbf{e}_h\right), \mathbf{S}^{(d)}\right) \text{ and } c = i, \\ r_h^{(d)} e^{-\alpha_d t} & \text{if } \mathbf{S}' = \left(\mathbf{S}^{(i)}, \left(\mathbf{W}^{(d)}, \mathbf{X}^{(d)} - \mathbf{e}_h\right)\right) \text{ and } c = d, \\ 0 & \text{otherwise,} \end{cases} \tag{2.3}$$

where $\mathbf{e}_k$ is a vector of zeros with only its kth component being one and its length is clear from the context. Define $P(\mathbf{S}'|\mathbf{S}, \mathbf{a})$ to be the probability that the system will end up in state $\mathbf{S}'$ at the beginning of the next decision epoch if decision $\mathbf{a} \in \mathbb{A}_{\mathbf{S}}$ is implemented at state $\mathbf{S}$. Define also $\beta(\mathbf{S}, \mathbf{a}) = \sum_{l \in \mathbb{L}} \sum_{h \in \mathbb{H}} \sum_{c \in \mathbb{C}} a_{lh}^{(c)} \lambda_{lh} + \sum_{h \in \mathbb{H}} \sum_{c \in \mathbb{C}} ((b_h^{(c)} \wedge x_h^{(c)}) \mu_h^{(c)})$ where $a \wedge b = \min(a, b)$. Then, under the assumption of exponential service and travel times, we have

$$P(\mathbf{S}'|\mathbf{S}, \mathbf{a}) = \begin{cases} a_{lh}^{(i)} \lambda_{lh} / \beta(\mathbf{S}, \mathbf{a}) & \text{if } \mathbf{S}' = \left(\left(\mathbf{W}^{(i)} - \mathbf{e}_l, \mathbf{X}^{(i)} + \mathbf{e}_h\right), \mathbf{S}^{(d)}\right), \\ & \forall l \in \mathbb{L}, h \in \mathbb{H}, \\ a_{lh}^{(d)} \lambda_{lh} / \beta(\mathbf{S}, \mathbf{a}) & \text{if } \mathbf{S}' = \left(\mathbf{S}^{(i)}, \left(\mathbf{W}^{(d)} - \mathbf{e}_l, \mathbf{X}^{(d)} + \mathbf{e}_h\right)\right), \\ & \forall l \in \mathbb{L}, h \in \mathbb{H}, \\ (b_h^{(i)} \wedge x_h^{(i)}) \mu_h^{(i)} / \beta(\mathbf{S}, \mathbf{a}) & \text{if } \mathbf{S}' = \left(\left(\mathbf{W}^{(i)}, \mathbf{X}^{(i)} - \mathbf{e}_h\right), \mathbf{S}^{(d)}\right), \forall h \in \mathbb{H}, \\ (b_h^{(d)} \wedge x_h^{(d)}) \mu_h^{(d)} / \beta(\mathbf{S}, \mathbf{a}) & \text{if } \mathbf{S}' = \left(\mathbf{S}^{(i)}, \left(\mathbf{W}^{(d)}, \mathbf{X}^{(d)} - \mathbf{e}_h\right)\right), \forall h \in \mathbb{H}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

The first two cases in (2.4) correspond to ambulance arrivals and the next two cases correspond to service completions at medical facilities. We next apply uniformation as in [43] with a uniformization constant

$$\beta = \mathcal{A} \sum_{l \in \mathbb{L}} \sum_{h \in \mathbb{H}} \lambda_{lh} + \sum_{h \in \mathbb{H}} \sum_{c \in \mathbb{C}} b_h^{(c)} \mu_h^{(c)}. \tag{2.5}$$

The new transition probabilities under uniformization are given by,

$$\tilde{P}(\mathbf{S}'|\mathbf{S},\mathbf{a}) = \begin{cases} (1-\beta(\mathbf{S},\mathbf{a}))/\beta & \text{if } \mathbf{S}'=\mathbf{S}, \\ P(\mathbf{S}'|\mathbf{S},\mathbf{a})\beta(\mathbf{S},\mathbf{a})/\beta & \text{if } \mathbf{S}'\neq\mathbf{S}. \end{cases} \tag{2.6}$$

By Theorem 11.5.2 in Puterman [57], there exists an optimal deterministic stationary policy to this MDP because $\mathcal{S}$ is discrete, action space is finite, rewards are bounded, and β is finite. Breaking ties in favour of prioritizing the immediate class and transporting to the closest medical facility, let π^* be the unique optimal deterministic stationary policy that takes action $\mathbf{a}^*_{\mathbf{S}}$ at state $\mathbf{S}$. Let $V^{(c)}(\mathbf{S})$ denote the total expected number of class c survivors starting from state $\mathbf{S}\in\mathcal{S}$ under π^*. And $V(\mathbf{S}) = V^{(i)}(\mathbf{S}) + V^{(d)}(\mathbf{S})$ denote the maximum expected number of survivors starting from state $\mathbf{S}$. Then, the optimality equations can be written as follows for all $\mathbf{S}\in\mathcal{S}$ and $\mathbf{a}\in\mathbb{A}_{\mathbf{S}}$,

$$V(\mathbf{S}) = \max_{\mathbf{a}\in\mathbb{A}_{\mathbf{S}}} \left\{ V^{(i)}(\mathbf{S},\mathbf{a}) + V^{(d)}(\mathbf{S},\mathbf{a}) \right\}, \tag{2.7}$$

$$\text{where } V^{(c)}(\mathbf{S},\mathbf{a}) = \sum_{\mathbf{S}'\in\mathcal{S}} \tilde{P}(\mathbf{S}'|\mathbf{S},\mathbf{a}) \int_0^\infty \left[f_h^{(c)}(t|\mathbf{S},\mathbf{a},\mathbf{S}') + e^{-\alpha_c t} V^{(c)}(\mathbf{S}') \right] \beta e^{-\beta t} dt, \\ \text{and } V^{(c)}(\mathbf{S}') = V^{(c)}(\mathbf{S}',\mathbf{a}^*_{\mathbf{S}'}) \text{ for } c\in\mathbb{C}. \tag{2.8}$$

The integrals in (2.7) can be rewritten as,

$$\int_0^\infty f^{(c)}(t|\mathbf{S},\mathbf{a},\mathbf{S}')\beta e^{-\beta t} dt = \begin{cases} r_h^{(i)}\beta/(\alpha_i+\beta) & \text{if } \mathbf{S}' = \left(\left(\mathbf{W}^{(i)}, \mathbf{X}^{(i)} - \mathbf{e}_h\right), \mathbf{S}^{(d)}\right) \\ & \text{and } c = i, \\ r_h^{(d)}\beta/(\alpha_d+\beta) & \text{if } \mathbf{S}' = \left(\mathbf{S}^{(i)}, \left(\mathbf{W}^{(d)}, \mathbf{X}^{(d)} - \mathbf{e}_h\right)\right) \\ & \text{and } c = d, \\ 0 & \text{otherwise,} \end{cases} \tag{2.9}$$

$$\int_0^\infty e^{-\alpha_c t} V^{(c)}(\mathbf{S}')\beta e^{-\beta t} dt = V^{(c)}(\mathbf{S}')\beta/(\alpha_c+\beta), \text{ for } c\in\mathbb{C}. \tag{2.10}$$

Plugging (2.4), (2.6), (2.9), and (2.10) into (2.7), we obtain

$$
\begin{aligned}
V^{(i)}(\mathbf{S},\mathbf{a}) = \frac{\beta V^{(i)}(\mathbf{S})}{\alpha_i+\beta} &+ \sum_{h\in\mathbb{H}} \left(b_h^{(i)} \wedge x_h^{(i)}\right) \mu_h^{(i)} \left(\frac{r_h^{(i)} + V^{(i)}(\mathbf{S}_{0,-h}^{(i)}, \mathbf{S}^{(d)}) - V^{(i)}(\mathbf{S})}{\alpha_i+\beta}\right) \\
&+ \sum_{h\in\mathbb{H}} \left(b_h^{(d)} \wedge x_h^{(d)}\right) \mu_h^{(d)} \left(\frac{V^{(i)}(\mathbf{S}^{(i)}, \mathbf{S}_{0,-h}^{(d)}) - V^{(i)}(\mathbf{S})}{\alpha_i+\beta}\right) \\
&+ \sum_{l\in\mathbb{L}} \sum_{h\in\mathbb{H}} a_{lh}^{(i)} \lambda_{lh} \left(\frac{V^{(i)}(\mathbf{S}_{-l,h}^{(i)}, \mathbf{S}^{(d)}) - V^{(i)}(\mathbf{S})}{\alpha_i+\beta}\right) \\
&+ \sum_{l\in\mathbb{L}} \sum_{h\in\mathbb{H}} a_{lh}^{(d)} \lambda_{lh} \left(\frac{V^{(i)}(\mathbf{S}^{(i)}, \mathbf{S}_{-l,h}^{(d)}) - V^{(i)}(\mathbf{S})}{\alpha_i+\beta}\right),
\end{aligned}
\tag{2.11}
$$

$$
\begin{aligned}
V^{(d)}(\mathbf{S},\mathbf{a}) = \frac{\beta V^{(d)}(\mathbf{S})}{\alpha_d+\beta} &+ \sum_{h\in\mathbb{H}} \left(b_h^{(i)} \wedge x_h^{(i)}\right) \mu_h^{(i)} \left(\frac{V^{(d)}(\mathbf{S}_{0,-h}^{(i)}, \mathbf{S}^{(d)}) - V^{(d)}(\mathbf{S})}{\alpha_d+\beta}\right) \\
&+ \sum_{h\in\mathbb{H}} \left(b_h^{(d)} \wedge x_h^{(d)}\right) \mu_h^{(d)} \left(\frac{r_h^{(d)} + V^{(d)}(\mathbf{S}^{(i)}, \mathbf{S}_{0,-h}^{(d)}) - V^{(d)}(\mathbf{S})}{\alpha_d+\beta}\right) \\
&+ \sum_{l\in\mathbb{L}} \sum_{h\in\mathbb{H}} a_{lh}^{(i)} \lambda_{lh} \left(\frac{V^{(d)}(\mathbf{S}_{-l,h}^{(i)}, \mathbf{S}^{(d)}) - V^{(d)}(\mathbf{S})}{\alpha_d+\beta}\right) \\
&+ \sum_{l\in\mathbb{L}} \sum_{h\in\mathbb{H}} a_{lh}^{(d)} \lambda_{lh} \left(\frac{V^{(d)}(\mathbf{S}^{(i)}, \mathbf{S}_{-l,h}^{(d)}) - V^{(d)}(\mathbf{S})}{\alpha_d+\beta}\right),
\end{aligned}
\tag{2.12}
$$

where $\mathbf{S}_{-l,h}^{(c)} = \left(\mathbf{W}^{(c)} - \mathbf{e}_l, \mathbf{X}^{(c)} + \mathbf{e}_h\right)$ and $\mathbf{S}_{0,-h}^{(c)} = \left(\mathbf{W}^{(c)}, \mathbf{X}^{(c)} - \mathbf{e}_h\right)$ for $c \in \mathbb{C}, l \in \mathbb{L}$, and $h \in \mathbb{H}$. In (2.11) and (2.12), $V^{(c)}(\mathbf{S}') - V^{(c)}(\mathbf{S})$ terms can be interpreted collectively as the marginal "benefit" by making the corresponding decision.

Remark 2.3.1. The MDP formulation given above can be easily extended to $k > 2$ classes of casualties. The formulation then will have k equations similar to (2.11) and (2.12) corresponding to k classes and there will be k terms within the maximization given in (2.7).

2.3 Greedy Algorithm for Optimal Policy

If we somehow know the values $V^{(c)}(\mathbf{S}, \mathbf{a})$ for all $\mathbf{S} \in \mathcal{S}$ and $\mathbf{a} \in \mathbb{A}_{\mathbf{S}}$, then, to find the optimal policy, we need to solve a integer program with the objective function (2.7), constraints (2.1) and (2.2), and decision variable $\mathbf{a}$. More specifically, for state $\mathbf{S} \in \mathcal{S}, l \in \mathbb{L}$, and $h \in \mathbb{H}$, let

$$\begin{aligned} m_{lh}^{(i)}(\mathbf{S}) &= \lambda_{lh} \left(\frac{V^{(i)}(\mathbf{S}_{-l,h}^{(i)}, \mathbf{S}^{(d)}) - V^{(i)}(\mathbf{S})}{\alpha_i + \beta} + \frac{V^{(d)}(\mathbf{S}_{-l,h}^{(i)}, \mathbf{S}^{(d)}) - V^{(d)}(\mathbf{S})}{\alpha_d + \beta} \right) \\ m_{lh}^{(d)}(\mathbf{S}) &= \lambda_{lh} \left(\frac{V^{(i)}(\mathbf{S}^{(i)}, \mathbf{S}_{-l,h}^{(d)} - V^{(i)}(\mathbf{S}))}{\alpha_i + \beta} + \frac{V^{(d)}(\mathbf{S}^{(i)}, \mathbf{S}_{-l,h}^{(d)} - V^{(d)}(\mathbf{S}))}{\alpha_d + \beta} \right). \end{aligned} \tag{2.13}$$

The constrained integer program for each $\mathbf{S} \in \mathcal{S}$ can be expressed as:

$$\begin{aligned} \max \quad & \sum_{l \in \mathbb{L}} \sum_{h \in \mathbb{H}} \sum_{c \in \mathbb{C}} a_{lh}^{(c)} m_{lh}^{(c)}(\mathbf{S}) + C \\ s.t. \quad & (2.1) \text{ and } (2.2), \end{aligned} \tag{2.14}$$

where C is a constant corresponding to the sum of first two lines in (2.11) and (2.12), which do not contain decision variable a. We will drop the state parameter $\mathbf{S}$ in $m_{lh}^{(c)}(\mathbf{S})$ when there is no risk of confusion in the rest of the chapter.

Algorithm 1 Finding optimal policy for given values of $V^{(c)}(\cdot)$

for all $l \in \mathbb{L}, h \in \mathbb{H}, c \in \mathbb{C}$ **do**
  $a_{lh}^{(c)} \leftarrow 0$
end for
$list \leftarrow \{(l,h,c) : l \in \mathbb{L}, h \in \mathbb{H}, c \in \mathbb{C}\}$
Sort-Descending(list, $m_{ih}^{(c)}(\mathbf{S})$) {Sort list in the descending order according to $m_{ih}^{(c)}(\mathbf{S})$}
for $k = 1$ to list.Length **do**
  $(l,h,c) \leftarrow list[k]$
  while $\sum_{h \in \mathbb{H}} a_{(lh)}^{(c)} < w_l^{(c)}$ **do**
    if $\sum_{l \in \mathbb{L}} \sum_{h \in \mathbb{H}} \sum_{c \in \mathbb{C}} a_{lh}^{(c)} < \mathcal{A}$ **then**
      $a_{lh}^{(c)} \leftarrow a_{lh}^{(c)} + 1$
    else
      BREAK
    end if
  end while
end for

Proposition 1. Algorithm 1 returns an optimal solution to (2.14) when $V^{(c)}(\mathbf{S})$'s are known for all $\mathbf{S} \in \mathcal{S}$ and $c \in \mathbb{C}$.

The action space, the feasible solutions, and the non-negative objective function of Problem (2.14) form a finite weighted matroid. Therefore, by [55] the optimal solution could be obtained using a greedy algorithm, such as Algorithm 1 which generalized Algorithm 1 in [49] to two classes. The proof of Proposition 1 is omitted.

Each triplet (l,h,c) in Algorithm 1 presents a potential dispatch of a class c casualty from location l to hospital h. According to Algorithm 1, ambulances will be dispatched in descending order of $m_{lh}^{(c)}$, where $m_{lh}^{(c)}$'s can be interpreted as the marginal reward of transporting a class c casualty from location l to medical facility h. Since finding exact values of $m_{lh}^{(c)}$'s can be computationally difficult even for reasonable problem sizes, we develop heuristic policies based on the greedy structure of (2.14) with different approximations for $m_{lh}^{(c)}$'s in Section 2.4.

2.4 Heuristics under Perfect Information on the State of Medical Facilities

In the aftermath of a mass casualty incident, figuring out the exact quantity and severity of all casualties at multiple casualty locations can be difficult. The situation at the scene may vary with time and the rescue progress. On the other hand, the status of the medical facilities is more likely to be accessible based on emergency drills, data on daily operations, emergency management plan, and information technology. Consequently, the state of the problem may be partially observable. In this section, we make use of the greedy structure from Proposition 1 to develop heuristic policies that only require information from medical facilities. We further modified the heuristic policies to accommodate uncertainty on the status of medical facilities in Section 2.5.

2.4 Myopic Heuristic (MYH)

The myopic policy takes into consideration a single casualty. It maximizes the expected survival probability for a casualty if he/she arrives at a medical facility at the next decision epoch. An ambulance that carries a class c casualty from location l towards facility h will reach the medical facility with probability λ_{lh}/β in the next decision epoch of the uniformized MDP defined in Section 2.3.2. If the state is $\mathbf{S} = \Big((\mathbf{W}^{(i)}, \mathbf{X}^{(i)}), (\mathbf{W}^{(d)}, \mathbf{X}^{(d)})\Big)$ and the next event is the arrival of this casualty at facility h, then he/she this casualty will become the $(x_h^{(c)} + 1)$th casualty at that facility. If $x_h^{(c)} + 1 > b_h^{(c)}$, this casualty will join the first-come-first-serve queue waiting for an available server and the time until service will be the sum of $x_h^{(c)} + 1 - b_h^{(c)}$ i.i.d.exponential random variables with rate $\tilde{\mu}_h^{(c)} = b_h^{(c)} \mu_h^{(c)}$. Otherwise, the casualty will start treatment right away. If the casualty fails to reach the facility, the expected survival probability in the next decision epoch will be 0. Then, we can approximate $m_{lh}^{(c)}(\mathbf{S})$ in Algorithm 1 by

$$m_G^{(c)}(l,h) = \lambda_{lh} \left(\frac{r_h^{(c)}}{\alpha_c + \beta}\right) \left(\frac{\mu_h^{(c)}}{\alpha_c + \mu_h^{(c)}}\right) \left(\frac{\tilde{\mu}_h^{(c)}}{\alpha_c + \tilde{\mu}_h^{(c)}}\right)^{\left[x_h^{(c)}+1-b_h^{(c)}\right]^+}, \tag{2.15}$$

where $[x]^+$ equals to x if $x \geq 0$ and 0 otherwise.

Remark 2.4.1. This is a generalization of the Myopic policy in [49] to multiple classes of casualties.

Remark 2.4.2. $m_G^{(c)}(l, h)$ values depend on the number of casualties at each medical facility while are independent of the numbers of casualties at casualty locations.

2.4 Policy Improvement Heuristic (PIH)

We generalize the policy improvement heuristic in [49] to the case with two casualty classes. We assume an infinite number of casualties in both immediate and delayed classes at all casualty locations and that all servers dedicated to the same class of casualties are pooled at each facility. Since we hypothetically divided each facility into two artificial facilities each serving only one class of casualties in our model, each artificial facility now can be approximated by a single server. To adopt the policy improvement method, we start with a simple static policy under which value functions i.e., the expected number of survivors, are easy to compute, and then apply one step of the policy improvement algorithm to get a state-dependent policy. We use Bernoulli splitting as the initial static policy: an empty ambulance will carry a class c casualty from location l to facility h with probability $\theta_{lh}^{(c)}$ independent of anything else at the beginning of each decision epoch. The constraints are: $\theta_{lh}^{(c)} > 0$ if and only if facility h is capable of treating class c casualties, $\theta_{lh}^{(c)} \geq 0, \forall c \in \mathbb{C}, l \in \mathbb{L}, h \in \mathbb{H}$, and $\sum_{c\in\mathbb{C}} \sum_{l\in\mathbb{L}} \sum_{h\in\mathbb{H}} \theta_{lh}^{(c)} \leq 1$. Both the initial splitting probabilities $\theta_{lh}^{(c)}$'s and service at facilities are independent of each other, therefore the Bernoulli splitting policy only depends on the number of casualties at each facility. We can then model each facility as a single server queue with arrival rate $\tilde{\lambda}_h^{(c)} = \mathcal{A} \sum_{l\in\mathbb{L}} \theta_{lh}^{(c)} \lambda_{lh}$ and service rate $\tilde{\mu}_h^{(c)} = b_h^{(c)} \mu_h^{(c)}$. Let $\mathbf{X} = (\mathbf{X}^{(i)}, \mathbf{X}^{(d)})$ represent the numbers of casualties at all medical facilities. Let $V_\infty^{(c)}(\mathbf{X})$ denote the expected total discounted survival probability for class c casualties under the Bernoulli splitting policy starting at initial state $\mathbf{X}$. By uniformizing the single server queue at each medical facility and incorporating the discount factor α_c, we can derive the following,

$$V_\infty^{(c)}(\mathbf{X}^{(i)}+\mathbf{e}_h,\mathbf{X}^{(d)})-V_\infty^{(c)}(\mathbf{X}^{(i)},\mathbf{X}^{(d)})=\begin{cases}\Delta^{(i)}(x_h^{(i)}) & \text{if } c=i,\\ 0 & \text{if } c=d,\end{cases}$$
$$V_\infty^{(c)}(\mathbf{X}^{(i)},\mathbf{X}^{(d)}+\mathbf{e}_h)-V_\infty^{(c)}(\mathbf{X}^{(i)},\mathbf{X}^{(d)})=\begin{cases}0 & \text{if } c=i,\\ \Delta^{(d)}(x_h^{(d)}) & \text{if } c=d,\end{cases}$$
$$\text{where } \Delta^{(c)}(x_h^{(c)})=\left(\frac{\tilde{\mu}_h^{(c)}r_h^{(c)}}{\tilde{\lambda}_h^{(c)}}\right)\left(\frac{\tilde{\mu}_h^{(c)}-\tilde{\lambda}_h^{(c)}+\alpha_c-\eta_h^{(c)}}{\tilde{\mu}_h^{(c)}-\tilde{\lambda}_h^{(c)}-\alpha_c-\eta_h^{(c)}}\right)\left(\frac{\tilde{\mu}_h^{(c)}+\tilde{\lambda}_h^{(c)}+\alpha_c-\eta_h^{(c)}}{2\tilde{\lambda}_h^{(c)}}\right)^{x_h^{(c)}},$$
$$\text{and } \eta_h^{(c)}=\sqrt{(\tilde{\mu}_h^{(c)}+\tilde{\lambda}_h^{(c)}+\alpha_c)^2-4\tilde{\lambda}_h^{(c)}\tilde{\mu}_h^{(c)}}. \tag{2.16}$$

We omit the derivation of (2.16) due to its similarity with the proof of Proposition 4 in Mills et al. [49]. Then, we approximate $m_{lh}^{(c)}(\mathbf{S})$ in (2.13) by

$$m_P^{(c)}(l,h)=\frac{\lambda_{lh}\Delta^{(c)}(x_h^{(c)})}{\alpha_c+\beta}. \tag{2.17}$$

We discuss the selection of the Bernoulli splitting probabilities $\theta_{lh}^{(c)}, l\in\mathbb{L}, h\in\mathbb{H}, c\in\mathbb{C}$ based on a fluid approximation in appledix B.

2.4 Ample Ambulances Heuristic (AAH)

Both myopic and policy improvement heuristics depend only on the number of casualties at medical facilities while ignoring the size of casualties at casualty locations. We develop a third heuristic taking into consideration the number of casualties at locations. To simplify the problem, we assume that there is ample transportation resource so that all casualties can be transported simultaneously and there is a single casualty location. Hence, we can drop the index associated with location.

The travel time from casualty location to facility h is exponentially distributed with rate λ_h. Now let $a_h^{(c)}$ denote the number of class c casualties transported to facility h. The number of dedicated servers for class c casualties at facility h is $b_h^{(c)}$ and service time follows an exponential distribution with rate $\mu_h^{(c)}$ as before. Suppose there are $x_h^{(c)}$ casualties at medical facility h at the decision epoch, then only $[b_h^{(c)} - x_h^{(c)}]^+$ of the servers are available initially. Suppose there are $w^{(i)}$ casualties triaged as immediate and $w^{(d)}$ casualties triaged as delayed at time 0. If $a_h^{(c)} \leq [b_h^{(c)} - x_h^{(c)}]^+$, all casualties sent to facility h will start treatment immediately upon arrival. Otherwise, there are more casualties sent to facility h than the number of servers available initially. Some of the casualties will have to join a queue. The one-period casualty allocation problem can be then formulated as

$$\begin{aligned} \max \quad & \sum_{c \in \mathbb{C}} \sum_{h \in \mathbb{H}} \Bigg[\left(a_h^{(c)} \wedge [b_h^{(c)} - x_h^{(c)}]^+ \right) r_h^{(c)} \int_0^\infty e^{-\alpha_c t} dF_{ch}^{(0)}(t) \\ & + r_h^{(c)} \sum_{k=1}^{\left[a_h^{(c)} - [b_h^{(c)} - x_h^{(c)}]^+\right]^+} \int_0^\infty e^{-\alpha_c t} dF_{ch}^{(k)}(t) \Bigg] \\ s.t. \quad & \sum_{h \in \mathbb{H}} a_h^{(c)} = w^{(c)}, \forall c \in \mathbb{C}, \end{aligned} \tag{2.18}$$

where $r_h^{(c)} e^{-\alpha_c t}$ represents the survival probability of a class c casualty who completes his/her treatment at facility h at time t, $F_{ch}^{(0)}(t)$ is the cumulative distribution function of the time that it takes starting from departure from the casualty location until service completion at facility h for a class c casualty if he/she starts treatment immediately upon arrival, and $F_{ch}^{(k)}(t)$ is the cumulative distribution of the time from departing the casualty location until service completion at facility h for the kth casualty in the queue not including those in service. Under exponential service and travel times, we have

$$\int_0^\infty e^{-\alpha_c t} dF_{ch}^{(0)}(t) = \left(\frac{\lambda_h}{\lambda_h + \alpha_c} \right) \left(\frac{\mu_h^{(c)}}{\mu_h^{(c)} + \alpha_c} \right). \tag{2.19}$$

We approximate the time until service completion for the kth casualty in the queue at a medical facility by the sum of his/her travel time, the service time for all casualties in front of him/her and his/her own service time. The actual time until service completion will be smaller since when he/she is traveling, casualties arrived earlier will already start their service. In this case, $F_{ch}^{(k)}(t)$ is the convolution of travel time distribution, first k service completion time distribution and her individual service time distribution which then leads to the following

$$\int_0^\infty e^{-\alpha_c t} dF_{ch}^{(k)}(t) = \left(\frac{\lambda_h}{\lambda_h + \alpha_c}\right)\left(\frac{\mu_h^{(c)}}{\mu_h^{(c)} + \alpha_c}\right)\left(\frac{\tilde{\mu}_h^{(c)}}{\tilde{\mu}_h^{(c)} + \alpha_c}\right)^k, \tag{2.20}$$

where $\tilde{\mu}_h^{(c)} = b_h^{(c)}\mu_h^{(c)}$ is the aggregate service rate for all $b_h^{(c)}$ class c servers at facility h.

Now we can evaluate and simplify (2.18) as

$$\begin{aligned} \max \quad & \sum_{c\in\mathbb{C}}\sum_{h\in\mathbb{H}} r_h^{(c)}\left(\frac{\lambda_h}{\lambda_h + \alpha_c}\right)\left(\frac{\mu_h^{(c)}}{\mu_h^{(c)} + \alpha_c}\right) \\ & \left\{\left(a_h^{(c)} \wedge [b_h^{(c)} - x_h^{(c)}]^+\right) + \left(\frac{\tilde{\mu}_h^{(c)}}{\alpha_c}\right)\left[1 - \left(\frac{\tilde{\mu}_h^{(c)}}{\tilde{\mu}_h^{(c)} + \alpha_c}\right)^{\left[a_h^{(c)} - [b_h^{(c)} - x_h^{(c)}]^+\right]^+}\right]\right\} \\ s.t. \quad & \sum_{h\in\mathbb{H}} a_h^{(c)} = w^{(c)}, \forall c \in \mathbb{C}. \end{aligned} \tag{2.21}$$

It is straight forward to modify the proof for Proposition 1 to prove that the optimal solution to (2.21) can be obtained by a greedy approach which assign casualties one by one to the facility with the largest $m_O(h)$, where

$$m_O(h) = \left(\frac{r_h^{(c)}\lambda_h}{\lambda_h + \alpha_c}\right)\left(\frac{\mu_h^{(c)}}{\mu_h^{(c)} + \alpha_c}\right)\left(\frac{\tilde{\mu}_h^{(c)}}{\tilde{\mu}_h^{(c)} + \alpha_c}\right)^{\left[\bar{a}_h^{(c)} - [b_h^{(c)} - x_h^{(c)}]^+ + 1\right]^+}, \tag{2.22}$$

and $\bar{a}_h^{(c)}$ denotes the number of class c casualties sent to facility h up until the casualty we are considering. We can apply this formulation when there are multiple casualty locations by letting $\bar{a}_h^{(c)}$ equals to the total number of casualties sent to facility h beforehand.

The reason we only considered one casualty location when developing this heuristic is that the assignment of casualties between multiple locations and multiple medical facilities under the ample ambulances assumption is a variation of the constrained weighted maximum bipartite problem which does not have an analytical solution and is complex to solve numerically. Furthermore, there is no guarantee that the modified numerical solution that works in the stochastic setting will perform well. Therefore, we focus on AAH developed based on a single casualty location assumption.

Notice the formulation in (2.22) is similar to the formulation for MYH in (2.15). There are two main differences: the myopic heuristic takes the travel rate as the probability of a casualty's arrival at a medical facility by the end of a decision epoch (the transportation rate λ_{lh} is a multiplier in (2.15)), while the transportation time is discounted in one-time allocation heuristic (corresponding to the $\lambda_{lh}/(\lambda_{lh}+\alpha_c)$ term in (2.22)). More importantly, the myopic heuristic considers a single casualty at each location and the real-time states at medical facilities while ignoring the casualties in transition. On the other hand, the one-time allocation heuristic considers the number of casualties sent to medical facilities as an approximation of the casualties in transition and in hospital. This difference will lead to two different formulations when we consider the uncertainty in the states of medical facilities in Session 2.5.

2.4 Shortest Completion Time Heuristic (SCH)

Shortest Completion Time Heuristic (SCH) was introduced in [49] under the name baseline dynamic policy, which always assigns casualties to the facility with the smallest expected time until service completion while ignoring all other casualties that have not yet arrived at medical facilities. Mathematically, let

$$m_S(l,h) = \lambda_{lh}^{-1} + \left(\tilde{\mu}_h^{(c)}\right)^{-1}\left[x_h + 1 - b_h^{(c)}\right]^{+} + \left(\mu_h^{(c)}\right)^{-1}. \tag{2.23}$$

A class c casualty will be transported to facility $h^{(c)} = \operatorname{argmin}_{\{k:b_k^{(c)}>0\}}\{m_S^{(c)}(l,k)\}$ from location l. Shortest completion time heuristic does not consider the different deterioration rates between the immediate and delayed casualty classes, thus does not prioritize casualties automatically. We assume priority is always given to the immediate class. At any casualty location, casualties in the delayed class will not be transported unless there is no immediate class left on scene. This assumption is consistent with the most commonly used triage method START [70].

2.4 Delta-Nearest Facility Heuristic (dNFH)

There is no standard for casualty distribution in practice to the best of our knowledge. However, in various past events, it is noted that some form of nearest hospital policy was used. Hence, we will compare the performance of heuristics policies proposed with a simple baseline policy called Delta-Nearest Facility Heuristic (dNFH) (2.4.5). dNFH is a static policy which transports casualties from a casualty location to medical facilities nearby. More precisely, for each casualty location $l \in \mathbb{L}$, we first identify the closest medical facility h^*, where the distance from l to h^* is $d(l, h^*)$. Then, we identify the set $\mathbb{H}_l$ that consists of all medical facilities such that the distance to casualty location l is within $d(l, h^*) + \delta$, where δ is a constant. Medical facilities in the set $\mathbb{H}_l$ are reasonably close to the casualty location. Finally, the next casualty will be transported to one of the medical facilities in the set $\mathbb{H}_l$ with probability proportional to the service capacity. Specifically, with a probability of $P_h = \frac{b_h^{(c)}}{\sum_{k \in \mathbb{H}_l} b_k^{(c)}}$ a class c casualty in location l will be transported to medical facility h. We again assume priority is always given to the immediate class as for SCH described in Section 2.4.4.

2.4 Shin and Lee's Heuristic (SLH)

We also compare our heuristics with the heuristic proposed by Shin and Lee [65]. In their model, there is only one casualty location, and immediate and delayed casualties share a single server at each medical facility with priority always assigned to immediate casualties. Medical facilities are classified into two tiers. A casualty treated at tier 1 (lower) facilities will have 20% less survival probability if treated at a tier 2 (high) facility. Service times are identically distributed for immediate and delayed casualties in their model. In order to implement their heuristic in our simulation frame work, we modified their heuristic as follows: firstly, at the casualty location (the heuristic does not work in multiple casualty locations setting), we find the target facilities h_c^* that have the smallest expected time until service completion for each class. Mathematically, $h_c^* = \text{argmin}_{h\in\mathbb{H}}\{\lambda_h^{-1} + \left(\mu_h^{(c)}\right)^{-1} + (x_h^{(c)} + Trans_h^{(c)})\tilde{\mu}_h^{(c)}\}$, where λ_h is the travel rate, $x_h^{(c)}$ is the number of class c casualties in the medical facility, $Trans_h^{(c)}$ is the number of class c casualties in transition (for a fair compassion with other heuristics, we ignore the information of casualties in transition for all heuristics i.e., $Trans_h^{(c)} = 0, \forall c \in \mathbb{C}, h \in \mathbb{H}$. Furthermore, for the same reason as mentioned in AAH in Section 2.4.3, extending the use of transition information to multiple casualty locations cases is not a trivial task), and $\tilde{\mu}_h^{(c)} = b_h^{(c)}\mu_h^{(c)}$ is the aggregate service rate for all $b_h^{(c)}$ class c servers at facility h. In the original heuristic, the authors used time until service. We changed it to time until service completion to be consistent with other heuristics. Notice, the optimal medical facility for each class is obtained in the same manner as in Shortest Completion Time Heuristic(SCH) (2.4.4).

After the target facilities are identified for each class, if only one class of casualties is left at the casualty location, then a casualty from that class will be selected. Otherwise, delayed casualties will be prioritized for transportation if $x_{(h_i^*)}^{(i)}\left(\lambda_{h_d^*}\tilde{\mu}_{h_d^*}^{(c)}\right) \geq x_{(h_d^*)}^{(d)}\left(\lambda_{h_i^*}\tilde{\mu}_{h_i^*}^{(c)}\right)$ or $x_{(h_i^*)}^{(i)}(\alpha_i + \alpha_d) \geq \tilde{\mu}_{h_i^*}^{(c)}$, immediate casualties will be prioritized otherwise. The selected casualty from class c will be transported to medical facility h_c^*.

2.5 Modified Heuristics under Incomplete Information on the State of Medical Facilities

All heuristic policies proposed (except for dNFH) in Section 2.4 rely on the state information of medical facilities. Such information may not available in real time due to the chaotic environment of a mass casualty incident and potentially damaged communication system.

In this section, we modify the heuristic policies discussed in Section 2.4 to accommodate for the uncertainty of state information of medical facilities. We assume, in the worst case, the medical facilities only announce their capacities for immediate and delayed classes at the very beginning of the mass casualty event and will never update this information again. We also assume the decision maker has the knowledge of the service rate at each facility from previous experiences. We do not consider the heuristic proposed by Shin and Lee [65] here since the scenario deviate significantly from their modeling assumption (perfect information on the states at facility and in transportation) and decision structure (prioritization and routing decision are make sequentially).

2.5 Myopic and Policy Improvement Heuristics under Incomplete Information on State

The number of casualties at each facility is the only information we need to use myopic and policy improvement heuristics as in (2.15) and (2.17). Let $\Lambda_h^{(c)}$ denote the decision maker's estimation of the number of class c casualties at facility h. We let $\Lambda_h^{(c)} = \left(x_h^{(c)} + y_h^{(c)} - \Psi_h^{(c)}(t)\right)^+$, where t is the time since the last update of information on state, $x_h^{(c)}$ denotes the number of class c casualties at facility h at the time of the last information update, y_h denotes the number of class c casualties sent to facility h up until this decision epoch since the last information update, and $\Psi(t)$ is a Poisson random variable with mean $b_h^{(c)}\mu_h^{(c)}t$ representing the number of possible departures from facility h assuming no server is idle during a period of length t. Note that $\Lambda_h^{(c)}$ under estimates the number of class c casualties at facility h because Poisson departure is assumed to occur at maximum rate. The assumption is reasonable as during a mass casualty event, medical facilities will be overwhelmed soon after casualties arrive.

We can use $\hat{m}_G^{(c)}(l,h)$ in place of $m_G^{(c)}(l,h)$ in (2.15) for MYH heuristic under incomplete state information.

$$\hat{m}_G^{(c)}(l,h) = \lambda_{l,h}\left(\frac{r_h^{(c)}}{\alpha_c+\beta}\right)\left(\frac{\mu_h^{(c)}}{\alpha_c+\mu_h^{(c)}}\right)E\left[\left(\frac{\tilde{\mu}_h^{(c)}}{\alpha_c+\tilde{\mu}_h^{(c)}}\right)^{[\Lambda_h^{(c)}+1-b_h^{(c)}]^+}\right], \tag{2.24}$$

and

$$E\left[\left(\frac{\tilde{\mu}_h^{(c)}}{\alpha_c+\tilde{\mu}_h^{(c)}}\right)^{(\Lambda_h^{(c)}+1-b_h^{(c)})^+}\right]$$
$$=\begin{cases} 1-\dfrac{F\left(x_h^{(c)}+y_h^{(c)}-b_h^{(c)};\tilde{\mu}_h^{(c)}t\right)-e^{\alpha_c t}\left(\frac{\tilde{\mu}_h^{(c)}}{\tilde{\mu}_h^{(c)}+\alpha_c}\right)^{\left(x_h^{(c)}+y_h^{(c)}+1-b_h^{(c)}\right)}F\left(x_h^{(c)}+y_h^{(c)}-b_h^{(c)};(\tilde{\mu}_h^{(c)}+\alpha_c)t\right)}{F\left(x_h^{(c)}+y_h^{(c)};\tilde{\mu}_h^{(c)}t\right)} & \text{if } x_h^{(c)}+y_h^{(c)}\geq b_h^{(c)} \\ 1 & \text{if } x_h^{(c)}+y_h^{(c)}< b_h^{(c)}, \end{cases}$$

where $F(\cdot,\xi)$ denotes the cumulative distribution function of a Poisson random variable with mean $\xi \geq 0$.

Similarly, $\hat{m}_P^{(c)}(l,h)$ will replace $m_P^{(c)}(l,h)$ in (2.17) for PIH heuristic:

$$\hat{m}_P^{(c)}(l,h) = \lambda_{i,j}\left(\frac{\tilde{\mu}_h^{(c)}r_h^{(c)}}{\tilde{\lambda}_h^{(c)}}\right)\left(\frac{\tilde{\mu}_h^{(c)}-\tilde{\lambda}_h^{(c)}+\alpha_c-\eta_h^{(c)}}{\tilde{\mu}_h^{(c)}-\tilde{\lambda}_h^{(c)}-\alpha_c-\eta_h^{(c)}}\right) \times E\left[\left(\frac{\tilde{\mu}_h^{(c)}+\tilde{\lambda}_h^{(c)}+\alpha_c-\eta_h^{(c)}}{2\tilde{\lambda}_h^{(c)}}\right)^{\Lambda_h^{(c)}}\right], \tag{2.25}$$

where

$$E\left[\left(\frac{\tilde{\mu}_h^{(c)}-\tilde{\lambda}_h^{(c)}+\alpha_c-\eta_h^{(c)}}{2\tilde{\lambda}_h^{(c)}}\right)^{\Lambda_h^{(c)}}\right]$$
$$=e^{-\tilde{\mu}^{(c)}t\left(\frac{\tilde{\mu}_h^{(c)}-\tilde{\lambda}_h^{(c)}+\alpha_c-\eta_h^{(c)}}{\tilde{\mu}_h^{(c)}+\tilde{\lambda}_h^{(c)}+\alpha_c-\eta_h^{(c)}}\right)}\left(\frac{\tilde{\mu}_h^{(c)}-\tilde{\lambda}_h^{(c)}+\alpha_c-\eta_h^{(c)}}{2\tilde{\lambda}_h^{(c)}}\right)^{x_h^{(c)}+y_h^{(c)}}\left[\frac{F\left(x_h^{(c)}+y_h^{(c)};\frac{2\tilde{\lambda}_h^{(c)}\tilde{\mu}_h^{(c)}t}{\tilde{\mu}_h^{(c)}+\tilde{\lambda}_h^{(c)}+\alpha_c-\eta_h^{(c)}}\right)}{F(x_h^{(c)}+y_h^{(c)};\tilde{\mu}_h^{(c)}t)}\right].$$

2.5 One-Time Allocation Heuristic under Incomplete Information on State

The one-time allocation heuristic is independent of the states of facilities except for their initial capacity which is assumed to be given. Thus AAH can easily adapt to this situation. We just need to interpret $\bar{b}_h^{(c)}$ as the capacity announced by facility h and $\bar{a}_h^{(c)}$ as the number of casualties sent to facility h right before the current one under consideration in (2.22).

2.5 Shortest Completion Time Heuristic under Incomplete Information on State

The shortest completion time heuristic needs to be modified as MYH and PIH. We will use the following formula to replace (2.23)

$$\hat{m}_S(l,h) = \lambda_{lk}^{-1} + \left(\mu_k^{(c)}\right)^{-1} + \left(\tilde{\mu}_k^{(c)}\right)^{-1} E\left[\left(\Lambda_k + 1 - b_k^{(c)}\right)^+\right] \tag{2.26}$$

and,

$$E\left[\left(\Lambda_k^{(c)} + 1 - b_k^{(c)}\right)^+\right]$$
$$= \begin{cases} \dfrac{\left(\frac{x_h^{(c)}+y_h^{(c)}+1-b_h^{(c)}}{b_h^{(c)}\mu_h^{(c)}}\right) F\left(x_h^{(c)}+y_h^{(c)}-b_h^{(c)};\tilde{\mu}_h^{(c)}t\right) - 1_{[x_h^{(c)}+y_h^{(c)}\geq b_h^{(c)}+1]} tF\left(x_h^{(c)}+y_h^{(c)}-b_h^{(c)}-1;\tilde{\mu}_h^{(c)}t\right)}{F\left(x_h^{(c)}+y_h^{(c)};\tilde{\mu}_h^{(c)}t\right)} & \text{if } x_h^{(c)}+y_h^{(c)} \geq b_h^{(c)}, \\ 0 & \text{if } x_h^{(c)}+y_h^{(c)} < b_h^{(c)}. \end{cases}$$

2.6 Simulation Study

Due to the high dimensionality of the state space of our MDP model, solving (2.11) and (2.12) even numerically to obtain the optimal solution requires an unrealistic amount of computational time and memory. Furthermore, the assumptions we made in order to have an analytically tractable MDP model may not reflect the reality. Such assumptions as preemptive ambulance dispatches, instantaneous return time from medical facilities to casually locations, exponential service time at medical facilities, and exponential ambulance travel times will be relaxed in our simulation. More realistic survival probability functions are used in the simulations as well instead of exponentially decaying survival probabilities. In this section, we compare all heuristics discussed in Section 2.4, namely MYH, PIH, AAH, SCH, SLH, and dNFH in the simulations as close to real mass casualty events as possible. Based on the characteristics of the events, such as major injury type, the composition of severities of casualties, casualty distribution, and the geographical scope, we categorize the mass casualty incidents into four types: single-location terrorist attack, multiple-location terrorist attack, major traffic accident, and earthquakes. More details of those events will be introduced in the remainder of this section.

2.6 Simulation Parameters

Casualty and medical facility locations are generated uniformly on a two-dimensional plane. A total number of x_l casualties at location l is generated randomly from a uniform distribution. The range of the number of casualties is event specific. We assume p percent of total casualties are triaged into the immediate class while the rest are in the delayed class. The percentage p is also generated uniformly between 0.1 to 0.4 (an estimate from Emergency Medical Doctor Lane M. Smith and J. Winslow). Therefore, the number of immediate casualties is $x_l^{(i)} = px_l$ and the number of delayed casualties is $x_l^{(d)} = (1-p)x_l$ for all $l \in \mathbb{L}$.

We focused on two specific types of traumatic injuries – penetrating and blunt in mass casualty incidents. Penetrating injuries are commonly seen at terrorist shooting events and blunt injuries occur frequently during traffic accidents and natural disasters such as earthquakes. To the best of our knowledge, the only work that provide survival probability estimates for penetrating and blunt injuries are [63, 64]. In these paper, the initial conditions of casualties are evaluated using the RPM score which ranges from 1 to 13 with 13 being the most severe. Then Delphi method is used to estimate the deterioration of casualties with different initial conditions. The survival probabilities are obtained, in 30 minutes intervals within the range of 6 hours for 13 possible initial RPM scores, using logistic regression. For both the penetrating and blunt injuries, we compute the survival probabilities of the delayed class using the mean of the survival probabilities with initial RPM scores between 1 and 4 and the survival probabilities of the immediate class using the mean of the survival probabilities with initial RPM scores between 5 and 9. We then fit the survival probabilities using a three-parameter shifted log-logistic function as follows

$$f(t) = \frac{\beta_0}{1 + \left(\frac{t}{\beta_1}\right)^{\beta_2}}. \tag{2.27}$$

We also used exponential functions of the form $f(t) = \beta_0 e^{-\beta_1 t}$ to get continuous survival probability functions used in our simulation.

The shifted log-logistic function provides a good fit on the immediate and the delayed class for both penetrating and blunt injuries in terms of mean square error (MSE). We will use the fitted log-logistic functions to generate survival probabilities in the simulation. The exponential function provides a good fit for immediate class but not such a good fit for the delayed class. Nevertheless, since our heuristics assumed exponential decay for health deterioration, we use these exponents as the discount factors in all heuristics. The fitted parameters are provided in Table 2.1.

		Shifted log-logistic				Exponential		
Injury	Triage	β_0	β_1	β_2	MSE	β_0	β_1	MSE
Pen.	Immediate	0.3510	35.838	1.9886	9.93e-05	0.3563	-0.0207	8.03e-05
	Delayed	0.9124	213.5976	2.3445	6.15e-04	1.0219	-0.0038	4.40e-03
Blt.	Immediate	0.6049	67.1604	1.5485	2.41e-04	0.6053	-0.0096	2.61e-04
	Delayed	0.9527	328.1880	2.3155	6.85e-05	1.0400	-0.0021	2.80e-03

Table 2.1: Parameter Estimation for Survival Probabilities

Since the Delphi estimates assume the deterioration of survival probabilities under minimum medical intervention, it is not reasonable to assume that the survival probability of a casualty under treatment at a medical facility also deteriorates at the same rate. Therefore, in the simulations, the survival probabilities are collected at the beginning of service instead of service completion.

The transportation resource is assumed to be ambulances that are all identical and travel at an average speed of 40 mph. The number of available ambulances is event dependent. In the simulation, we explicitly take into account the travel time from facilities back to casualty location. The travel time from two locations are assumed to be lognormal distributed (according to Ingolfsson et al. [31]) with mean travel time equals the distance divided by the average speed. l_1-norm (commonly known as the Manhattan distance) is used to compute the distance between any location-facility pair.

We considered two type of medical facilities – level 1 trauma centers and level 3 trauma centers that differ mainly in terms of capacity. Level 1 trauma centers are expected to be able to handle more casualties than level 3 trauma centers. The capacities of immediate class casualties at level 3 trauma centers are very limited. In our simulation, a medical facility will be a level 1 trauma center with probability 0.58 according to McLay and Mayorga [46]. The number of servers for immediate casualties is uniformly distributed between 5 to 8 and the number of servers for delayed casualties is uniformly distributed from 12 to 20. With probability 0.42, a medical facility will be a community hospital with 2 to 3 servers for immediate casualties and 6 to 10 servers for delayed casualties also uniformly distributed. We assume the service time distribution is identical for a given injury type and triage class at level 1 trauma center and community hospital. The service time for both penetrating and blunt injuries will follow an exponential distribution with mean service time of 90 minutes for the immediate class and mean service time of 180 minutes for the delayed class. The reason the mean service time for immediate class is shorter than the mean service time for the delayed class is that we considered only the life-saving procedures for immediate casualties. The surgical procedures for immediate casualties tend to be damage control surgery and further treatments which are not emergent will be deferred thus not included as service time. Other parameters are based on the discussions with Emergency Medical Doctors Lane M. Smith and James Winslow of Wake Forest University, NC.

2.6 Implementation of Heuristic Policies in Simulations

Since we relaxed the assumptions that ambulance assignments are preemptive and return time from medical facilities to casualty locations are instantaneous in the simulations, we modify our heuristic policies accordingly.

When an ambulance arrives at casualty location $l \in \mathbb{L}$ and prefect information on the states of medical facilities are available at all facilities $h \in \mathbb{H}$, heuristics for complete information on the state of medical facilities derived in Section 2.4 will be used in the corresponding simulations.

We explicitly consider the return time from facilities to casualty locations in the simulations. Under the assumption of perfect state information on medical facilities, when an ambulance arrives at facility $h \in \mathbb{H}$, MYH will send the ambulance back to location $l^* = \text{argmax}_l\{\lambda_{lh} \max_{q\in\mathbb{H},c\in\mathbb{C}} m_G^{(c)}(l,q)\}$; PIH will assign the ambulance to location $l^* = \text{argmax}_l\{\lambda_{lh} \max_{q\in\mathbb{H},c\in\mathbb{C}} m_P^{(c)}(l,q)\}$; AAH will dispatch the ambulance to location $l^* = \text{argmax}_l\{\lambda_{lh} \max_{q\in\mathbb{H},c\in\mathbb{C}} m_O^{(c)}(l,q)\}$; SCH will send the ambulance will go to location $l^* = \text{argmax}_l\{\lambda_{lh}/ \max_{q\in\mathbb{H},c\in\mathbb{C}} m_S^{(c)}(l,q)\}$; and the ambulance following dNFH will return to the closest casualty location where casualties are still waiting. SLH is designed for single casualty location, hence the return decision is irrelevant. When the information on the state of medical facilities is incomplete, $\hat{m}$ indices defined in Section2.5 will be used in place of m indices for all heuristics.

We also need to update the estimated arrival rate $\tilde{\lambda}_h^{(c)}$ in the policy improvement heuristic to consider the travel time back to casualty locations. The new approximated arrival rate to facility h becomes

$$\tilde{\lambda}_h^{(c)} = \mathcal{A} \sum_{l\in\mathbb{L}} \left[\theta_{lh}^{(c)} \sum_{c'\in\mathbb{C}} \sum_{q\in\mathbb{H}} \sum_{l'\in\mathbb{L}} \theta_{l'q}^{(c')} (\lambda_{lh}^{-1} + \lambda_{lq}^{-1})^{-1} \right]$$

Comparing this with the $\tilde{\lambda}_h^{(c)}$ in Section 2.4.2, we used the average round trip travel time for all possible 'facility-casualty-facility h' route to estimate the arrival rate to medical facility h.

2.6 Simulation Scenarios and Results

We designed four types of MCIs in our simulation study from observing all kind of MCIs in reality – single-location terrorist attacks, multiple-location terrorist attacks, single-location major traffic accidents, and earthquakes. These four types of MCIs are distinguished by different characteristics, such as the numbers of casualties, the geographical distribution of casualties, main injury type, and the number of available medical facilities. Within each type of MCIs, we randomized the number of casualty locations (for those types involving multiple casualty locations), the numbers of casualties at those locations, the triage outcome (the composition of the immediate and delayed casualties), the number of ambulances, the number of medical facilities and their capacities, and the geographical locations of casualty locations and medical facilities to generate 300 scenarios. We select the locations for casualty location(s) and medical facilities uniformly without replacement from a two-dimensional integer lattice. The distance between each location pair is computed using l_1-norm (commonly known as the Manhattan distance). The generation of other parameters is MCI type-dependent thus described in the respective subsections.

We repeated the simulations under the assumption of both perfect and incomplete information on the state of medical facilities using the generated scenarios. For each individual scenario, we replicated the simulation 100 times. In each replication, the total numbers of survivors are recorded for all heuristic policies with a synchronized stream of random numbers. We compared all other heuristics against the dNFH heuristic and reported the percentage improvement as the results. We provided statistics including the minimum ("Min"), first quantile ("Q1"), median ("Med"), third quartile ("Q3"), maximum ("Max"), mean, 95% confidence interval ("Lower" and "Upper") of the percentage improvement. In addition,

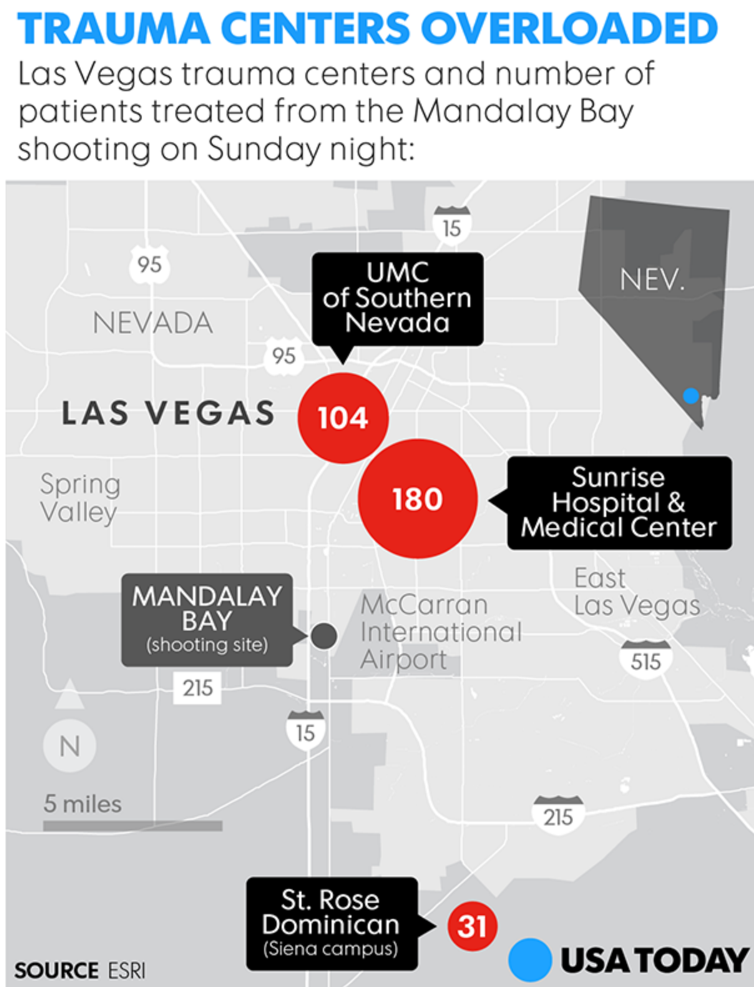

Figure 2.2: Casualty distribution in 2017 Las Vegas shooting

"#sig" denotes the number of scenarios where each heuristic is better than dNFH policy at a statistical significance level of 5%. When the state of medical facilities is known perfectly, we compared SLH heuristic and all other heuristics introduced in Section 2.4 against the dNFH heuristic. When the state of medical facilities is incomplete, we compared all other heuristics introduced in Section 2.5 against the dNFH heuristic. All simulations were coded and executed using Matlab.

2.6.3.1 Single-location Terrorist Attacks

Most terrorist attacks happen at a single location. Recent examples include the 2016 Orlando nightclub shooting in Florida, U.S., the 2017 London Bridge attack in England,

and the 2019 Christchurch mosque shooting in New Zealand, the 2017 Las Vegas shooting in Nevada, U.S. (refer to Figure 2.3). During these terrorist attacks, almost all injured casualties concentrated at a small neighborhood. A majority of the casualties suffered penetrating injuries. We categorize these events with a single incident location and predominantly penetrating injuries as single-location terrorist attacks.

In our simulation experiments, the number of casualties varied from dozens to hundreds. The percentage of immediate casualties is uniformly distributed between 10% and 40%. We assumed the casualty location and medical facilities are within a region of 10 by 10 miles^2. We varied the numbers of ambulances from 10 to 50 and the number of medical facilities from 2 to 4 in the simulations.

The results when the information on the state of medical facilities is available to the decision-maker are presented in Tables 2.2, 2.3, 2.4 and 2.5. The results when the information on the state of medical facilities is incomplete to the decision-maker are presented in Tables 2.6, 2.7 and 2.8. The first three columns provide the range of the number of casualties, the number of ambulances, and the number of medical facilities. When the number of casualties is small, existing medical resources are sufficient. According to the simulation results for both perfect information and incomplete information on the state of medical facilities (refer to the first row of Tables 2.2 and 2.6), SCH and SLH perform slightly better than dNFH, while, MYH, PIH, and AAH perform worse than dNFH.

As the number of casualties increases (refer to rows 2 through 4 in Table 2.2 followed by rows 1 through 3 in Table 2.3, and rows 2 through 4 in Table 2.6 followed by rows 1 through 3 in Table 2.7), on average all dynamic heuristics (PIH, AAH, MYH, SLH, and SCH) perform better than the static policy dNFH. AAH performs the best for all scenarios except the last one where there are 100 to 150 casualties, 30 ambulances and 3 medical facilities whose states are not up to date (refer to row 3 in Table 2.7), where SCH out performs AAH on average by less than 1%. Among all 12 scenarios, AAH performs the best in terms of the median. As the number of ambulances increases, while fixing the number of medical facilities (refer to rows

2 through 4 in Table 2.2 or rows 1 through 3 in Table 2.3 and rows 2 through 4 in Table 2.6 or rows 1 through 3 in Table 2.7), the advantages of dynamic heuristics over dNFS shrink. This could be because service capacity at medical facilities becomes the main bottleneck as the number of ambulances increases. Therefore, the routing of ambulances has a lesser impact. Notice when the numbers of ambulances and medical facilities are relatively small, SCH performs worse than AAH and PIH (refer to row 2 in Table 2.2, row 1 in Table 2.3, row 2 in Table 2.6 and row 1 in Table 2.7). As the numbers of ambulances and medical facilities increase, the performance of SCH improves and become comparable with AAH (refer to row 2 in Table 2.2, row 1 in Table 2.3, row 2 in Table 2.6 and row 1 in Table 2.7). It implies, even though medical resources are overwhelmed compared to the demand for all MCIs, AAH and PIH are more suitable in situations where the medical resources are extremely limited while SCH is suitable in situations where more medical resources are available.

When hundreds of casualties are involved (see Tables 2.4, 2.5 and 2.8), the advantages of dynamic heuristics are obvious. AAH is always the best performer when the state of medical facilities are incomplete (refer to Table 2.4 and 2.5). When the state is known, AAH performs the best except when there are 10 ambulances and 4 medical facilities (refer to row 4 in Table 2.8) where PIH out performs AAH. PIH performs almost as good as AAH in many other scenarios as can be seen in rows 2, 3, and 6 of Table 2.8.

2.6.3.2 Multiple-location Terrorist Attacks

In the events like the 2015 Paris terrorist attack in France, the attacks occurred almost simultaneously. Casualties were distributed at six different locations miles away from each other. More than ten hospitals were involved [29] (refer to Figure 2.3). Similar events also include 2015 London bombings in England. Among these events, a majority of critical casualties have penetrating injuries due to use if bombs and guns during these attacks. We categorized those events as multiple-location terrorist attacks. We simulated the multiple-location terrorist attacks with 3 or 4 casualty locations in accordance with past events. At each casualty location, we assumed the total number of casualties is uniformly distributed between 30 to 150. Among those casualties, the percentage of immediate casualties is uniformly distributed between 10% to 40%. We again assumed all casualty locations and medical facilities are located within a 10 by 10 miles^2 region. We only simulate under the assumption of imperfect information on the state of medical facilities here. The simulation results are presented in Table 2.9. The first column of Table 2.9 represents the number of casualty locations, the number of ambulances, and the number of medical facilities.

All dynamic heuristics perform better than dNFH in terms of median and mean. The improvement of PIH, AAH, and MYH over dNFH decreases as the number of ambulances increases while fixing the number of medical facilities (refer to rows 1, 2, and 3 or 4, 5, and 6 in Table 2.9). AAH performs the best among the proposed heuristics in terms of the median and the mean, and the performance of PIH is close to AAH in terms of the mean especially when the number of ambulances is small. The performance of SCH is far behind other dynamic heuristics when the number of ambulances is small but comparable with other dynamic heuristics as the number of ambulances gets larger.

No. Pat.	No. Amb.	No. Hsp.	Heuristic Policies	Min	Q1	Med	Q3	Max	Confidence Interval Lower	Mean	Upper	#sig.
20-50	10	2	PIH	-20.96%	-3.11%	-1.29%	0.63%	41.82%	-0.89%	-0.22%	0.45%	79
			MYH	-25.50%	-5.16%	-2.35%	-0.31%	38.86%	-2.96%	-2.30%	-1.64%	63
			AAH	-7.07%	-2.37%	-0.86%	2.14%	73.74%	1.15%	2.15%	3.15%	115
			SLH	-7.78%	-0.58%	0.53%	2.08%	72.88%	2.04%	3.03%	4.02%	167
			SCH	-7.67%	-0.56%	0.79%	2.36%	72.69%	1.89%	2.88%	3.86%	190
100-150	10	2	PIH	-8.36%	6.82%	17.62%	45.41%	171.01%	28.22%	31.57%	34.92%	273
			MYH	-9.69%	5.96%	18.29%	47.10%	175.32%	28.28%	31.72%	35.16%	262
			AAH	-0.88%	9.38%	21.73%	54.97%	185.48%	33.22%	36.88%	40.54%	286
			SLH	-5.05%	5.61%	15.30%	41.61%	175.06%	26.09%	29.48%	32.86%	280
			SCH	-6.71%	0.00%	3.08%	9.52%	165.57%	11.80%	14.79%	17.79%	195
100-150	20	2	PIH	-9.71%	1.13%	5.13%	13.73%	230.76%	13.92%	17.02%	20.12%	219
			MYH	-10.66%	0.22%	4.68%	14.14%	226.48%	13.23%	16.33%	19.44%	200
			AAH	-1.22%	2.50%	6.77%	16.08%	233.88%	16.23%	19.47%	22.72%	255
			SLH	-7.03%	0.18%	4.14%	12.54%	223.19%	13.37%	16.55%	19.73%	195
			SCH	-8.03%	-3.12%	-0.56%	1.58%	217.01%	7.44%	10.56%	13.68%	86
100-150	30	2	PIH	-12.11%	-0.34%	2.18%	8.73%	159.52%	11.47%	14.45%	17.43%	166
			MYH	-12.16%	-0.67%	1.77%	8.08%	161.12%	10.89%	13.87%	16.86%	156
			AAH	-1.62%	1.15%	3.36%	9.37%	165.09%	14.10%	17.30%	20.50%	219
			SLH	-8.01%	-1.62%	0.81%	6.95%	163.40%	11.23%	14.37%	17.51%	144
			SCH	-9.46%	-3.70%	-1.90%	0.14%	160.97%	7.90%	11.05%	14.19%	60

Table 2.2: Percentage improvement over dNFH in single-location terrorist attacks with perfect information on the state of medical facilities (Part 1).

No. Pat.	No. Amb.	No. Hsp.	Heuristic Policies	Min	Q1	Med	Q3	Max	Confidence Interval Lower	Mean	Upper	#sig.
100-150	10	3	PIH	-9.95%	7.75%	17.28%	36.04%	186.92%	23.94%	26.99%	30.04%	276
			MYH	-16.52%	4.47%	15.30%	33.00%	189.34%	22.08%	25.23%	28.39%	261
			AAH	0.52%	14.18%	28.28%	47.98%	219.67%	34.50%	38.05%	41.60%	300
			SLH	-1.96%	9.31%	21.34%	34.55%	205.57%	27.85%	31.20%	34.55%	296
			SCH	-5.47%	4.98%	11.57%	24.05%	200.10%	19.74%	22.97%	26.20%	277
100-150	20	3	PIH	-16.82%	1.15%	7.57%	23.21%	245.73%	15.72%	18.88%	22.04%	231
			MYH	-18.80%	-0.68%	5.82%	22.54%	248.02%	13.85%	17.01%	20.18%	204
			AAH	-1.73%	4.40%	13.13%	33.60%	255.10%	22.78%	26.32%	29.85%	272
			SLH	-6.24%	2.53%	10.55%	29.98%	250.65%	20.17%	23.64%	27.11%	245
			SCH	-6.78%	-0.47%	2.98%	26.75%	241.78%	16.47%	19.94%	23.41%	184
100-150	30	3	PIH	-13.85%	-0.77%	2.45%	17.32%	225.13%	12.22%	15.74%	19.27%	177
			MYH	-19.72%	-1.80%	0.95%	16.09%	222.84%	10.79%	14.31%	17.82%	150
			AAH	-2.63%	1.15%	4.51%	25.23%	229.99%	17.25%	21.04%	24.83%	237
			SLH	-7.86%	-1.14%	1.93%	22.78%	221.46%	14.22%	17.94%	21.66%	172
			SCH	-9.21%	-3.14%	-0.54%	20.97%	218.25%	11.80%	15.53%	19.25%	129

Table 2.3: Percentage improvement over dNFH in single-location terrorist attacks with perfect information on the state of medical facilities (Part 2).

No. Pat.	No. Amb.	No. Hsp.	Heuristic Policies	Min	Q1	Med	Q3	Max	Confidence Interval			#sig.
									Lower	Mean	Upper	
150-500	10	3	PIH	2.20%	41.07%	83.57%	155.92%	629.22%	103.19%	112.79%	122.39%	300
			MYH	0.13%	41.17%	81.40%	154.97%	685.52%	105.04%	115.35%	125.67%	299
			AAH	5.36%	50.33%	100.45%	176.32%	735.38%	120.34%	131.28%	142.21%	300
			SLH	2.69%	41.05%	81.06%	152.82%	694.30%	99.46%	108.70%	117.94%	300
			SCH	-13.47%	9.80%	24.10%	49.64%	258.81%	31.70%	35.43%	39.15%	282
150-500	30	3	PIH	-2.85%	6.40%	22.60%	51.43%	308.40%	34.78%	39.73%	44.68%	277
			MYH	-3.69%	5.67%	21.02%	50.66%	311.06%	33.87%	38.81%	43.76%	272
			AAH	-0.05%	8.22%	26.05%	54.56%	321.57%	37.46%	42.54%	47.63%	292
			SLH	-6.23%	3.94%	20.69%	49.81%	315.05%	32.89%	37.82%	42.75%	260
			SCH	-9.92%	-2.07%	2.52%	33.83%	260.61%	19.20%	23.52%	27.85%	163
150-500	50	3	PIH	-7.23%	3.41%	13.80%	45.42%	317.76%	27.04%	31.42%	35.80%	264
			MYH	-8.05%	2.88%	12.74%	44.26%	318.31%	26.39%	30.75%	35.12%	254
			AAH	-0.47%	4.72%	14.99%	48.47%	331.32%	28.92%	33.42%	37.92%	277
			SLH	-7.81%	0.36%	10.43%	42.14%	312.06%	23.78%	28.10%	32.42%	212
			SCH	-9.10%	-3.72%	-0.50%	34.30%	313.21%	16.27%	20.51%	24.75%	137

Table 2.4: Percentage improvement over dNFH in single-location terrorist attacks with perfect information on the state of medical facilities (Part 1).

No. Pat.	No. Amb.	No. Hsp.	Heuristic Policies	Min	Q1	Med	Q3	Max	Confidence Interval Lower	Confidence Interval Mean	Confidence Interval Upper	#sig.
150-500	10	4	PIH	3.51%	37.97%	73.39%	139.34%	479.76%	94.39%	102.86%	111.34%	300
			MYH	0.71%	34.21%	73.13%	143.95%	514.02%	96.19%	105.42%	114.65%	299
			AAH	8.20%	55.96%	94.14%	171.54%	541.55%	118.24%	128.19%	138.14%	300
			SLH	4.92%	42.45%	76.40%	141.12%	494.85%	97.90%	106.62%	115.35%	300
			SCH	-4.47%	17.07%	29.22%	50.40%	277.04%	37.83%	41.83%	45.83%	296
150-500	30	4	PIH	-4.86%	9.17%	26.99%	53.66%	429.04%	35.48%	40.08%	44.68%	276
			MYH	-9.09%	7.91%	25.02%	51.96%	424.88%	34.22%	38.81%	43.40%	267
			AAH	-0.08%	12.44%	30.41%	58.43%	436.10%	39.77%	44.55%	49.33%	297
			SLH	-5.12%	8.24%	26.31%	53.28%	426.86%	35.25%	39.88%	44.51%	273
			SCH	-6.52%	0.74%	16.08%	37.92%	424.51%	25.56%	30.05%	34.54%	224
150-500	50	4	PIH	-5.86%	6.10%	23.39%	49.24%	477.59%	36.30%	42.38%	48.45%	266
			MYH	-6.73%	5.45%	22.59%	48.86%	480.25%	35.45%	41.51%	47.57%	260
			AAH	-1.27%	7.66%	26.10%	53.51%	490.04%	39.75%	46.04%	52.32%	286
			SLH	-6.76%	3.37%	21.65%	46.68%	466.31%	34.24%	40.24%	46.24%	247
			SCH	-9.83%	-1.81%	16.11%	40.90%	456.40%	28.12%	34.05%	39.99%	192

Table 2.5: Percentage improvement over dNFH in single-location terrorist attacks with perfect information on the state of medical facilities (Part 2).

No. Pat.	No. Amb.	No. Hsp.	Heuristic Policies	Min	Q1	Med	Q3	Max	Confidence Interval Lower	Confidence Interval Mean	Confidence Interval Upper	#sig.
20-50	10	2	PIH	-16.00%	-3.10%	-1.47%	0.83%	62.93%	-0.10%	0.86%	1.82%	79
			AAH	-7.33%	-2.90%	-0.74%	2.17%	83.68%	1.73%	2.97%	4.21%	116
			MYH	-27.86%	-7.07%	-3.49%	-1.18%	60.45%	-4.19%	-3.29%	-2.40%	46
			SCH	-0.37%	0.64%	1.23%	2.16%	74.67%	3.64%	4.78%	5.93%	276
100-150	10	2	PIH	-10.35%	3.46%	11.90%	31.44%	224.07%	19.11%	21.91%	24.72%	256
			AAH	-6.23%	4.79%	13.06%	36.50%	219.40%	22.09%	24.97%	27.85%	276
			MYH	-16.95%	0.88%	8.97%	27.11%	231.40%	15.54%	18.32%	21.09%	220
			SCH	-15.74%	0.63%	1.47%	2.90%	153.40%	5.69%	7.49%	9.30%	196
100-150	20	2	PIH	-20.49%	-0.16%	2.47%	12.26%	190.81%	10.84%	13.67%	16.50%	176
			AAH	-2.90%	0.64%	3.33%	14.09%	195.90%	13.53%	16.54%	19.55%	206
			MYH	-21.88%	-2.51%	0.59%	8.39%	172.50%	8.06%	10.84%	13.62%	133
			SCH	-1.25%	0.71%	1.25%	2.03%	175.49%	9.98%	12.77%	15.56%	185
100-150	30	2	PIH	-19.74%	-0.55%	1.33%	6.78%	211.99%	8.28%	11.13%	13.99%	150
			AAH	-2.97%	0.36%	2.02%	7.70%	214.12%	11.04%	14.02%	17.00%	189
			MYH	-29.57%	-2.62%	-0.26%	3.15%	209.35%	5.74%	8.60%	11.47%	103
			SCH	-1.14%	0.58%	1.10%	1.82%	199.11%	9.43%	12.32%	15.20%	167

Table 2.6: Percentage improvement over dNFH in single-location terrorist attacks with incomplete information on the state of medical facilities (Part 1).

No. Pat.	No. Amb.	No. Hsp.	Heuristic Policies	Min	Q1	Med	Q3	Max	Confidence Interval Lower	Confidence Interval Mean	Confidence Interval Upper	#sig.
100-150	10	3	PIH	-15.46%	2.23%	10.07%	25.27%	219.07%	15.08%	17.49%	19.89%	246
			AAH	-5.26%	4.53%	13.40%	28.43%	221.52%	18.82%	21.36%	23.90%	276
			MYH	-24.13%	-3.84%	2.95%	15.34%	188.65%	7.06%	9.29%	11.52%	166
			SCH	-12.01%	1.98%	3.27%	11.01%	176.81%	7.75%	9.60%	11.45%	258
100-150	20	3	PIH	-17.42%	-1.44%	2.77%	11.67%	202.52%	8.77%	11.52%	14.27%	183
			AAH	-5.57%	0.95%	5.35%	17.56%	222.01%	13.68%	16.76%	19.85%	228
			MYH	-22.07%	-6.00%	-1.30%	6.65%	196.48%	3.37%	6.07%	8.78%	113
			SCH	-2.54%	1.49%	2.24%	16.12%	215.40%	12.00%	14.84%	17.68%	278
100-150	30	3	PIH	-17.98%	-1.32%	1.29%	12.52%	290.81%	10.43%	14.17%	17.90%	161
			AAH	-5.04%	0.31%	2.84%	19.20%	301.38%	15.25%	19.31%	23.37%	207
			MYH	-29.78%	-4.30%	-1.59%	5.06%	286.47%	5.89%	9.62%	13.34%	101
			SCH	-0.30%	1.37%	2.03%	19.68%	299.37%	15.47%	19.46%	23.44%	281

Table 2.7: Percentage improvement over dNFH in single-location terrorist attacks with incomplete information on the state of medical facilities (Part 2).

No. Pat.	No. Amb.	No. Hsp.	Heuristic Policies	Min	Q1	Med	Q3	Max	Confidence Interval Lower	Mean	Upper	#sig.
150-500	10	3	PIH	0.70%	36.79%	71.05%	147.75%	862.19%	98.55%	109.12%	119.69%	299
			AAH	4.55%	38.58%	75.22%	154.87%	826.44%	100.22%	110.55%	120.88%	300
			MYH	-3.28%	30.18%	65.75%	131.82%	745.73%	89.46%	99.36%	109.25%	294
			SCH	-35.25%	-2.72%	0.57%	3.46%	135.02%	4.01%	6.18%	8.35%	116
150-500	30	3	PIH	-2.10%	12.63%	25.92%	55.84%	352.64%	39.11%	44.10%	49.08%	286
			AAH	0.01%	13.06%	26.94%	60.13%	362.28%	40.85%	45.97%	51.10%	292
			MYH	-4.77%	9.27%	23.02%	55.63%	334.39%	35.91%	40.80%	45.69%	270
			SCH	-3.54%	0.68%	1.43%	31.41%	246.04%	19.56%	23.48%	27.39%	177
150-500	50	3	PIH	-2.19%	4.06%	14.98%	48.39%	353.42%	34.61%	41.09%	47.57%	267
			AAH	-1.04%	4.62%	15.98%	50.24%	362.78%	36.25%	42.86%	49.47%	277
			MYH	-4.57%	2.49%	12.66%	45.53%	348.13%	32.59%	39.03%	45.47%	249
			SCH	-1.36%	0.41%	1.24%	37.81%	323.15%	25.16%	30.65%	36.14%	160
150-500	10	4	PIH	0.95%	44.09%	71.57%	149.14%	512.81%	94.64%	102.74%	110.83%	300
			AAH	5.93%	44.77%	68.10%	147.94%	482.95%	93.57%	101.34%	109.12%	300
			MYH	-5.02%	33.37%	56.62%	129.49%	514.05%	81.69%	89.37%	97.05%	298
			SCH	-49.98%	-5.81%	0.12%	3.94%	150.67%	-2.45%	-0.94%	0.57%	136
150-500	30	4	PIH	-5.55%	11.53%	28.10%	56.06%	609.81%	43.05%	50.14%	57.23%	290
			AAH	0.39%	12.23%	31.30%	57.99%	646.52%	46.24%	53.58%	60.93%	296
			MYH	-8.09%	7.29%	23.29%	52.00%	604.01%	38.73%	45.74%	52.74%	265
			SCH	-2.89%	1.22%	10.43%	32.47%	358.36%	24.97%	30.59%	36.21%	239
150-500	50	4	PIH	-9.02%	4.35%	20.19%	42.77%	408.52%	29.82%	34.83%	39.84%	262
			AAH	-0.13%	5.08%	21.45%	45.83%	409.17%	31.69%	36.80%	41.91%	281
			MYH	-10.60%	1.66%	17.33%	40.05%	403.45%	26.95%	31.92%	36.88%	230
			SCH	-1.17%	0.67%	12.33%	32.68%	398.84%	22.73%	27.38%	32.03%	210

Table 2.8: Percentage improvement over dNFH in single-location terrorist attacks with incomplete information on the state of medical facilities.

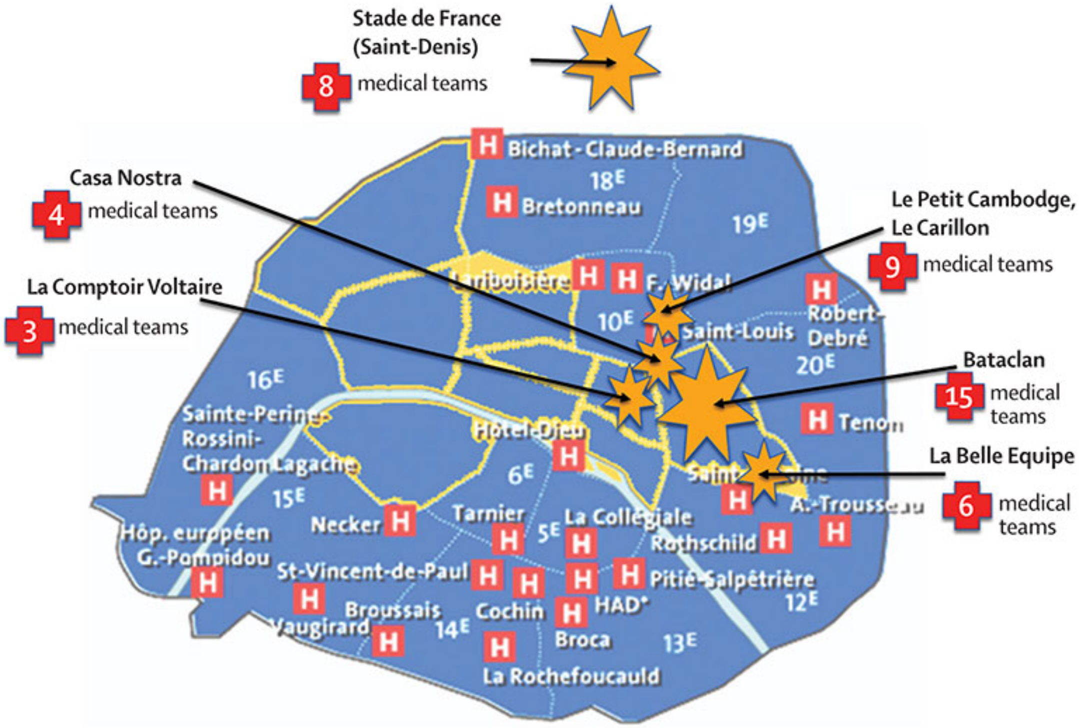

Figure 2.3: Casualty distribution in 2015 Paris attack

2.6.3.3 Single-location Major Traffic Accidents

Major traffic accidents, such as multiple pile-up accidents on the highway and train derailments, usually happen at a single location, for example the 2015 Philadelphia train derailment in Pennsylvania, U.S., and the accident happened on Interstate 95 at Stafford County, Virginia, U.S. in which over a hundred vehicles slammed into each other at high speeds led to more than one hundred injuries. Major traffic accidents may take place in between cities, therefore, the medical facilities may be further away than those MCIs that happen in cities. Most casualties during vehicle accidents suffer from blunt injuries. We categorized those events as single-location major traffic accidents. In our simulations, we study major traffic accidents where the casualty location and medical facilities are located within a larger area (25 by 25 miles^2 region) due to the possibility that these events may happen farther away from metropolitan area unlike terrorist attacks. The number of casualties is generated uniformly between 100 and 300 with the percentage of the immediate

No. Loc.	No. Amb.	No. Hsp.	Heuristic Policies	Min	Q1	Med	Q3	Max	Confidence Interval Lower	Mean	Upper	#sig.
3	10	3	PIH	-2.89%	8.03%	15.97%	28.11%	107.44%	18.90%	20.51%	22.12%	293
			AAH	-0.21%	9.47%	17.48%	29.70%	108.76%	20.14%	21.76%	23.38%	296
			MYH	-9.98%	3.80%	10.76%	21.57%	102.44%	13.71%	15.31%	16.90%	256
			SCH	-29.81%	-3.93%	1.83%	8.91%	71.64%	1.75%	3.10%	4.45%	163
3	30	3	PIH	-1.58%	1.57%	3.71%	9.01%	232.18%	6.97%	8.65%	10.32%	239
			AAH	-2.49%	2.16%	4.27%	9.82%	233.39%	7.83%	9.54%	11.26%	257
			MYH	-7.29%	-0.40%	1.23%	6.99%	226.10%	4.63%	6.27%	7.90%	154
			SCH	-12.45%	1.03%	2.98%	8.67%	218.87%	6.13%	7.73%	9.33%	223
3	50	3	PIH	-2.15%	0.44%	2.46%	8.09%	96.51%	5.32%	6.48%	7.65%	193
			AAH	-1.39%	0.90%	3.39%	8.94%	102.00%	6.31%	7.54%	8.76%	217
			MYH	-8.10%	-0.93%	1.18%	6.60%	90.42%	3.71%	4.84%	5.98%	148
			SCH	-2.59%	0.92%	3.00%	8.81%	101.87%	5.88%	7.06%	8.23%	221
4	10	3	PIH	-0.72%	9.04%	18.03%	31.94%	162.47%	23.05%	25.45%	27.86%	292
			AAH	-0.05%	9.75%	19.34%	33.55%	166.32%	24.17%	26.62%	29.07%	296
			MYH	-7.15%	4.79%	12.73%	25.59%	158.72%	17.71%	20.07%	22.43%	261
			SCH	-37.23%	-4.84%	3.05%	11.35%	78.19%	2.97%	4.54%	6.10%	173
4	30	3	PIH	-1.44%	1.56%	4.17%	8.71%	137.66%	6.10%	7.53%	8.97%	233
			AAH	-1.10%	2.03%	4.76%	9.40%	142.61%	6.77%	8.26%	9.75%	241
			MYH	-5.17%	-0.44%	2.35%	6.06%	135.22%	4.00%	5.43%	6.86%	176
			SCH	-7.76%	0.49%	2.28%	7.53%	141.82%	4.86%	6.37%	7.87%	191
4	50	3	PIH	-1.72%	0.70%	2.55%	6.73%	99.84%	5.79%	7.10%	8.41%	206
			AAH	-1.83%	1.45%	3.74%	7.56%	102.02%	6.76%	8.10%	9.45%	232
			MYH	-4.12%	-0.49%	1.44%	5.27%	93.43%	4.54%	5.83%	7.13%	152
			SCH	-3.35%	1.06%	2.83%	6.85%	100.75%	5.98%	7.31%	8.64%	221

Table 2.9: Percentage improvement over dNFH in multiple-location terrorist attacks.

casualties uniformly generated between 10% and 40%. We simulated the scenarios under the assumption of both perfect and imperfect information on the state of medical facilities. The results when the information on the state of medical facilities is available are presented in Tables 2.10 and 2.11 and the results when the information on the state of medical facilities is imperfect are presented in Tables 2.12 and 2.13.

All dynamic heuristics perform better than dNFH as expected. AAH performs the best in all scenarios in both settings (weather or not the information on the state of medical facilities is perfect) in terms of the median and the mean.

2.6.3.4 Earthquake Scenarios

Earthquakes usually affect a large area. In the aftermath of an earthquake, a huge number of blunt casualties spread over the entire region [30, 44]. In our simulation, we restricted ourselves to an urban area of size 50 by 50 miles^2. We generate a total of 100 to 200 casualty locations uniformly. At each casualty location, the number of casualties is generated from a uniform distribution between 6 and 15. We assume casualties have already been triaged into immediate and delayed classes at locations waiting for transportation. The percentage of immediate casualties is uniformly distributed between 10% and 40%. We only simulate under the assumption of incomplete information on the state of medical facilities for the earthquake scenario. The results are presented in Table 2.14.

Among the earthquake scenarios, SCH is doing the best in terms of the median and the mean. PIH, AAH and MYH perform even worse than dNFH in some scenarios in terms of the median (refer to rows 1 and 2 in Table 2.14). One possible reason for AAH not performing well is the number of casualty locations in earthquake scenarios is over a hundred, which dramatically violates the underlying assumption of a single casualty location for AAH.

No. (Pat.).	No. Amb.	No. Hsp.	Heuristic Policies	Min	Q1	Med	Q3	Max	Confidence Interval Lower	Mean	Upper	#sig.
100-300	30	2	PIH	-2.82%	9.91%	33.63%	74.03%	317.80%	42.97%	47.41%	51.86%	283
			MYH	-3.95%	8.93%	31.95%	71.82%	317.00%	41.58%	46.01%	50.43%	274
			AAH	-0.94%	10.98%	41.11%	82.61%	333.08%	47.85%	52.56%	57.28%	288
			SLH	-7.85%	6.88%	36.13%	76.27%	313.91%	42.37%	46.80%	51.23%	265
			SCH	-14.42%	-3.40%	1.32%	62.51%	173.57%	26.44%	30.45%	34.46%	151
100-300	50	2	PIH	-10.31%	3.07%	33.61%	79.48%	242.46%	44.36%	49.63%	54.91%	250
			MYH	-10.81%	2.59%	32.99%	78.10%	240.56%	43.54%	48.80%	54.06%	242
			AAH	-2.20%	4.06%	38.19%	86.03%	247.84%	48.26%	53.70%	59.13%	269
			SLH	-13.54%	0.50%	33.64%	80.19%	224.31%	42.38%	47.54%	52.70%	219
			SCH	-18.35%	-4.08%	28.07%	70.85%	207.41%	34.03%	38.84%	43.65%	160
100-300	30	3	PIH	-10.11%	19.81%	47.01%	107.98%	311.86%	63.63%	69.75%	75.87%	289
			MYH	-13.15%	16.09%	43.72%	103.14%	302.47%	60.28%	66.33%	72.39%	278
			AAH	-5.05%	29.18%	58.17%	120.52%	338.06%	75.59%	82.26%	88.92%	293
			SLH	-6.61%	24.72%	51.55%	116.38%	327.04%	69.35%	75.75%	82.15%	287
			SCH	-12.59%	12.99%	37.07%	105.34%	312.43%	56.25%	62.38%	68.51%	250
100-300	50	3	PIH	-3.62%	16.82%	46.47%	99.11%	392.82%	63.33%	70.00%	76.66%	283
			MYH	-6.68%	14.59%	42.99%	96.83%	388.04%	61.08%	67.74%	74.40%	277
			AAH	-4.34%	23.03%	55.68%	114.06%	404.68%	73.17%	80.19%	87.21%	287
			SLH	-12.83%	17.21%	48.22%	107.40%	376.56%	65.20%	71.81%	78.42%	275
			SCH	-13.19%	13.18%	38.83%	100.61%	350.73%	58.22%	64.63%	71.04%	256

Table 2.10: Percentage improvement over dNFH in single-location traffic accidents with perfect information on the state of medical facilities (Part 1).

No. (Pat.).	No. Amb.	No. Hsp.	Heuristic Policies	Min	Q1	Med	Q3	Max	Confidence Interval Lower	Mean	Upper	#sig.
100-300	30	4	PIH	-18.09%	14.97%	37.67%	83.64%	367.73%	58.49%	65.68%	72.87%	280
			MYH	-19.55%	10.08%	34.53%	78.11%	371.82%	53.91%	61.04%	68.16%	262
			AAH	-4.86%	25.22%	55.23%	108.60%	402.26%	76.26%	84.31%	92.36%	297
			SLH	-5.71%	21.95%	49.59%	101.47%	378.38%	70.46%	78.15%	85.84%	293
			SCH	-11.24%	12.07%	39.69%	94.01%	332.58%	60.28%	67.77%	75.26%	268
100-300	50	4	PIH	-12.34%	13.16%	43.20%	99.07%	312.65%	59.27%	65.80%	72.33%	279
			MYH	-13.81%	10.85%	39.24%	94.91%	310.13%	55.65%	62.13%	68.61%	264
			AAH	-7.11%	23.00%	59.83%	124.80%	346.91%	76.46%	83.88%	91.30%	291
			SLH	-8.74%	20.02%	54.41%	116.45%	334.69%	69.86%	76.98%	84.11%	282
			SCH	-13.03%	15.00%	47.09%	114.06%	332.76%	65.14%	72.24%	79.34%	276

Table 2.11: Percentage improvement over dNFH in single-location traffic accidents with perfect information on the state of medical facilities (Part 2).

No. (Pat.).	No. Amb.	No. Hsp.	Heuristic Policies	Min	Q1	Med	Q3	Max	Confidence Interval Lower	Mean	Upper	#sig.
100-300	30	2	PIH	-4.92%	9.49%	35.97%	83.65%	253.29%	46.52%	51.15%	55.78%	281
			AAH	-1.20%	10.03%	41.29%	87.85%	259.34%	49.38%	54.13%	58.89%	288
			MYH	-6.52%	5.75%	32.60%	75.46%	246.42%	41.66%	46.14%	50.61%	258
			SCH	-5.57%	0.22%	25.58%	60.60%	169.18%	31.51%	35.43%	39.36%	181
100-300	50	2	PIH	-1.95%	6.32%	49.89%	91.59%	214.42%	52.19%	57.21%	62.24%	268
			AAH	-1.43%	6.26%	53.48%	95.32%	216.15%	54.27%	59.41%	64.54%	270
			MYH	-4.35%	3.96%	46.58%	87.63%	207.33%	48.86%	53.83%	58.80%	252
			SCH	-2.15%	0.66%	42.60%	81.93%	188.25%	46.38%	51.32%	56.26%	206
100-300	30	3	PIH	-4.11%	17.06%	47.42%	109.39%	320.81%	64.87%	71.51%	78.16%	289
			AAH	-0.45%	23.49%	59.32%	120.19%	338.17%	73.65%	80.58%	87.50%	295
			MYH	-10.61%	8.20%	37.96%	97.44%	283.52%	54.37%	60.63%	66.89%	255
			SCH	-5.96%	9.26%	31.67%	103.97%	265.71%	53.68%	59.70%	65.72%	261
100-300	50	3	PIH	-9.57%	10.52%	31.80%	87.42%	390.03%	53.86%	60.64%	67.42%	277
			AAH	-2.68%	15.90%	37.83%	103.03%	403.20%	61.03%	68.02%	75.02%	288
			MYH	-11.17%	3.26%	25.39%	79.99%	376.95%	47.43%	54.03%	60.64%	240
			SCH	-1.12%	11.76%	33.47%	96.78%	406.06%	54.25%	60.79%	67.32%	274

Table 2.12: Percentage improvement over dNFH in single-location traffic accidents with incomplete information on the state of medical facilities (Part 1).

No. (Pat.).	No. Amb.	No. Hsp.	Heuristic Policies	Min	Q1	Med	Q3	Max	Confidence Interval Lower	Mean	Upper	#sig.
100-300	30	4	PIH	-7.80%	17.91%	41.01%	113.85%	434.67%	63.63%	70.77%	77.92%	281
			AAH	-5.67%	25.64%	52.19%	131.15%	456.25%	75.50%	83.12%	90.73%	293
			MYH	-17.15%	8.29%	31.18%	88.79%	423.31%	48.85%	55.44%	62.04%	254
			SCH	-18.78%	13.76%	37.13%	93.47%	332.34%	56.99%	63.35%	69.71%	292
100-300	50	4	PIH	-19.45%	10.40%	34.45%	97.30%	384.35%	63.66%	72.34%	81.01%	266
			AAH	-6.02%	14.81%	44.45%	121.21%	399.69%	75.45%	84.65%	93.86%	290
			MYH	-26.03%	1.75%	25.94%	85.61%	374.10%	54.07%	62.34%	70.60%	233
			SCH	-5.37%	14.26%	40.44%	105.77%	398.87%	69.26%	77.84%	86.42%	291

Table 2.13: Percentage improvement over dNFH in single-location traffic accidents with incomplete information on the state of medical facilities (Part 2).

2.7 Conclusion

In the aftermath of a mass casualty event, enormous numbers of casualties requiring immediate medical attention overwhelm the limited medical resources available. The prioritization and distribution of casualties to medical facilities is not only an intriguing operations research problem but also plays a vital role in practice. In this work, we formulated this problem as an MDP. Based on the insight derived from the MDP model, we proposed multiple heuristics that provide combined prioritization and distribution decisions. The heuristics could be useful to the emergency medical responders when making real-time decisions at the scene of a mass casualty incident or to create response plans.

In an extensive simulation study, we considered hypothetical MCIs abstracted from real events to examine the performance of proposed heuristics under a variety of conditions. In particular, we grouped MCIs into four main categories: single-location terrorist attacks, multiple-location terrorist attacks, single-location major traffic accidents, and earthquakes. From the simulation results, the dynamic heuristics, which take advantage of the available or estimated information on the state of medical facilities, such as MYH, PIH, AAH, and SCH perform reasonable well in comparison with the static heuristic, dNFH (which does not take into consideration the availability of medical facilities) except when the size of the event is very small relative to the availability of resources. This implies that the mass casualty events indeed requires different response plans from daily emergency incidents (which usually involve much fewer casualties than MCIs). In addition, we found that different heuristics derived based on various assumptions suit different types of MCIs. While AAH performs strongly at single-location terrorist attack scenarios, multiple-location terrorist attack scenarios, and single-location major traffic accident scenarios, SCH has non-negligible advantages in earthquake scenarios. Our simulation studies indicate that it is beneficial to categorize MCIs based on their characteristics such as scale and injury types, and utilize dif-

No. Amb.	No. Hsp.	Heuristic Policies	Quantile Min	Q1	Med	Q3	Max	Confidence Interval Lower	Mean	Upper	#sig.
30	3	PIH	-2.06%	0.25%	1.23%	3.92%	18.99%	2.03%	2.37%	2.70%	167
		AAH	-5.33%	-2.33%	-1.14%	1.99%	18.80%	-0.36%	0.01%	0.37%	83
		MYH	-3.88%	-1.40%	-0.29%	1.76%	17.85%	0.32%	0.63%	0.93%	96
		SCH	-1.24%	1.43%	2.42%	4.88%	23.59%	3.27%	3.62%	3.96%	255
40	3	PIH	-1.77%	0.21%	0.89%	2.55%	17.67%	1.65%	1.98%	2.30%	148
		AAH	-4.91%	-2.06%	-1.10%	0.47%	15.57%	-0.39%	-0.05%	0.29%	61
		MYH	-3.21%	-1.53%	-0.51%	0.85%	15.91%	0.02%	0.32%	0.63%	76
		SCH	-0.96%	0.77%	2.00%	3.54%	21.08%	2.70%	3.07%	3.44%	215
50	3	PIH	-2.22%	0.06%	1.81%	5.26%	36.86%	3.34%	3.94%	4.53%	191
		AAH	-4.38%	-1.72%	0.24%	3.45%	38.06%	1.76%	2.40%	3.04%	128
		MYH	-4.99%	-2.01%	0.35%	3.22%	32.69%	1.40%	1.99%	2.58%	128
		SCH	0.12%	2.35%	4.49%	7.97%	39.42%	5.95%	6.59%	7.22%	288
30	5	PIH	-0.85%	2.86%	4.04%	5.55%	15.72%	4.31%	4.56%	4.80%	297
		AAH	-4.06%	0.81%	2.88%	4.89%	13.26%	2.78%	3.09%	3.40%	230
		MYH	-1.89%	1.56%	2.53%	4.36%	11.48%	2.82%	3.03%	3.24%	271
		SCH	-1.53%	3.06%	4.85%	7.35%	17.53%	5.16%	5.48%	5.80%	289
40	5	PIH	-0.62%	2.58%	4.27%	6.53%	18.68%	4.65%	4.95%	5.26%	293
		AAH	-3.10%	2.05%	4.22%	6.81%	21.58%	4.45%	4.81%	5.18%	271
		MYH	-2.42%	1.02%	2.28%	4.42%	13.14%	2.52%	2.77%	3.01%	244
		SCH	0.78%	5.56%	7.38%	9.96%	26.08%	7.74%	8.12%	8.50%	300
50	5	PIH	-0.57%	2.71%	4.84%	7.47%	27.19%	5.16%	5.57%	5.98%	283
		AAH	-1.78%	2.01%	4.77%	8.06%	30.82%	5.14%	5.62%	6.10%	257
		MYH	-2.23%	0.58%	2.40%	4.56%	17.02%	2.72%	3.05%	3.37%	224
		SCH	1.73%	5.67%	8.47%	12.07%	36.25%	8.97%	9.48%	9.99%	300

Table 2.14: Comparison with dNFH in earthquakes with incomplete information on the state of medical facilities.

ferent heuristics for response. We recommend AAH when the number of casualty locations is small and SCH when the casualties are spread out in a larger area.

The heuristics proposed in this chapter are fairly easy to implement and could easily be deployed into an emergency management system to automate decision making. With the development of the internet of things and wearable medical equipment, more and more information will become accessible to the decision-maker. With more detailed states of casualties, ambulances and medical facilities on hand, the dynamic heuristics that take more information under consideration will have bigger advantages over simple static policies such as nearest hospital policy.

Although our simulation experiments are much more realistic and comprehensive than those in prior work, we acknowledge that they are still not up-to-the level we desire mainly due to lack of available data. Therefore, further investigation of the performance of the proposed heuristics is necessary. Future research avenues include but not limited to: the use of more realistic survival probability functions or even real-time vital signals to estimate survival probabilities empirically, consideration of different transportation methods such as helicopters that carry multiple casualties at a faster rate than ambulances, and incorporating errors in triage.

CHAPTER 3

URBAN SEARCH AND RESCUE RESOURCE ALLOCATION IN FLOODING

3.1 Introduction

The relationship between the climate change and the more frequent extreme events, such as heatwaves, extreme precipitation, and coastal flooding has already been observed in the Fifth Assessment Report of the Intergovernmental Panel on Climate Change from 2014 (IPCC) [22]. River flooding, in particular, affects more people globally than any other natural disaster and causes billions of dollars lost annually [14]. Storms and their direct implications, including heavy precipitation, floods, and landslides, were one major cause of damage in 2017 [20]. In May 2017, heavy landslides and floods in Sri Lanka caused more than 200 deaths. In Nepal, Bangladesh, and India, massive rainfalls and the consequential floods affected more than 40 million people including 1,200 deaths also in 2017 [25, 66]. In North America and Europe, the frequency and intensity of precipitation events has increased in recent years [67].

U.S. is one of the top ten countries with the highest climate risk index score hit that were by tropical cyclones in 2017 (see Table 3.1). Take Hurricane Harvey, the costliest tropical cyclone on record, that hit Texas in 2017 as an example. Houston metropolitan area observed at least 30 inches of precipitation [71] and a maximum of 60.58 inches in Nederland [62]. This makes Harvey the wettest tropical cyclone on record for both Texas and the United States [7]. An estimated 25 to 30 percent of Harris County, roughly 444 mi^2 of land, and home to 4.5 million people in Houston and its suburbs, was submerged [39]. Approximately 10 percent of Texas was flooded [72]. During Hurricane Harvey, devastating winds and catastrophic flooding necessitated 21,433 searches and rescue personnel from the

Ranking CRI	Country	CRI score	Death toll	Deaths per 100 000 inhabitants	Absolute losses in US$ million (PPP)	Losses per unit GDP in %
1	Puerto Rico	1.50	2 978	90.242	82 315.24	63.33
3	Dominica	9.33	31	43.662	1 686.89	215.44
6	Vietnam	13.50	298	0.318	4 052.31	0.62
7	Madagascar	15.00	89	0.347	693.04	1.74
12	United States	19.83	389	0.119	177 981.95	0.91
13	Antigua and Barbuda	20.67	3	3.297	1 101.44	45.93
20	Philippines	33.17	250	0.238	505.78	0.06
21	Costa Rica	33.83	11	0.221	273.68	0.33
25	Nicaragua	36.00	23	0.370	52.70	0.14
27	Haiti	37.33	18	0.164	88.87	0.44

PPP = Purchasing Power Parities. GDP = Gross Domestic Product. Note: this table includes impacts from all disasters, not only tropical cyclones. However, tropical cyclones are responsible for a significant share of the disasters due to major storm events.

Table 3.1: Ten countries with the highest CRI score, which were hit by tropical cyclones in 2017 [20].

nation to evacuate 35,046 people, rescue 12,982 people and 2,055 animals [72]. However, 103 people still died in storm-related incidents including 68 due to direct effects of the storm including flooding [12].

The urban search and rescue operations are of vital importance in saving people's lives during extreme weather events like hurricanes. However managing these operations is difficult due to hazardous weather conditions, large numbers of rescue requests, and often limited resources. To the best of our knowledge, no standard guideline exists for coordinating the urban search and rescue operations at least in the U.S. Based on [3, 72], the urban search and rescue operations decisions are made in an “ad hoc” fashion based on experience from previous events. The Harris County Fire Marshals Office included the follows in their report [3]: 1. The establishment of triage protocols for determining the priority for rescue by dispatchers has been identified as “best Practice”. Multiple departments found that triage protocols helped them in managing limited resources even though there are no unified standards. 2. A better coordination and management between public safety agencies including but not limited to fire agencies, emergency medical services, task forces, and other enforcement agencies. In addition, Alexander [2] enumerated eighteen principles expert on emergency planning and identified the optimal allocation of urgent needs with appropriate resources as one of the important criteria. These point to a need for systematic planning of urban search and rescue operations.

We considered particularly the urban search and rescue operations that follow flooding such as the one that happened in Texas following Hurricane Harvey. We aim at creating a decision support tool that facilitates the dispatch of helicopters for rescuing people. Specifically, a large number of rescue requests occurred across the entire Harris county in a short period when the flood took place. The number of people and animals need to be rescued differ from request to request. The flood developed at a different rate depends on the terrain. Thus the urgent level varies from place to place. The required times to research, rescue, and transport people requesting rescue also vary from location to location. The emergency response coordinators need to dispatch helicopters to request in an optimal order such that most people could be rescued in time. Rescued people will be sent to major shelters capable of landing helicopters. Our heuristics prescribe the assignment of helicopters to request to maximize the number of accomplished requests before their deadlines.

The proposed analytical model is a generic scheduling model which could be also of interest to the scheduling community. Our model has the following key features: 1. We allow class-dependent deadlines which will eliminate all jobs in the corresponding class once the deadline is reached. 2. Both class-dependent rewards and holding costs are considered in our model. 3. We provide a full characteristic of the optimal policy for the two-class model. 4. We proposed easy-to-implement and near-optimal heuristic policies. The heuristics are examined in realistic simulations based on hurricane Harvey.

The remaining of this chapter is organized as follows. We review operations research and urban search and rescue literature relevant to this work in Section 3.2. We characterized the optimal policy for the two-class prioritization problem with one server in Section 3.3.

Heuristics for the general problem involving N classes and M servers were proposed in Section 3.4. Numerical studies that compared heuristics to the optimal policy were presented in Section 3.5. A simulation case study based on the air rescue operation during the aftermath of Hurricane Harvey was presented in Section 3.6. We conclude the chapter by a discussion of the main insights gained from this study in Section 3.7.

3.2 Literature Review

The problem of interest in this article – the allocation of insufficient resources or service capacities to overwhelming requests has long been studied in the literature. This problem has been referred to by various names, such as prioritization of impatient jobs, resource allocation, scheduling with due dates. This problem has also been studied using various models under different modeling assumptions. However, there is not much research on applying operations research techniques on resource allocation in Urban Search and Rescue during flooding events. Therefore, we only reviews papers that are tangentially relevant in the remainder of this section together.

Pinedo [56] considered three single machine stochastic scheduling models where the processing times are independent exponentially distributed, the release dates have arbitrary joint distributions. He showed the optimality of the $c\mu$ (cost times service rate) type rule among all dynamic policies for minimizing the expected weighted sum of job completion times, the expected weighted sum of job tardiness, and the expected weighted number of late jobs, when the random due dates satisfying certain conditions on their joint distributions.

Ross [61] considered the scheduling of n jobs with distinct exponential deadlines and general service times on a single server. Sufficient conditions were provided under which the list policy is optimal. Cao [15] provided sufficient conditions for a list policy to be optimal for both preemptive and non-preemptive multiple servers scheduling problem with exponential service rate and deadlines.

Glazebrook et al. [26], Li and Glazebrook [42] and James et al. [33] focused on finding close to optimal solutions. Glazebrook et al. [26] investigated three schedule models with impatient jobs. The authors proved that the deviation of $dc\mu$ (the product of expiration rate, cost, and service rate) rule from the optimal permutation policy is bounded by $O(\theta^2)$ (θ is the common expiration rate) in a single non-preemptive server system with exponential service time. The abandon probability for each job possesses the memoryless property. The authors generalized the Gittins index policy for the preemptive server case. Policy improvement policy is adopted to obtain an index policy for the multiple-class queuing model with independent Poisson arrival, preemptive service with exponential service times, and exponentially distributed deadlines. Li and Glazebrook [42] applied a policy improvement step on the approximated $dc\mu$ heuristic proposed in [26] to obtain a near-optimal heuristic for the multiple class clearing system with impatient jobs. James et al. [33] further showed that the $dc\mu$ rule proposed in [26] is asymptotically optimal when the expiration rate approaches zero. An approximate policy improvement method based on using bias functions from simulation to approximate the value functions in the dynamic programming recursion is proposed and tested in a five classes system.

Ayesta et al. [8] establish a model considering linear holding costs, job completion rewards, and abandonment penalties. The optimal solutions were identified for the cases where one completing with an alternative task or two jobs completing with each other. Heuristic policy based on Whittle's relaxation was proposed for multiple jobs cases.

Similar clearing models have been applied to improving medical system. Argon et al. [4] and Jacobson et al. [32] modeled the casualty prioritization problem in the aftermath of mass casualty incidents as a scheduling problem for a clearing system. The objective is to maximize the expected number of survivors. Casualties belong to two types may have different service times. One casualty reneges every time the deadline of the respective type is reached. Stochastic comparisons were used to identify conditions under which state-independent policies are optimal and optimal policies were partially characterized when they depend on the system state through an MDP formulation. [32] generalized [4] with type-dependent reward and developed heuristics which allowed multiple servers. Sun et al. [68] also used a single server clearing model to study the triage and prioritization problem under austere conditions. In their model, all patients are of unknown status at the beginning and the service provider has the option to treat a patient directly without triage or triage a patient to figure out his/her urgent level. The authors proved the optimal dynamic policy can be expressed by a switching curve.

Clearing models have also been used widely in designing evacuation plans. Childers et al. [17] modeled patients in health-care facility evacuation as impatient jobs in a clearing system. The authors suggest a threshold policy begin with noncritical patients and then switch to critical patients based on a simulation study. In [16], Childers et al. declared that "all-or-nothing" policy, where one prioritized group of patients are emptied before another group of patients starts the evacuation, is not always optimal through solving a Markov decision process model numerically. The authors also provided a $c\mu$ type threshold for determining the priority among critical and noncritical patients.

Many other methods have been used to study the allocation of limited allocation with different domains of applications. Kamali et al. [38] looked at a similar prioritization problem in the context of mass casualty incidents. An integer program is proposed to find the optimal service order for patients triaged into multiple classes.

3.3 Two-Class Prioritization Problem

3.3 Model Assumptions and the MDP Formulation

We now consider a model involving two heterogeneous classes of jobs. x jobs in class 1 and y jobs in class 2 are waiting for service at time zero with future arrivals for neither class. A common deadline for class $k \in \{1, 2\}$ jobs follows an exponential random distribution with a positive rate of D_k. All jobs, including those in service, who have not completed their service before their respective deadline will leave the system. A single server serves one job at a time non-preemptively. The service time for each individual job follows an exponential distribution with a positive rate of μ_k for $k \in \{1, 2\}$. A reward of r_k is generated at the service completion for each job in class $k \in \{1, 2\}$ after which the job exits the system. The non-negative holding cost is c_k per unit of time for each class $k \in \{1, 2\}$ job in the system. The goal is to find the optimal schedule for the processing of jobs such that the total expected reward is maximized. We refer to this problem as the prioritization problem for two classes.

We formulate the problem as a Markov Decision Process. The state (x, y) consists of the number of jobs in each class. The prioritization decisions are made at service completions. Let $\mathcal{A} = \{1, 2\}$ denote the set of actions where $a \in \mathcal{A}$ denotes the action of serving a job in class a next. Let $V(x, y)$ denote the maximum expected reward starting from state (x, y). The optimality equation can be written as

$$\begin{aligned}
&V(x,y) = \max_{a\in\mathcal{A}} \left\{V(a;x,y)\right\}, \text{ for } x \geq 1, y \geq 1, \text{ where}\\
&V(1;x,y) = \frac{\mu_1\left[r_1 + V(x-1,y)\right] + D_1V(0,y) + D_2V(x,0) - c_1x - c_2y}{\mu_1 + D_1 + D_2}, \text{for } x \geq 1, y \geq 0;\\
&V(2;x,y) = \frac{\mu_2\left[r_2 + V(x,y-1)\right] + D_1V(0,y) + D_2V(x,0) - c_1x - c_2y}{\mu_2 + D_1 + D_2}, \text{for } x \geq 0, y \geq 1;\\
&V(x,0) = V(1;x,0), \text{ for } x \geq 1; V(0,y) = V(2;0,y), \text{ for } y \geq 1; \text{ and } V(0,0) = 0.
\end{aligned} \tag{3.1}$$

3.3 Classification of the Optimal Policy

In this section, we show that for a fixed number of jobs in one class, the optimal policy for the prioritization problem is a threshold policy on the number of jobs in the other class. We need the following Lemmas to prove the main results.

Lemma 2. When $x = 0, y \geq 1$ or $x \geq 1, x = 0$, there is a single class of jobs in the system. We can rewrite the recurrence using a single state variable and one set of parameters as follows,

$$V(x) = \frac{\mu[r + V(x-1)] - xc}{\mu + D}, \text{ and } V(0) = 0.$$

Solving the recurrence explicitly gives us the following result,

$$V(n) = \left[1 - \left(\frac{\mu}{\mu + D}\right)^x\right] \left(\frac{\mu r}{D} + \frac{c\mu}{D^2}\right) - xcD^{-1}, \text{ for } x \geq 1. \tag{3.2}$$

Lemma 3. Consider a state (x, y) such that $x \geq 1$ and $y \geq 1$,

(a) If $a = 2$ is optimal at state $(x-1, y)$ and the following equality holds

$$\begin{aligned} &\mu_1 \left(r_1 + c_1 D_1^{-1}\right) \left[D_2 \left(\frac{\mu_1}{\mu_1 + D_1}\right)^x - (D_1 + D_2)\right] \\ \geq &\mu_2 \left(r_2 + c_2 D_2^{-1}\right) \left[D_1 \left(\frac{\mu_2}{\mu_2 + D_2}\right)^y - (D_1 + D_2)\right], \end{aligned} \tag{3.3}$$

then $a = 2$ is the optimal action at state (x, y). We say the action $a = 2$ is strictly optimal (strictly better than $a = 1$) if and only if condition (3.3) holds as strictly inequality.

(b) If $a = 1$ is optimal at state $(x, y-1)$ and the following equality holds

$$\begin{aligned} &\mu_1 \left(r_1 + c_1 D_1^{-1}\right) \left[D_2 \left(\frac{\mu_1}{\mu_1 + D_1}\right)^x - (D_1 + D_2)\right] \\ \leq &\mu_2 \left(r_2 + c_2 D_2^{-1}\right) \left[D_1 \left(\frac{\mu_2}{\mu_2 + D_2}\right)^y - (D_1 + D_2)\right] \end{aligned} \tag{3.4}$$

then $a = 1$ is the optimal action at state (x, y). We say the action $a = 1$ is strictly optimal (strictly better than $a = 2$) if and only if condition (3.3) holds as a strictly inequality.

Proof. (a) The expected reward for taking action $a = 1$ at state (x, y) can be express as,

$$V(1; x, y) = \frac{\mu_1 \left[r_1 + V(x-1, y)\right] + D_1 V(0, y) + D_2 V(x, 0) - c_1 x - c_2 y}{\mu_1 + D_1 + D_2} \tag{3.5}$$

By our assumption that $a = 2$ is optimal at $(x-1, y)$, we have

$$\begin{aligned} &V(x-1,y) = V(2;x-1,y) \\ &= \frac{\mu_2\left[r_2 + V(x-1,y-1)\right] + D_1V(0,y) + D_2V(x-1,0) - c_1(x-1) - c_2y}{\mu_2 + D_1 + D_2}. \end{aligned} \tag{3.6}$$

Substitute 3.6 into 3.5 we get,

$$\begin{aligned} V(1;x,y) = &\frac{\mu_1 r_1 + D_1V(0,y) + D_2V(x,0) - c_1x - c_2y}{\mu_1 + D_1 + D_2} + \frac{\mu_1}{\mu_1 + D_1 + D_2}\cdot \\ &\frac{\mu_2\left[r_2 + V(x-1,y-1)\right] + D_1V(0,y) + D_2V(x-1,0) - c_1(x-1) - c_2y}{\mu_2 + D_1 + D_2}. \end{aligned} \tag{3.7}$$

On the other hand,

$$V(2;x,y) = \frac{\mu_2\left[r_2 + V(x,y-1)\right] + D_1V(0,y) + D_2V(x,0) - c_1x - c_2y}{\mu_2 + D_1 + D_2}, \tag{3.8}$$

and,

$$\begin{aligned} &V(x,y-1) \geq V(1;x,y-1) \\ &= \frac{\mu_1\left[r_1 + V(x-1,y-1)\right] + D_1V(0,y-1) + D_2V(x,0) - c_1x - c_2(y-1)}{\mu_2 + D_1 + D_2}, \end{aligned} \tag{3.9}$$

The equality in (3.9) holds when $a = 1$ is optimal at state $(x, y-1)$. Substitute (3.9) into (3.8), we have

$$\begin{aligned} V(2;x,y) \geq \hat{V} = &\frac{\mu_2 r_2 + D_1V(0,y) + D_2V(x,0) - c_1x - c_2y}{\mu_2 + D_1 + D_2} + \frac{\mu_2}{\mu_2 + D_1 + D_2}\cdot \\ &\frac{\mu_1\left[r_1 + V(x-1,y-1)\right] + D_1V(0,y-1) + D_2V(x,0) - c_1x - c_2(y-1)}{\mu_1 + D_1 + D_2} \end{aligned} \tag{3.10}$$

If $\hat{V} \geq V(1;x,y)$, then $V(2;x,y) \geq \hat{V} \geq V(1;x,y)$ which provides an sufficient condition for $a = 2$ being the optimal action at state (x, y). Further, the condition is an if and only if condition if $a = 1$ is optimal at state $(x, y-1)$.

$$
\begin{aligned}
&\hat{V} - V(1; x, y) \\
=&\frac{\mu_2 r_2 + D_1 V(0,y) + D_2 V(x,0) - c_1 x - c_2 y}{\mu_2 + D_1 + D_2} + \frac{\mu_2}{\mu_2 + D_1 + D_2} \cdot \\
&\frac{\mu_1 \left[r_1 + V(x-1, y-1)\right] + D_1 V(0, y-1) + D_2 V(x,0) - c_1 x - c_2(y-1)}{\mu_1 + D_1 + D_2} \\
-&\frac{\mu_1 r_1 + D_1 V(0,y) + D_2 V(x,0) - c_1 x - c_2 y}{\mu_1 + D_1 + D_2} - \frac{\mu_1}{\mu_1 + D_1 + D_2} \cdot \\
&\frac{\mu_2 \left[r_2 + V(x-1, y-1)\right] + D_1 V(0, y) + D_2 V(x-1,0) - c_1 (x-1) - c_2 y}{\mu_2 + D_1 + D_2} \\
=&\frac{(D_1 + D_2)\mu_2 r_2 + \mu_2 c_2 - \mu_2 D_1 \left[V(0,y) - V(0,y-1)\right]}{\left[\mu_1 + D_1 + D_2\right]\left[\mu_2 + D_1 + D_2\right]} \\
-&\frac{(D_1 + D_2)\mu_1 r_1 + \mu_1 c_1 - \mu_1 D_2 \left[V(x,0) - V(x-1,0)\right]}{\left[\mu_1 + D_1 + D_2\right]\left[\mu_2 + D_1 + D_2\right]} \\
=&\frac{1}{\left[\mu_1 + D_1 + D_2\right]\left[\mu_2 + D_1 + D_2\right]} \cdot \\
&\left\{ \mu_1 \left[r_1 + c_1 D_1^{-1}\right] \left[D_2 \left(\frac{\mu_1}{\mu_1 + D_1} \right)^x - (D_1 + D_2) \right] \right. \\
&\left. - \mu_2 \left[r_2 + c_2 D_2^{-1}\right] \left[D_1 \left(\frac{\mu_2}{\mu_2 + D_2} \right)^y - (D_1 + D_2) \right] \right\},
\end{aligned}
\tag{3.11}
$$

where the marginal rewards when a single class of casualties exist can be obtained using Lemma 2, as the follows,

$$
\begin{aligned}
V(x,0) - V(x-1,0) &= \left(r_1 + c_1 D_1^{-1}\right) \left(\frac{\mu_1}{\mu_1 + D_1} \right)^x - c_1 D_1^{-1}, x \geq 1, \\
V(0,y) - V(0,y-1) &= \left(r_2 + c_2 D_2^{-1}\right) \left(\frac{\mu_2}{\mu_2 + D_2} \right)^y - c_2 D_2^{-1}, y \geq 1,
\end{aligned}
$$

Since μ_1, μ_2, D_1, and D_2 are all positive, the denominator in equation (3.11) is positive. Therefore, $\hat{V} - V(1; x, y) \geq 0$ if and only if the numerator in equation (3.11) is non-negative (which is equivalent as the condition in the lemma). $\hat{V} - V(1; x, y) \geq 0$, implies $V(2; x, y) \geq V(1; x, y)$ which implies $a = 2$ is optimal at state (x, y). The condition is an if and only if condition if $a = 1$ is optimal at state $(x, y-1)$ in which case $V(2; x, y) = \hat{V}$.

(b) We can prove the correctness with an argument similar to the proof of part (a), and hence, is omitted.

□

We now use Lemma 3 to show that the optimal policy is an index policy for an arbitrary state (x, y), where $x \geq 1$ and $y \geq 1$.

Proposition 4. For any state (x, y) with $x \geq 1$ and $y \geq 1$, the optimal action $a^*(x, y) = 1$ if

$$\begin{aligned} &\mu_1 \left(r_1 + c_1 D_1^{-1}\right) \left[(D_1 + D_2) - D_2 \left(\frac{\mu_1}{\mu_1 + D_1}\right)^x\right] \\ \geq &\mu_2 \left(r_2 + c_2 D_2^{-1}\right) \left[(D_1 + D_2) - D_1 \left(\frac{\mu_2}{\mu_2 + D_2}\right)^y\right]; \end{aligned} \tag{3.12}$$

Otherwise, $a^*(x, y) = 2$.

Proof. Suppose inequality (3.3) holds at (x, y) for $x \geq 1$ and $y \geq 1$, since $0 < \mu_1 < \mu_1 + D_1$, inequality (3.3) holds for all states $(\hat{x}, y)$ for $1 \leq \hat{x} \leq x$. By default, $a = 2$ is optimal for state $(0, y)$ with $y \geq 1$. Then $a = 2$ is optimal for state $(1, y)$ for $y \geq 1$ by Lemma 3. By induction on x, based on Lemma 3, we can show that $a = 2$ is optimal at state (x, y).

On the other hand if inequality (3.3) does not hold at (x, y) for $x \geq 1$ and $y \geq 1$, inequality (3.4) must holds at (x, y) for $x \geq 1$ and $y \geq 1$. Since $0 < \mu_2 < \mu_2 + D_2$, inequality (3.4) holds for all states $(x, \hat{y})$ for $1 \leq \hat{y} \leq y$. By default, $a = 1$ is optimal for state $(x, 0)$ with $x \geq 1$. Then by an induction based on Lemma 3, we have $a = 1$ is optimal at state (x, y). □

Proposition 5. For a fixed $x \geq 0$, there exists a threshold $y^* \geq 0$ such that action $a = 1$ is optimal at state (x, y) for all $0 \leq y < y^*$, and $a = 2$ is the optimal action at state (x, y) for all $y \geq y^*$, where

$$y^* = \left\lceil \frac{\ln\left(\frac{\mu_1\left(r_1+c_1D_1^{-1}\right)D_2\left(\frac{\mu_1}{\mu_1+D_1}\right)^x+(D_1+D_2)\left\{\mu_2\left(r_2+c_2D_2^{-1}\right)-\mu_1\left(r_1+c_1D_1^{-1}\right)\right\}}{\mu_2\left(r_2+c_2D_2^{-1}\right)D_1}\right)}{\ln\left(\frac{\mu_2}{\mu_2+D_2}\right)} \right\rceil . \tag{3.13}$$

Proof. When $x = 0$, by default, the optimal action is $a = 2$ at state $(0, y)$ for all $y \geq 1$, i.e., $y^* = 1$. For state $(x, 0)$ with $x \geq 1$, $a = 1$ is the optimal action also by default. Then for an arbitrary state (x, y) with $x \geq 1$ and $y \geq 1$, let y^* be the smallest positive integer such that the inequality (3.3) holds, by proposition 4, $a = 2$ is optimal at state (x, y^*). Since $0 \leq \mu_2 \leq \mu_2 + D_2$, $\mu_2 > 0$, $r_2 \geq 0$, $c_2 \geq 0$, and $D_2 > 0$, inequality (3.3) holds for all $y \leq y^*$. Therefore, $a = 2$ is the optimal action for all states (x, y) with $y \geq y^*$.

On the other hand, since by our construction, y^* is the smallest positive integer such that the inequality (3.3) holds, the inequality (3.4) holds for all $1 \leq y < y^*$. By proposition 4, $a = 1$ is optimal at states (x, y) for $1 \leq y < y^*$. And equation (3.13) is a reformulation of inequality (3.3) in Lemma 3. □

3.3 Asymptotic Results for Deadlines

We express the class dependent deadline rate as a multiple of a common rate D as $D_1 = d_1 D$ and $D_2 = d_2 D$. We then study the optimality condition given in (3.12) when the deadline rate D goes to zero (deadline goes to infinity). From Proposition 4, we know $a = 1$ is optimal if and only if,

$$\frac{\mu_1\left(r_1 + c_1 d_1^{-1} D^{-1}\right)\left[(d_1 D + d_2 D) - d_2 D\left(\frac{\mu_1}{\mu_1+d_1 D}\right)^x\right]}{\mu_2\left(r_2 + c_2 d_2^{-1} D^{-1}\right)\left[(d_1 D + d_2 D) - d_1 D\left(\frac{\mu_2}{\mu_2+d_2 D}\right)^y\right]} \geq 1.$$

Now consider the fraction as D approaches 0 for $c_1, c_2 > 0$:

$$\begin{aligned}
&\lim_{D\to 0} \frac{\mu_1\left(r_1 + c_1 d_1^{-1} D^{-1}\right)\left[(d_1 D + d_2 D) - d_2 D\left(\frac{\mu_1}{\mu_1 + d_1 D}\right)^x\right]}{\mu_2\left(r_2 + c_2 d_2^{-1} D^{-1}\right)\left[(d_1 D + d_2 D) - d_1 D\left(\frac{\mu_2}{\mu_2 + d_2 D}\right)^y\right]} \\
=&\lim_{D\to 0} \frac{\mu_1\left(r_1 D + c_1 d_1^{-1}\right)\left[(d_1 + d_2) - d_2\left(\frac{\mu_1}{\mu_1 + d_1 D}\right)^x\right]}{\mu_2\left(r_2 D + c_2 d_2^{-1}\right)\left[(d_1 + d_2) - d_1\left(\frac{\mu_2}{\mu_2 + d_2 D}\right)^y\right]} \\
=&\frac{c_1\mu_1}{c_2\mu_2}.
\end{aligned} \tag{3.14}$$

The result in (3.14) implies our dynamic policy agrees asymptotically with $c\mu$ rule as the deadlines go to infinity. The rewards no longer make a difference in this case since for sufficiently large deadlines, all jobs will eventually be completed.

In many applications, the holding costs are ambiguous and difficult to quantify. The deadlines to some extend also indicate the urgency. Therefore, we are interested in models where the holding costs are omitted. If we set $c_1 = c_2 = 0$ before letting D approaches 0, we have

$$\begin{aligned}
&\lim_{D\to 0} \frac{\mu_1 r_1\left[(d_1 D + d_2 D) - d_2 D\left(\frac{\mu_1}{\mu_1 + d_1 D}\right)^x\right]}{\mu_2 r_2\left[(d_1 D + d_2 D) - d_1 D\left(\frac{\mu_2}{\mu_2 + d_2 D}\right)^y\right]} \\
=&\lim_{D\to 0} \frac{\mu_1 r_1\left[(d_1 + d_2) - d_2\left(\frac{\mu_1}{\mu_1 + d_1 D}\right)^x\right]}{\mu_2 r_2\left[(d_1 + d_2) - d_1\left(\frac{\mu_2}{\mu_2 + d_2 D}\right)^y\right]} \\
=&\frac{\mu_1 r_1 d_1}{\mu_2 r_2 d_2}.
\end{aligned} \tag{3.15}$$

In this case, the result is similar to the conclusion in [26] which is also obtained by assuming a small common deadline rate.

3.4 Heuristics for Multi-Class and Multi-Server Problem

In a more generic model involving N heterogeneous classes of jobs severed by M identical jobs finding an explicit solution analytically becomes impossible. Due to the curse of dimensionality, the numerical method only works for small scale problems. Therefore, we designed heuristic policies which provide good solutions to the complex general model in this section based on the solution structure of the analytical solution we derived for two-class single-server case in Section 3.3.

Let $\mathbb{C}$ denotes the set of all classes and $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$ is a vector representing the number of jobs in each class waiting for service at time zero. There will be no future arrivals for any class. A common deadline exists for all jobs in a given class $k \in \mathbb{C}$ which follows an exponential random distribution with a positive rate of $D_k, k \in \mathbb{C}$. All jobs, including those in service, who have not completed their service before their respective deadline will expire (leave the system). There is a single machine serves one job at a time non-preemptively. The service time for each individual job follows an exponential distribution with a positive rate of μ_k for $k \in \mathbb{C}$. The machine generates a reward of r_k at the service completion for each job in class k for $k \in \mathbb{C}$ after that the job exits the system. The non-negative holding cost is c_k per unit of time for each class k job in the system for $k \in \mathbb{C}$. We aim to schedule the jobs dynamically for service so as to maximize the total expected reward until the system is empty.

We formulate the system as a Markov Decision Process. The actions corresponding to which class will be in service next are made every time the service becomes idle and the system is not empty. Let $\mathcal{A} = \{1, 2, \ldots\ldots, N\}$ denote the actions space while $a \in \mathcal{A}$ denotes the action of serving a job in class a. In state $\mathbf{x}$, an action $a \in \mathcal{A}$ is feasible if and only if $x_a > 0$. Let $\mathcal{A}_{\mathbf{x}}$ denotes the set of feasible actions in state $\mathbf{x}$. Let $V(\mathbf{x})$ denote the maximum expected reward starting from state $\mathbf{x}$. The optimality equation can be expressed as,

$$V(\mathbf{x}) = \max_{a \in \mathcal{A}_{\mathbf{x}}} \left\{ \frac{\mu_a \left[r_a + V(\mathbf{x} - e_a) \right] + \sum_i D_i V(\mathbf{x} - x_i e_i) - \sum_i c_i x_i}{\mu_a + \sum_i D_i} \right\}, \text{ and } V(\mathbf{0}) = 0. \quad (3.16)$$

Consider two feasible actions p and q at state $\mathbf{x}$, following a similar calculation as in the proof for Lemma 3, we have the following condition for action p more preferable than action q,

$$\begin{aligned} &\mu_p \left\{ c_p + r_p \sum_i D_i - \sum_{i \neq p} D_i \left[V(\mathbf{x} - x_i e_i) - V(\mathbf{x} - x_i e_i - e_p) \right] \right\} \\ \geq &\mu_q \left\{ c_q + r_q \sum_i D_i - \sum_{i \neq q} D_i \left[V(\mathbf{x} - x_i e_i) - V(\mathbf{x} - x_i e_i - e_q) \right] \right\}. \end{aligned} \tag{3.17}$$

Notice the computation of the marginal values $V(\mathbf{x} - x_i e_i) - V(\mathbf{x} - x_i e_i - e_p), \forall i \neq p$ or $i \neq q$ are performed at a state space one dimension lower than the original state $V(\mathbf{x})$ as in each computation at least one class will vanish. Although it is possible to follow the path as in Section 3.3 to obtain the optimal policy for the general N classes case, due to the curse of dimensionality, find a closed form result is impossible. Even numerical computation will quickly become infeasible as the number of classes increases. Therefore, we focus on constructing easy-to-implement heuristics based on the analytical results we established in Section 3.3 and equation (3.17). We also incorporate multiple servers when building our heuristics.

3.4 Dynamic Heuristics

3.4.1.1 Two-Class Approximation Heuristic (TAH)

We approximate the difference of V values in equation (3.17) assuming there are only two classes. In that case, the difference in V values only involves a single class of jobs. Therefore, we can use (3.2) to approximate the V values explicitly as,

$$\begin{aligned} &V(\mathbf{x} - x_i e_i) - V(\mathbf{x} - x_i e_i - e_p) \\ \approx &V(e_p x_p) - V(e_p(x_p - 1)) \\ = &(r_p + c_p D_p^{-1}) \left(\frac{\mu_p}{\mu_p + D_p} \right)^{x_p} - c_p D_p^{-1}. \end{aligned} \tag{3.18}$$

The next available server will be assigned to a job in class a^* as follows,

$$a^*(\mathbf{x}) = \underset{p\in\mathcal{A}}{\text{argmax}} \left\{ \mu_p \left(r_p + c_p D_p^{-1}\right) \left[\sum_i D_i - \sum_{i\neq p} D_i \left(\frac{\mu_p}{\mu_p + D_p} \right)^{x_p} \right] \right\}. \tag{3.19}$$

3.4.1.2 Bi-Class Approximation Heuristics (BAH-mean, BAH-max)

We adopt a similar idea as in TAH by considering one class at a time while treating all other classes as another class in order to approximate the V values in equation (3.17). For every class of p, we construct another class q by averaging all other class except p. The class q will have a reward of $\bar{r}_q = \sum_{i\neq p} r_i/(N-1)$, holding cost of $\bar{r}_q = \sum_{i\neq p} c_i/(N-1)$, service rate of $\bar{\mu}_q = \left(\sum_{i\neq p} \mu_i^{-1}/(N-1)\right)^{-1}$ (Harmonic mean), and deadline rate of $\bar{D}_q = \left(\sum_{i\neq p} D_i^{-1}/(N-1)\right)^{-1}$. Then, we can use the results in the two-class approximation heuristic to identify the next class to serve when a server becomes available next.

$$a^*(\mathbf{x}) = \underset{p\in\mathcal{A}}{\text{argmax}} \left\{ \mu_p \left(r_p + c_p D_p^{-1}\right) \left[(D_p + \bar{D}_q) - \bar{D}_q \left(\frac{\mu_p}{\mu_p + D_p} \right)^{x_p} \right] \right\}. \tag{3.20}$$

Note that the deadline for class q appears in the formula.

We can also use the maximum of the deadlines rate instead of the average among all other classes to create another variation of the heuristic. In that case, for every class of p, we construct another class q with a a deadline rate of $\bar{D}_q = \max_{i\neq p} D_i$. Then, we can obtain the class to server by replacing the mean deadline rate $\bar{D}_q$ in formula (3.20) by the maximum deadlines rate $\bar{D}_q$. BAH-mean and BAH-max perform similarly in the simulation. Therefore, we only include the results of BAH-mean in numerical studies and simulations.

3.4 Static Heuristics

3.4.2.1 Dedicate Assignment Heuristic (DAH)

The dedicate assignment heuristic evenly distributes all M identical servers among N classes. Specifically, an available server will be assigned to serve class i, which maximizes

x_i/M_i and where M_i is the number of servers currently serving class i. Assigned servers will keep serving designated class until no jobs left. Until then, all servers assigned to that class will be re-assigned one by one to other classes still need service.

3.4.2.2 $dc\mu$ Heuristic (DCM)

Glazebrook et. al. [26] proposed a static policy which prioritize class $k^* = \underset{a\in\{1,2\}}{\text{argmax}}\{D_i c_i \mu_i\}$. We modified the cost term c_i in the original formula to $r_i - c_i\mu_i^{-1}$ to incorporate both the reward and holding cost. Specifically, the available server will be assigned to class a^* jobs given by the follows,

$$a^*(\mathbf{x}) = \underset{a\in\mathcal{A}}{\text{argmax}} \left\{ D_a \left(r_a - c_a \mu_a^{-1} \right) \mu_a \right\}. \tag{3.21}$$

3.4.2.3 Infinite Jobs Heuristic (IFH)

For this heuristic, we assume there are infinite number of jobs in each class. We use the formula in (3.19) to determine the optimal action by setting x_p to ∞. More specifically, the available server will be assigned to class a^* jobs given by the follows,

$$a^*_\infty = \underset{a\in\mathcal{A}}{\text{argmax}} \left\{ \mu_a \left(r_a + c_a D_a^{-1} \right) \right\}. \tag{3.22}$$

3.5 Numerical Studies

The results obtained for the two-class prioritization problem in Section 3.3 could be of theoretical interests. Therefore, we dedicated this numerical study to investigate the performance of the proposed heuristics in Section 3.4 comparing the optimal policy in an environment similar to the analytical model. We consider a model with three classes of jobs served by a single server in the numerical study. Assume both service times and deadline are exponentially distributed with rate $\mu_k > 0$ and $D_k > 0$ for all jobs in class $k \in \{1, 2, 3\}$. We generated the initial numbers of jobs $x_k, k \in \{1, 2, 3\}$ independently and uniformly over the set of integers between 1 and 30. We fixed the holding cost to be zero for all classes. The expected rewards $r_k, k \in \{1, 2, 3\}$ are drawn independently from uniform distributions with ranges $(0, 1)$. We considered several experiments setting with different deadline rates $D_k, k \in \{1, 2, 3\}$ and service rate $\mu_k, k \in \{1, 2, 3\}$. For each experiment setting, we generated 10,000 random scenarios. In each scenario, we computed the expected total reward for each heuristic policies using the optimality equation defined in (3.1) while replacing the action at each state by the action generated by the respective heuristics. The optimal expected total reward was computed using backward induction. Due to the curse of dimensionality, the optimal policy could only be computed when the number of classes in small and the number of jobs in each class is moderate. Then, we recorded the percentage deviation on the total reward of each heuristic from the optimal total reward for each scenario. We provided statistics including the 95% confidence interval (C.I.), minimum "Min"), first quantile ("Q1"), median ("Med"), third quartile ("Q3"), maximum ("Max"), of the percentage improvement. In addition, "#best" denotes the number of scenarios where each heuristic performs better than other heuristics (in case of a draw, all heuristics performs as good as the best one are all counted toward the best).

For the first set of experiments, we generate the service rates $\mu_k, k \in \{1, 2, 3\}$ uniformly from $(0, 3)$. On the other hand, we change the range used to generate the deadline rates $D_k, k \in \{1, 2, 3\}$. The results are presented in Table 3.2.

In first block of Table 3.2, the deadlines are generated uniformly from $(0.05, 1)$. TAH performs close to the optimal policy and better than all other heuristics at 5% significant level. Both DCM and INF are within 1% to the optimal policy. While DCM and INF have the largest number of best-performing scenarios, the performance of TAH is the most consistent with a maximum deviation from the optimal policy less than 5%. Other heuristics deviate from the optimal policy by at least 10% in the worst cases. In block two and three, as the deadline rates decrease, the performance of DAH, DCM, and TAH improved monotonically.

Heuristics	95% C.I.	Min.	Q1	Median	Q2	Max	#best.
			$D_i \sim$ Uniform $(0.05, 0.1)$				
DAH	14.88 ± 0.26	0.00	0.73	8.63	26.44	72.82	1436
INF	0.32 ± 0.02	0.00	0.00	0.00	0.09	11.84	7337
DCM	0.24 ± 0.01	0.00	0.00	0.00	0.05	12.73	7663
TAH	0.14 ± 0.01	0.00	0.00	0.01	0.11	4.19	5964
BCH	2.88 ± 0.05	0.00	0.58	1.77	4.12	22.55	195
			$D_i \sim$ Uniform $(0.01, 0.05)$				
DAH	11.43 ± 0.23	0.00	0.68	5.73	17.75	76.68	1446
INF	0.60 ± 0.03	0.00	0.00	0.00	0.28	17.93	6572
DCM	0.09 ± 0.01	0.00	0.00	0.00	0.00	10.01	8394
TAH	0.07 ± 0.00	0.00	0.00	0.00	0.03	3.72	7062
BCH	3.32 ± 0.06	0.00	0.75	2.07	4.67	23.95	149
			$D_i \sim$ Uniform $(0.005, 0.01)$				
DAH	6.24 ± 0.16	0.00	0.28	2.42	7.91	65.13	1645
INF	0.07 ± 0.00	0.00	0.00	0.00	0.00	4.97	8124
DCM	0.01 ± 0.00	0.00	0.00	0.00	0.00	5.15	9334
TAH	0.01 ± 0.00	0.00	0.00	0.00	0.00	2.93	8531
BCH	2.87 ± 0.06	0.00	0.52	1.39	3.64	26.90	82

Table 3.2: Percentage deviation from the optimum when deadlines vary.

When deadline rates are between 0.005 and 0.01, both DCM and TAH are on average 0.01 % away from the optimal policy. INF and BCH on the other hand, do exhibit monotone improvements as the deadlines getting larger.

For the second set of experiments, we generate the deadline rates $D_k, k \in \{1, 2, 3\}$ uniformly from $(0.005, 0.1)$. On the other hand, we change the range used to generate the service rates $\mu_k, k \in \{1, 2, 3\}$. The results are presented in Table 3.3.

In first block of Table 3.3, the service rates are generated uniformly from $(0, 1)$. TAH on average performs close to the optimal policy and better than all other heuristics at 5% significant level. DCM on average is also within 1% to the optimal policy and has the largest number of best-performing scenarios. The performance of TAH is the most consistent with the smallest maximum deviation from the optimal policy of 7.29% and also a large number of best-performing scenarios. Other heuristics deviate from the optimal policy by at least 1.5% on average and 17% in the worst cases. In block two and three, we increase the

Heuristics	95% C.I.	Min.	Q1	Median	Q2	Max	#best.
			$\mu_i \sim$ Uniform $(0,1)$				
DAH	16.37 ± 0.30	0.00	0.87	9.29	27.13	93.07	1245
INF	1.52 ± 0.06	0.00	0.00	0.00	0.92	29.74	5355
DCM	0.61 ± 0.03	0.00	0.00	0.00	0.16	17.03	6739
TAH	0.14 ± 0.01	0.00	0.00	0.00	0.07	7.29	6469
BCH	1.79 ± 0.04	0.00	0.11	0.63	2.28	20.69	666
			$\mu_i \sim$ Uniform $(1,2)$				
DAH	10.49 ± 0.20	0.00	0.56	5.66	16.76	80.81	1411
INF	2.21 ± 0.06	0.00	0.00	0.20	2.71	24.46	4278
DCM	0.21 ± 0.01	0.00	0.00	0.00	0.03	10.10	7532
TAH	0.07 ± 0.00	0.00	0.00	0.00	0.05	3.82	6807
BCH	1.37 ± 0.03	0.00	0.25	0.76	1.80	14.63	390
			$\mu_i \sim$ Uniform $(2,3)$				
DAH	8.89 ± 0.17	0.00	0.59	4.87	13.82	64.56	1248
INF	1.86 ± 0.05	0.00	0.00	0.16	2.31	23.39	4365
DCM	0.09 ± 0.01	0.00	0.00	0.00	0.00	4.12	8045
TAH	0.03 ± 0.00	0.00	0.00	0.00	0.02	1.25	7423
BCH	1.12 ± 0.02	0.00	0.25	0.71	1.54	10.71	370

Table 3.3: Percentage deviation from the optimum when service rates vary.

service rates to $(1, 2)$ and $(2, 3)$, the performance of DAH, DCM, TAH, and BCH improved monotonically. INF, on the other hand, performs closest to optimal when the service rate is small. This phenomenon makes sense as INF assume an infinite number of jobs thus favors slower services.

3.6 Simulation Studies

3.6 Simulation Setting

We test the performance of proposed heuristics in a hypothetical flooding scenario. The scenario is abstract from Hurricane Harvey and the resulting flooding in Harris County during the period of late August until early September in 2017.

We considered the area in between longitude of 29.35 and 30.35 and latitude of -94.85 and -95.85. It corresponds to a 69 by 69 miles^2 region contains the city of Houston and its suburbs. We discretized the region with a 24 by 24 grid to obtain the parameters for our simulations studies from data.

We derive the distribution of rescue requests using the rescue requests data collected during Hurricane Harvey from Data World [6]. After cleaning the data, we obtained 1862 data points each representing a rescue request with legitimate geographical coordinate which falls within our pre-specified region. Base on the 1862 pairs of longitude and latitude, we construct a discrete 2-dimensional empirical probability function. Let p_{ij} denotes the probability of a request from (i, j)'s cell in the 24 by 24 grid. Although this data set includes all kinds of emergency requests, we used the empirical probability function obtained from it to generate the distribution of requests to be rescued by air resources in our simulations.

We could not identify where exactly those people rescued by helicopters were transported to exactly. Instead, we identified the major shelters and assume all rescued people will be transported to the nearest one. We considered four shelters listed in [1]: Cypress Ranch High School, M.O.Campbell Education Center, George R. Brown Convention Center, and Bay Harbour United Methodist Church (Refer to Figure 3.1 for more details). Then, for cell (i, j) in the grid, we compute the distance from the center of the cell to the nearest shelter among those four shelters and used it as the mean distance d_{ij} for each cell in the simulation.

We only considered helicopter as servers in our simulation and denote the number of servers by M. According to Texas Department of Public Safety, at least 26 helicopters were assigned to the state of Texas Emergency Support Function 9 Search and Rescue (ESF 9 SAR) and 1056 people in total were rescued by air resources [72]. We adopted the cruise speed of a UH-60 Blackhawk in our simulation as the mean speed of all helicopters which is 170 miles per hour according to its Wikipedia page. We assume if a helicopter is assigned to a request at cell (i, j), it always starts and ends at the nearest shelter to the request location. The re-allocation of helicopters between the shelters are neglected. We specifically considered the round trip travel time and time for loading and unloading in our simulation. Therefore, the mean single trip travel time for a request at cell (i, j) equals $t_{i,j} = d_{ij}/170 + 10$

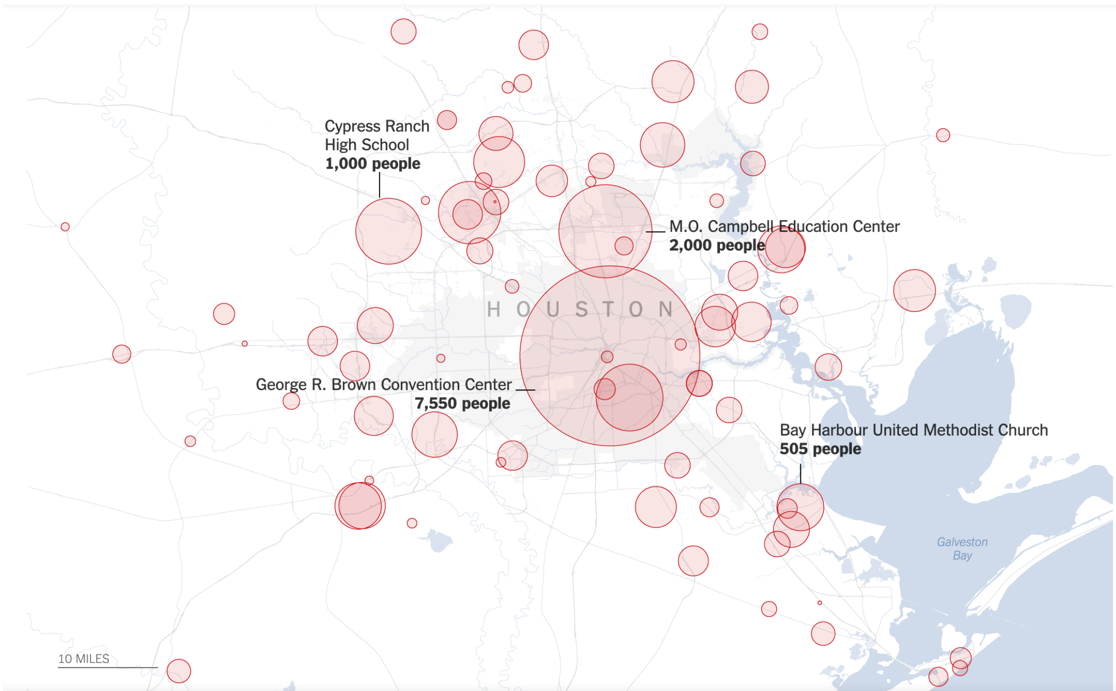

Figure 3.1: Number of people in shelters [1]

minutes. 10 minutes represents the loading time at rescue locations or the unloading time at shelters. We assume the time required for a single trip during a rescue mission follows a log-normal distribution with mean $t_{i,j}$ and variance $0.1t_{i,j}$.

3.6 Parameters in Heuristics

To adopt the heuristics introduced in Section 3.4, we divide all requests into 9 classes. We partitioned the region of interest into 9 classes according to geographical locations in our simulations. More specifically, we divide the squared region using 3 by 3 grid and each small square corresponds to a class. Each class now contains 64 cells defined in Section 3.6.1. The parameters for each class will be generated based on its containing cells.

There are $M > 0$ helicopters in our simulation. We used the pooled round trip service rate averaged across all cells belong to a given class k as the service rate for that class, i.e. $\mu_k = 0.5M \sum_{(i,j)\in\mathbb{K}} t_{ij}^{-1}$, where $\mathbb{K}$ is the set of all cells belongs to class $k \in \mathbb{C}$.

There is no clear specified deadline in search and rescue mission. We mainly view the deadline in our model as a measure of different level of urgent across the classes. When a deadline is reached for a class, even though all jobs in that class leave the system in our model, we do not mean that all people requesting rescue in that sub-area are dead. When the deadline is reached, those people have not been rescued will facing higher risks or more rigorous environment. We utilized the channel status on the Harrison County Flood Warning System (FWS) [28] to generate the deadlines for each class. Figure 3.2 show the main information panel of the FWS website. More specifically, we selected a few creeks within the sub-region of each class. For each creek, we identified the time interval for the stream elevation between the lastest time it exceeded the height of “Flooding Possible” and the first time it exceeded the height of “Flooding Likely” after the landfall of Hurricane Harvey. A short interval length indicates the rapid increasing of stream elevation and the flood will occur soon. If either the elevation never reached the height of “Flooding Likely” or the length of the interval is larger than 24 hours, we cap the interval at 24 hours since we only consider the rescue operation within one day. For example, stream elevation and other information of Cypress creek at I-45 can be found in Figure 3.3, the “Flooding Possible” height is 82.50 feet which occurred at 00:19 a.m. on August 27, and the “Flooding Likely” height is 85.50 feet which occurred at 2:55 a.m. on August 27. Thus, the length of the interval was 2.6 hours. The average length of all such intervals within each class is used as the deadline multiplier in the simulation. In the base case of our simulation, a deadline unit correspondence to an hour. The actual deadline used in the simulations for each class is deterministic and equals the deadline multiplier times the deadline unit. For example, a class with deadline multiplier of 5 will have 5 hours before the deadline arrives. When we vary the length of the deadlines, we change the length of the deadline unit. For example, when the deadline unit becomes 10 minutes, a class with deadline multiplier of 5 will have 50 minutes before the deadline arrives.

The total number of requests in each class equals the sum of all requests belongs to the class. We set the reward to be one for all requests completed before their deadlines and the costs are set to zero. The reward is collected when the helicopter arrives at the request location in the simulations.

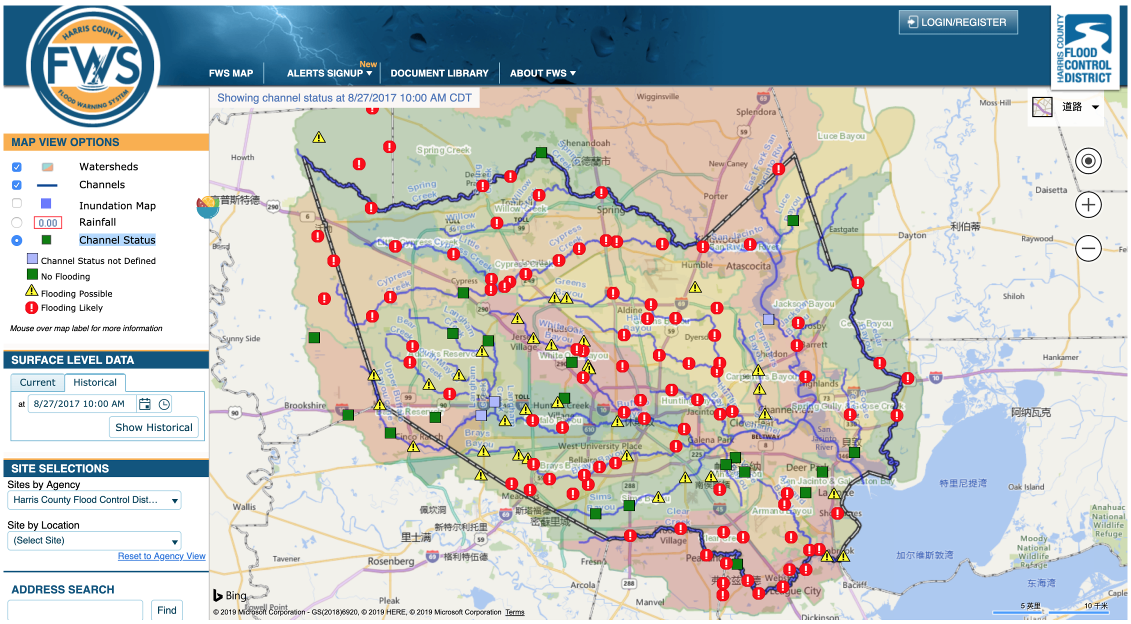

Figure 3.2: Channel status and flood warming on FWS [28]

3.6 Simulation Results

We test the heuristics proposed in Section 3.4 in various simulations scenarios. Each simulations scenario has a different set of parameters such as total number of request, number of helicopters, and deadlines. We replicate each scenario 100 times with random request distribution and travel time. Common random numbers are used across all heuristics. We report the mean percentage of rescue requests completed and its confidence interval as the result. All simulation were coded using Matlab 2017b.

We first vary the total number of requests while fixing the number of helicopters at 25 and deadline unit to an hour. We increase the number of requests from 10 to 1000 with a step size of 10. The results are presented in Table 3.4 and Figure 3.4.

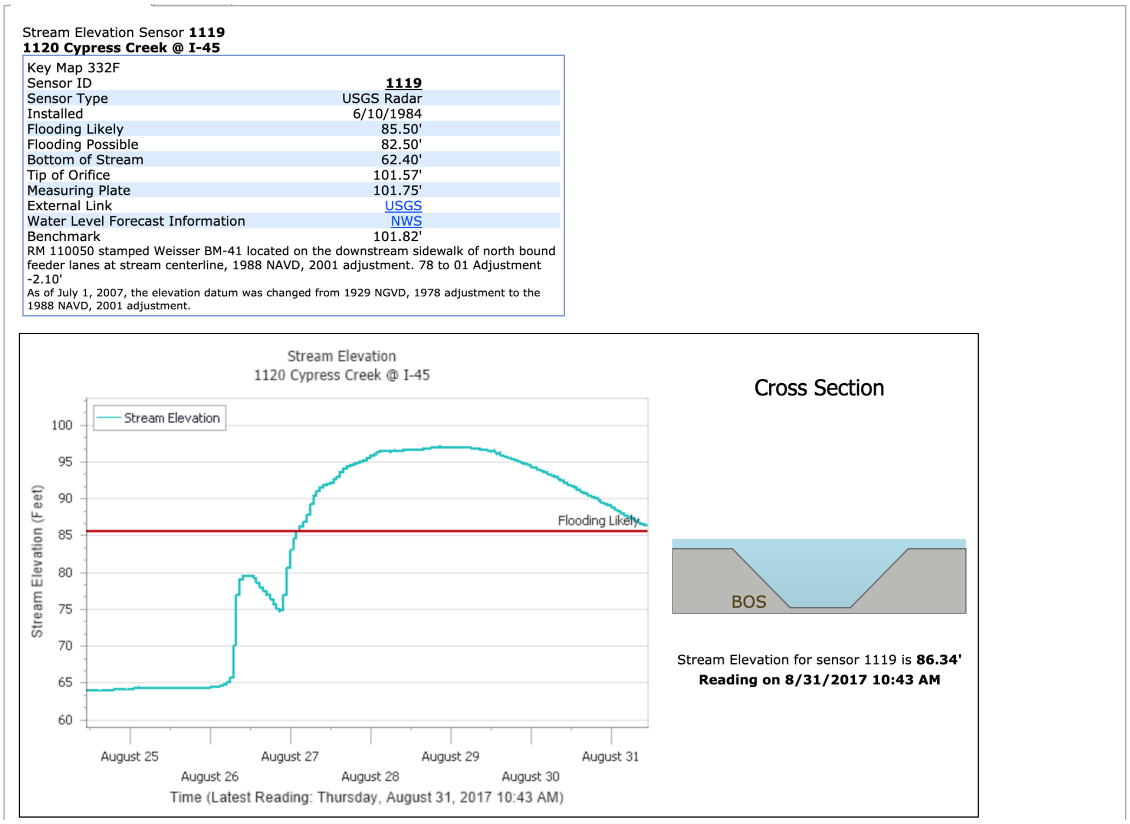

Figure 3.3: Information on Cypress creek at I-45 [28]

From Figure 3.4, we recognize that the percentage of completed requests decreases as the total number of requests increase for all heuristics. When the number of requests is small (around 200), both DCM and BAH-mean perform the best. As the number of requests increases, TAH performs the best and DAH and INF converge to TAH when the number of requests is close to 1000. The percentage of the uniquely best performing heuristic (at 5% significant level) is highlighted in bold for each scenario in Table 3.4.

Next, we fixed the total number of request to 300 and the deadline unit to an hour and vary the number of helicopters from 1 to 100. The results are presented in Table 3.5 and Figure 3.5.

The percentage of completed requests increases as the number of servers increases. From Figure 3.5, intuitively, TAH perform the best when the number of servers is small (around 20). DCM and BAH-mean become the best performer as the number of servers increases to

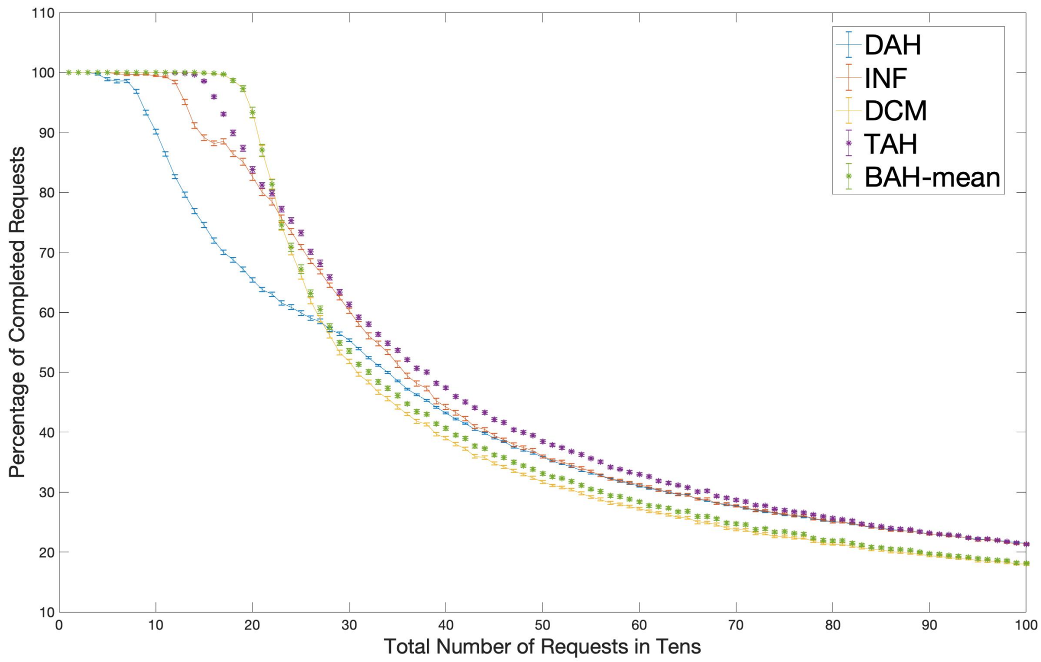

Figure 3.4: Request completion percentages v.s. No. requests

around 40. INF, DCM, TAH, and BAH-mean converge to 100 percent when more than 60 servers are available. In Table 3.5, the percentage of the uniquely best performing heuristic (at 5% significant level) is highlighted in bold for each scenario. Only TAH is the outperform all other heuristics at 5% significant level when the number of servers is from 15 to 30. DCM and BAH-mean perform similarly but better than other heuristics when the number of servers is between 35 and 50.

Last but not least, we fixed the total number of request to 300 and the number of helicopters at 25 while varying the deadline unit from 1 minutes to 60 minutes. The results are presented in Table 3.6 and Figure 3.6. We also look at the case where the number of servers increases to 35 while keep all other setting the same. The results are presented in Table 3.7 and Figure 3.7.

An interesting observation from Figures 3.6 and 3.7 is that the percentage of completed requests for all heuristics except DAH increases as the deadlines increase in the shape of ”step functions”. This might due to that the performance of heuristics will only be affected

Table 3.4: Request completion percentages v.s. No. requests (Fix No. server at 25 and deadline unit equals to 60 minutes).

No.Req.	Heuristics				
(×10)	DAH	INF	$Dc\mu$	TAH	BCH-Mean
5	98.86±0.26	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
10	90.15±0.39	99.45±0.11	100.00±0.00	99.99±0.02	100.00±0.00
15	74.56±0.42	89.13±0.48	99.89±0.05	98.57±0.20	99.89±0.05
20	65.38±0.39	82.57±0.57	93.31±0.87	83.78±0.53	93.33±0.88
25	59.86±0.41	70.88±0.42	66.19±0.66	**73.22±0.49**	67.18±0.71
30	55.33±0.23	60.28±0.43	51.79±0.38	**61.30±0.41**	53.54±0.41
35	48.54±0.16	51.31±0.53	44.22±0.34	**53.65±0.34**	46.11±0.38
40	43.19±0.14	44.20±0.43	38.99±0.28	**47.41±0.28**	40.64±0.31
45	38.98±0.14	39.45±0.35	34.77±0.26	**42.09±0.32**	36.20±0.32
50	35.81±0.14	35.96±0.25	31.68±0.24	**38.41±0.28**	33.09±0.29
55	33.15±0.13	33.48±0.21	29.18±0.20	**35.64±0.22**	30.47±0.23
60	31.02±0.16	31.18±0.22	27.22±0.21	**32.95±0.25**	28.38±0.26
65	29.49±0.14	29.55±0.22	25.75±0.22	**30.74±0.26**	26.80±0.29
70	27.61±0.12	27.67±0.18	23.78±0.22	**28.67±0.23**	24.71±0.28
75	26.20±0.11	26.36±0.16	22.58±0.21	**26.97±0.25**	23.43±0.27
80	24.99±0.14	25.07±0.21	21.34±0.20	**25.63±0.21**	21.87±0.23
85	23.99±0.13	23.95±0.17	20.34±0.17	24.25±0.21	20.75±0.21
90	22.95±0.10	22.94±0.15	19.38±0.18	23.16±0.18	19.75±0.20
95	22.13±0.12	21.97±0.17	18.53±0.20	22.18±0.19	18.89±0.23
100	21.42±0.10	21.35±0.15	17.99±0.19	21.27±0.18	18.16±0.20

when the small changes in deadlines cumulated. Also, for some heuristics such as INF and TAH, the improvement in the percentage of completed requests is not monotonic. This might due to the deadlines are exponential in the model but deterministic in the simulation. From Figure 3.6 and Table 3.6, when there are 25 servers, we observed that DAH performs the best when the deadlines are small. TAH is the first heuristic that outperforms DAH and all other heuristics at 5% significant level as the deadline unit exceed 55. From Figure 3.7 and Table 3.7, when the number of servers increases to 35 while keeping all other settings untouched, we observed that DAH again performs the best when the deadlines are small. TAH is the first heuristic that outperforms DAH and all other heuristics at 5% significant level as the deadline unit exceed 35 but less than or equals to 45. INF, DCM, DAH, and BAH-mean perform much better than DAH when the deadline unit is more than 50.

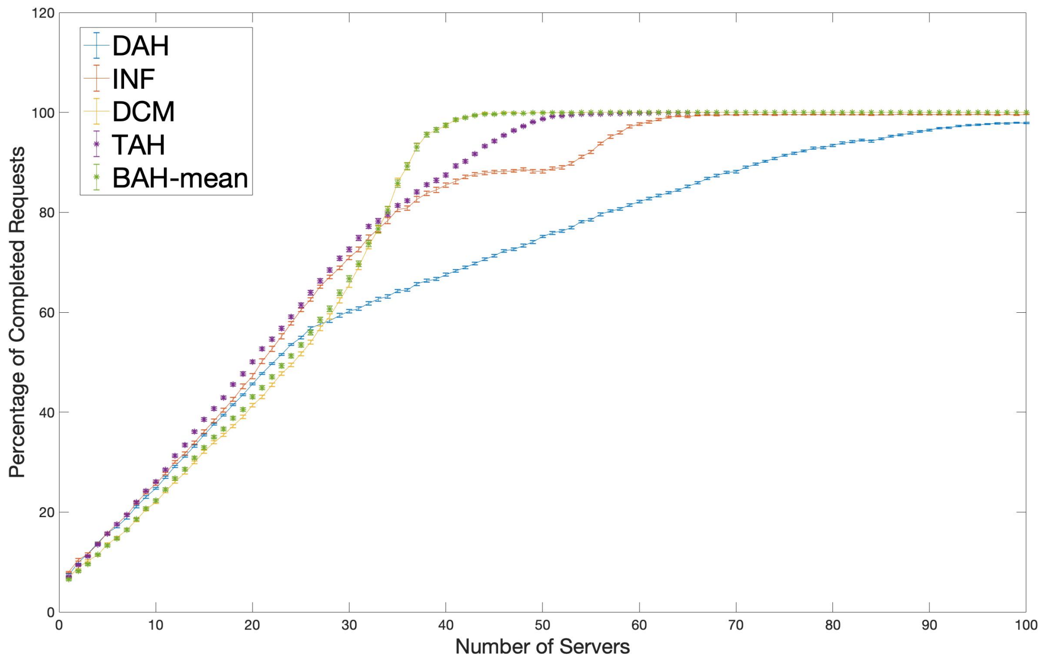

Figure 3.5: Request completion percentages v.s. No. servers

Overall speaking, TAH performs strongly when the number of jobs is large, the number of servers is small, and the deadline is moderate. DCM and BAH-mean often perform similarly. They perform well when the number of jobs is small, the number of servers is larger, and when deadlines are larger while the number of servers is relatively large. DAH perform plausible only when the deadlines are small.

3.7 Conclusion

In urban search and rescue operations, especially during catastrophic natural disasters, the huge number of requests overwhelms the limited resources. This study investigated the optimal allocation of limited resources to urgent demands. First of all, our generic scheduling model consists of a single server two classes of jobs with potentially distinct service rates, holding costs, rewards, and a common deadline for each class. We identified explicitly the optimal policy for the proposed model. Furthermore, based on the structure of the optimal policy, we developed multiple heuristics for a more general system with multiple servers and

Table 3.5: Request completion percentages v.s. No. servers (Fix no. requests at 300 and deadline unit equals to 60 minutes).

No.	Heuristics				
Ser.	DAH	INF	$Dc\mu$	TAH	BCH-Mean
5	15.83±0.20	15.78±0.24	13.57±0.25	15.67±0.25	13.32±0.27
10	24.75±0.21	25.61±0.26	22.02±0.28	26.08±0.30	22.30±0.29
15	35.44±0.20	36.10±0.34	32.08±0.30	**38.55±0.33**	32.90±0.31
20	45.65±0.17	47.20±0.50	41.43±0.34	**50.07±0.38**	43.07±0.38
25	54.94±0.25	60.46±0.35	51.71±0.40	**61.40±0.45**	53.45±0.44
30	60.21±0.34	70.94±0.42	65.55±0.58	**72.58±0.48**	66.72±0.63
35	64.25±0.31	80.59±0.46	86.08±0.76	81.39±0.37	85.76±0.76
40	67.51±0.31	85.47±0.44	97.39±0.39	87.46±0.38	97.44±0.39
45	71.30±0.28	88.09±0.30	99.61±0.11	94.26±0.31	99.63±0.11
50	75.16±0.23	88.24±0.33	99.91±0.03	98.68±0.11	99.95±0.02
55	78.55±0.26	92.08±0.28	100.00±0.00	99.67±0.05	99.99±0.01
60	82.14±0.26	97.69±0.26	100.00±0.00	99.85±0.04	100.00±0.00
65	85.20±0.26	99.14±0.14	100.00±0.00	99.96±0.02	100.00±0.00
70	88.14±0.22	99.51±0.08	100.00±0.00	100.00±0.00	100.00±0.00
75	91.42±0.19	99.57±0.06	100.00±0.00	100.00±0.00	100.00±0.00
80	93.41±0.22	99.54±0.07	100.00±0.00	100.00±0.00	100.00±0.00
85	94.73±0.20	99.52±0.07	100.00±0.00	100.00±0.00	100.00±0.00
90	96.47±0.17	99.56±0.06	100.00±0.00	100.00±0.00	100.00±0.00
95	97.51±0.13	99.61±0.06	100.00±0.00	100.00±0.00	100.00±0.00
100	97.88±0.12	99.62±0.06	100.00±0.00	100.00±0.00	100.00±0.00

more than two classes of jobs. The proposed heuristics perform close to the optimal in our numerical study.

Moreover, we designed hypothetical simulations abstracted from the real urban search and rescue operations during Hurricane Harvey in Houston, TX, U.S. in 2017. Unfortunately, our simulations were not as realistic as we want due to lack of data. Especially the common deadlines were generated using the elevation of the creeks and water channels rather than flood levels and the rewards are set to one rather than the actual number of people rescued in each operation. We still believe they reflect some aspects of the urban search and rescue operations carried out during Hurricane Harvey and are capable of providing meaningful feedback. The proposed heuristics performed well in the simulation as well. From the simulation results, we believe the proposed heuristics could be utilized by the commanders

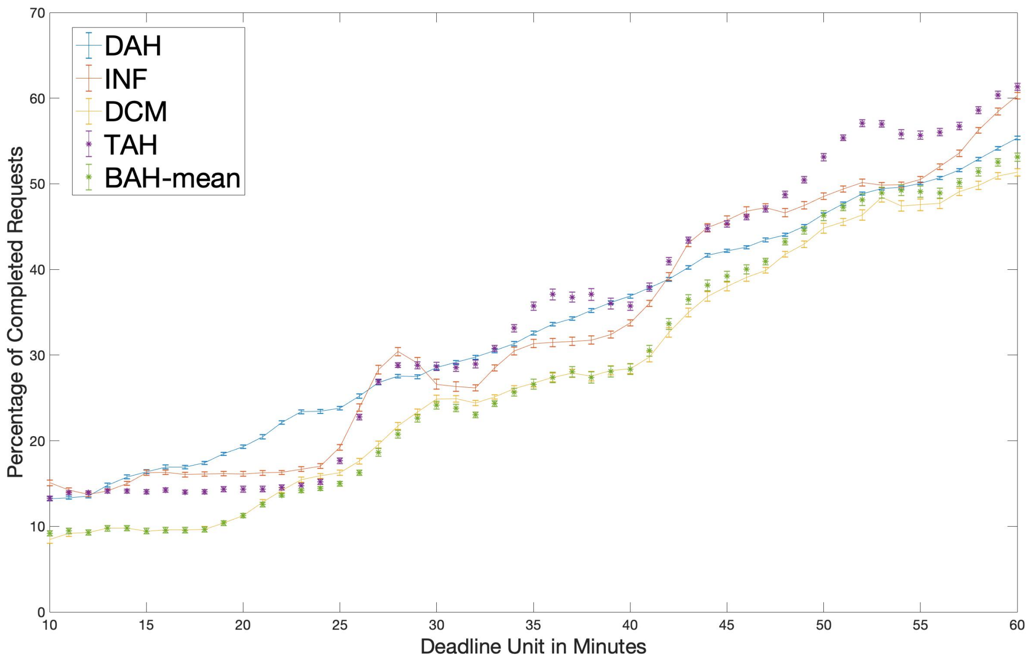

Figure 3.6: Request completion percentages v.s. deadline unit (25 servers)

of the urban search and rescue task forces in coordinating resources or be used in designing response plans.

Base on the simulations, we recommend applying TAH, a dynamic heuristic takes into consideration the number of requests, when resources are confined while exercising DCM, a simple static policy, for relative ample resources setting although further experiments in more detailed simulation or test events are necessary to support our claim. The simulation setup could be improved with a better classification of requests based on more details of the request, more concrete deadlines from weather forecasting or real-time monitoring data.

Our analytical findings could be interested in the scheduling community. Our approach has the potential to be extended to identify the structure-property for the optimal policy with more than two classes. Our heuristics could also be applied in many other situations beyond urban search and rescue, such as healthcare – coordinate emergency medical response personnel and resources and triage casualties in the aftermath of mass casualty incidents, service, and production – appointment scheduling with impatient customers and production

Table 3.6: Request Completion Percentages v.s. Deadlines (Fix No. server at 25 and No. request at 300).

Minutes Per Unit.	Heuristics DAH	INF	$Dc\mu$	TAH	BCH-Mean
10	13.23±0.23	**15.08±0.31**	8.46±0.44	13.26±0.26	9.19±0.31
15	16.38±0.24	16.25±0.29	9.44±0.32	14.04±0.25	9.44±0.32
20	**19.30±0.23**	16.12±0.28	11.26±0.27	14.33±0.35	11.26±0.27
25	**23.80±0.22**	19.23±0.33	16.27±0.30	17.66±0.32	14.98±0.27
30	28.54±0.24	26.60±0.60	24.88±0.42	28.67±0.46	24.15±0.46
35	32.56±0.22	31.36±0.48	26.73±0.44	**35.72±0.49**	26.60±0.59
40	**36.88±0.22**	33.76±0.36	28.39±0.56	35.71±0.49	28.37±0.62
45	42.18±0.18	45.75±0.53	38.02±0.51	45.33±0.38	39.20±0.58
50	46.47±0.15	48.54±0.38	44.83±0.58	**53.11±0.45**	46.28±0.58
55	50.06±0.16	50.52±0.35	47.55±0.68	**55.68±0.48**	49.09±0.67
60	55.35±0.20	60.31±0.39	51.36±0.43	**61.34±0.40**	53.11±0.46

planning with perishable materials, military rescue operations – evacuate wounded soldiers from battlefields, and humanitarian aid – evacuate refugees from war zone and distribute food and medicines to areas in need.

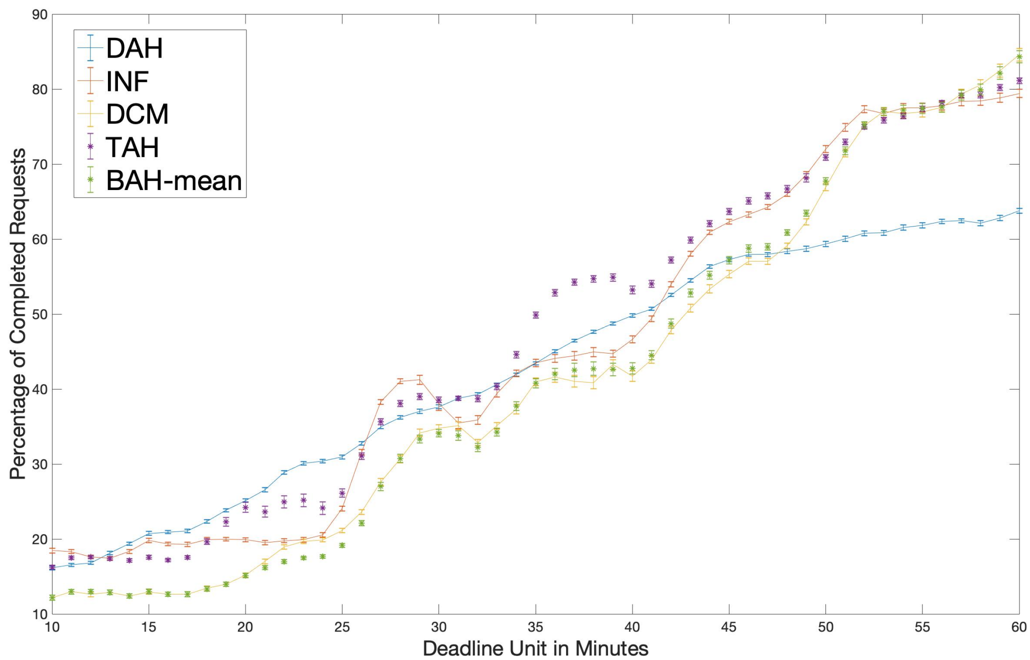

Figure 3.7: Request completion percentages v.s. deadline unit (35 servers)

Table 3.7: Request Completion Percentages v.s. Deadlines (Fix No. server at 35 and No. request at 300).

Minutes	Heuristics				
Per Unit.	DAH	INF	*Dcμ*	TAH	BCH-Mean
10	16.14± 0.26	**18.44± 0.30**	12.18± 0.36	16.23± 0.26	12.18± 0.36
15	**20.74± 0.25**	19.77± 0.29	12.92± 0.32	17.56± 0.26	12.99± 0.31
20	**25.14± 0.23**	19.91± 0.26	15.18± 0.28	24.20± 0.65	15.11± 0.27
25	**30.94± 0.25**	24.06± 0.32	21.11± 0.28	26.12± 0.57	19.15± 0.26
30	37.60± 0.26	38.05± 0.91	34.76± 0.49	38.54± 0.36	34.10± 0.50
35	43.43± 0.19	43.48± 0.49	40.95± 0.51	**49.86± 0.41**	40.76± 0.62
40	49.78± 0.23	46.60± 0.47	41.70± 0.70	**53.19± 0.52**	42.75± 0.77
45	57.25± 0.26	62.30± 0.32	55.31± 0.51	**63.67± 0.43**	57.17± 0.50
50	59.33± 0.33	**71.98± 0.46**	66.93± 0.49	70.87± 0.40	67.72± 0.46
55	61.81± 0.33	77.51± 0.62	76.98± 0.71	77.36± 0.34	77.41± 0.67
60	63.77± 0.32	79.41± 0.56	84.57± 0.82	81.11± 0.39	84.30± 0.78

APPENDIX A: PROOF OF PROPOSITION 1

To show that greedy algorithm provides the optimal solution to problem (2.14), we only need to proof that the action space, the set of feasible solutions, and the non-negative objective function form a finite weighted matroid.

Proof of Proposition 1. Let E denote the action space, i.e. the set of all possible values for all $a_{lh}^{(c)}$'s and obviously E is a finite set since $\mathbb{L}, \mathbb{H}$ and $\mathbb{C}$ are all finite set. Let $\mathcal{I}$ denote the collection of feasible set of the problem. We need to check the following conditions in order to proof the pair $(E, \mathcal{I})$ forms a finite matroid.

1. $\mathcal{I}$ set is not empty since empty set is a feasible solution.

2. Let A be a feasible solution and $A' \subset A$, then A' is also a feasible solution. This also holds trivially.

3. Let A and B be two feasible solutions and A has more elements than B, then there exists $a \in A \backslash B$ such that $B \cup \{a\}$ is also feasible. This also holds true by the following argument: A has more elements than B implies $\sum_{l \in \mathbb{L}} \sum_{h \in \mathbb{H}} \sum_{c \in \mathbb{C}} b_{lh}^{(c)} < \sum_{l \in \mathbb{L}} \sum_{h \in \mathbb{H}} \sum_{c \in \mathbb{C}} a_{lh}^{(c)} \leq \mathcal{A}$, for $\{b_{lh}^{(c)} : c \in \mathbb{C}, l \in \mathbb{L}, h \in \mathbb{H}\} \in B, \{a_{lh}^{(c)} : c \in \mathbb{C}, l \in \mathbb{L}, h \in \mathbb{H}\} \in A$. Suppose there exist a $b_{l'h'}^{(c')} < a_{l'h'}^{(c')}$, since $\sum_{h \in \mathbb{H}} a_{l'h}^{(c')} \leq w_{l'}^{(c')}$ and $\sum_{h \in \mathbb{H}} b_{l'h}^{(c')} \leq w_{l'}^{(c')}$, $\sum_{h \in \mathbb{H}, h \neq h'} b_{l'h}^{(c')} + a_{l'h}^{(c')} \leq w_{l'}^{(c')}$. Also, $\sum_{l \in \mathbb{L}} \sum_{h \in \mathbb{H}} \sum_{c \in \mathbb{C}} b_{lh}^{(c)} - b_{l'h}^{(c')} + a_{l'h}^{(c')} \leq \mathcal{A}$. Therefore, $B \cup \{a\}$ e.g., $\{b_{lh}^{(c)} : c \in \mathbb{C}, l \in \mathbb{L}, h \in \mathbb{H}\} \backslash \{b_{l'h}^{(c')}\} \cup \{c_{l'h}^{(c')}\}$ is also a feasible solution.

Hence, $(E, \mathcal{I})$ forms a finite matroid. Furthermore, the objective function is non-negative and together we have a weighted matroid. Then, following the result in [55], the greedy algorithm gives us an optimal solution to the problem defined in (2.14). □

APPENDIX B: BERNOULLI SPLITTING POLICY FOR POLICY IMPROVEMENT HEURISTIC

The policy improvement heuristic depends on the Bernoulli splitting policy as the initial static casualty distribution. Ideally, we would want to find the optimal splitting, i.e., the set of $\theta_{lh}^{(c)}$, which maximizes $V_{\infty}^{(c)}(\mathbf{X}_0)$ for any given initial casualty level $\mathbf{X}_0$ at facilities under some feasibility constraints. However, due to the non-linearity of the objective function such a constrained optimization problem is hard solve except for trivial cases. Instead, we obtain the optimal initial Bernoulli splitting from a fluid approximation of the problem based on the fluid model in [49]. In the fluid approximation, casualties are transported to medical facilities from casualty locations continuously. The casualties in class c from location l will be transported to facility h at a rate of $\mathcal{A}\theta_{lh}^{(c)}\lambda_{lh}$ per unit of time from λ_{lh}^{-1} on and generating a reward (survival probability) of $r_h^{(c)}$ per unit of time. Thus, the total discounted rewards (survival probabilities) of class c casualties from location l to facility h will be

$$\int_{\lambda_{lh}^{-1}}^{\infty} \mathcal{A}\theta_{lh}^{(c)}\lambda_{lh} r_h^{(c)} e^{-\alpha_c t} dt = \left(\frac{\mathcal{A}\theta_{lh}^{(c)}\lambda_{lh} r_h^{(c)}}{\alpha_c} \right) e^{-\alpha_c/\lambda_{lh}}.$$

Assuming all facilities are empty initially, the fluid approximation is then given as the follows:

$$\begin{aligned}
\max \quad & \mathcal{A}\sum_{l\in\mathbb{L}}\sum_{h\in\mathbb{H}}\sum_{c\in\mathbb{C}} \left(\frac{\theta_{lh}^{(c)}\lambda_{lh} r_h^{(c)}}{\alpha_c} \right) e^{-\alpha_c/\lambda_{lh}} \\
s.t. \quad & \mathcal{A}\sum_{l\in\mathbb{L}} \theta_{lh}^{(c)}\lambda_{lh} \leq \tilde{\mu}_h^{(c)}, \forall h \in \mathbb{H}, c \in \mathbb{C} \\
& \sum_{l\in\mathbb{L}}\sum_{h\in\mathbb{H}}\sum_{c\in\mathbb{C}} \theta_{lh}^{(c)} \leq 1 \\
& \theta_{lh}^{(c)} \geq 0, \forall l \in \mathbb{L}, h \in \mathbb{H}, c \in \mathbb{C}
\end{aligned} \tag{23}$$

The first constraint is the stability constraint for each queue at a medical facility. The second and third constraints are the feasibility constraints for the Bernoulli splitting of ambulances. The constrained maximization problem in (23) can be solved using linear programming. If we further simplify the problem by relaxing the first stability constraint, the problem in (23) can be solved using a greedy approach. We trim the solution obtained from the greedy approach with the stability constraints to obtain a feasible solution to (23). For details, refer to Algorithm 2.

Algorithm 2 Greedy Algorithm for Fluid Approximation Initialization of PIH

for all $l \in \mathbb{L}, h \in \mathbb{H}, c \in \mathbb{C}$ **do**
    $\theta_{lh}^{(c)} \leftarrow 0$
end for
$list \leftarrow \{(l, h, c), l \in \mathbb{L}, h \in \mathbb{H}, c \in \mathbb{C}\}$
Sort-Descending(list, $\alpha_c^{-1}\lambda_{lh}r_h^{(c)}e^{-\alpha_c/\lambda_{lh}}$)
for $k = 1$ to list.Length **do**
    $(l, h, c) \leftarrow list[k]$
    $\theta_{ij}^{(c)} \leftarrow \min\{1 - \sum_{l\in\mathbb{L}}\sum_{h\in\mathbb{H}}\sum_{c\in\mathbb{C}}\theta_{lh}^{(c)}, \tilde{\mu}_h^{(c)}/(\mathcal{A}\lambda_{lh}) - \sum_{l\in\mathbb{L}}\theta_{lh}^{(c)}\}$
end for

Proposition 6. Algorithm 2 yields a feasible solution to the optimization problem in (23).

The proof is similar to the proof for Algorithm 2 in Mills et al. [49]and hence omitted. The numerical experiments in [49] demonstrate that the performance of greedy solution to the problem in (23) is comparable to the optimal solution obtained by solving a linear program. Therefore, we will use the greedy solution to find the Bernoulli splitting probabilites for the policy improvement heuristics.

BIBLIOGRAPHY

[1] G. Aisch, S. Almukhtar, J. Ashkenas, M. Bloch, J. Burgess, A. Carlsen, F. Fessenden, T. Griggs, K. R. Lai, J. C. Lee, J. K. Patel, A. Pearce, B. Saget, A. Singhvi, J. Ward, J. White, and J. Williams. Maps: Tracking harveys destructive path through texas and louisiana. Available at https://www.nytimes.com/interactive/2017/08/24/us/hurricane-harvey-texas.html, 8 2017. [Accessed on April 4, 2019].

[2] D. Alexander. Towards the development of a standard in emergency planning. *Disaster Prevention and Management: An International Journal*, 14(2):158–175, 2005.

[3] D. Anderson. Hurricane harvey fire/ems after-action report. Available at http://www.hcffa.org/wp-content/uploads/2018/01/Hurricane-Harvey-AAR_Fire-EMS_Final.pdf, 2017. [Accessed on February 11, 2019].

[4] N. T. Argon, S. Ziya, and R. Righter. Scheduling impatient jobs in a clearing system with insights on patient triage in mass casualty incidents. *Probability in the Engineering and Informational Sciences*, 22(3):301–332, 2008.

[5] R. Aringhieri, M. E. Bruni, S. Khodaparasti, and J. Van Essen. Emergency medical services and beyond: Addressing new challenges through a wide literature review. *Computers & Operations Research*, 78:349–368, 2017.

[6] S. Arrazolo. Master rescue doc port arthur area, 8 2017. Data retrieved from Data World, https://data.world/sya/master-rescue-doc-port-arthur-area.

[7] L. Avila. Tropical storm harvey advisory number 37. Available at https://www.nhc.noaa.gov/archive/2017/al09/al092017.public.037.shtml, 8 2017. [Accessed on April 10, 2019].

[8] U. Ayesta, P. Jacko, and V. Novak. Scheduling of multi-class multi-server queueing systems with abandonments. *Journal of Scheduling*, 20(2):129–145, 2017.

[9] J. Bazyar, M. Farrokhi, and H. Khankeh. Triage systems in mass casualty incidents and disasters: a review study with a worldwide approach. *Open Access Macedonian Journal of Medical Sciences*, 7(3):482, 2019.

[10] V. Bélanger, A. Ruiz, and P. Soriano. Recent optimization models and trends in location, relocation, and dispatching of emergency medical vehicles. *European Journal of Operational Research*, 2018.

[11] M. Benson, K. L. Koenig, and C. H. Schultz. Disaster triage: Start, then savea new method of dynamic triage for victims of a catastrophic earthquake. *Prehospital and Disaster Medicine*, 11(2):117–124, 1996.

[12] E. S. Blake and D. A. Zelinsky. Hurricane harvey (al092017). Available at https://www.nhc.noaa.gov/data/tcr/AL092017_Harvey.pdf, 1 2017. [Accessed on July 10, 2019].

[13] M. L. Brandeau, J. H. McCoy, N. Hupert, J.-E. Holty, and D. M. Bravata. Recommendations for modeling disaster responses in public health and medicine: a position paper of the society for medical decision making. *Medical Decision Making*, 29(4):438–460, 2009.

[14] W. Cai, S. Borlace, M. Lengaigne, P. Van Rensch, M. Collins, G. Vecchi, A. Timmermann, A. Santoso, M. J. McPhaden, L. Wu, et al. Increasing frequency of extreme el niño events due to greenhouse warming. *Nature climate change*, 4(2):111, 2014.

[15] Y. Cao. Multiple server preemptive scheduling with impatience. *Probability in the Engineering and Informational Sciences*, 31(2):226–238, 2017.

[16] A. K. Childers, M. E. Mayorga, and K. M. Taaffe. Prioritization strategies for patient evacuations. *Health care management science*, 17(1):77–87, 2014.

[17] A. K. Childers, G. Visagamurthy, and K. Taaffe. Prioritizing patients for evacuation from a health-care facility. *Transportation Research Record*, 2137(1):38–45, 2009.

[18] D. C. Cone and D. S. MacMillan. Mass-casualty triage systems: a hint of science. *Academic Emergency Medicine*, 12(8):739–741, 2005.

[19] M. D. Dean and S. K. Nair. Mass-casualty triage: distribution of victims to multiple hospitals using the save model. *European Journal of Operational Research*, 238(1):363–373, 2014.

[20] D. Eckstein, M.-L. Hutfils, and M. Winges. Global climate risk index 2019. *Germanwatch, Bonn*, 2019.

[21] S. Einav, L. Aharonson-Daniel, C. Weissman, H. R. Freund, K. Peleg, I. T. Group, et al. In-hospital resource utilization during multiple casualty incidents. *Annals of Surgery*, 243(4):533, 2006.

[22] C. B. Field. *Climate change 2014–Impacts, adaptation and vulnerability: Regional aspects*. Cambridge University Press, 2014.

[23] E. R. Frykberg. Triage: principles and practice. *Scandinavian Journal of Surgery*, 94(4):272–278, 2005.

[24] A. Garner, A. Lee, K. Harrison, and C. H. Schultz. Comparative analysis of multiple-casualty incident triage algorithms. *Annals of Emergency Medicine*, 38(5):541–548, 2001.

[25] J. Gettleman. More than 1,000 died in south asia floods this summer. *The New York Times*, 2017.

[26] K. Glazebrook, P. Ansell, R. Dunn, and R. Lumley. On the optimal allocation of service to impatient tasks. *Journal of Applied Probability*, 41(1):51–72, 2004.

[27] Q. Gong and R. Batta. Allocation and reallocation of ambulances to casualty clusters in a disaster relief operation. *IIE Transactions*, 39(1):27–39, 2007.

[28] Harrison County Flood Warning System. Channel status. Available at `https://www.harriscountyfws.org/`. [Accessed on June 1, 2019].

[29] M. Hirsch, P. Carli, R. Nizard, B. Riou, B. Baroudjian, T. Baubet, V. Chhor, C. Chollet-Xemard, N. Dantchev, N. Fleury, et al. The medical response to multisite terrorist attacks in paris. *The Lancet*, 386(10012):2535–2538, 2015.

[30] G. Hotz, E. Ginzburg, G. Wurm, V. DeGennaro, D. Andrews, S. Basavaraju, V. Coronado, L. Xu, T. Dulski, D. Moffett, et al. Post-earthquake injuries treated at a field hospital-Haiti, 2010. *Morbidity and Mortality Weekly Report*, 59(51/52):1673–1677, 2010.

[31] A. Ingolfsson, S. Budge, and E. Erkut. Optimal ambulance location with random delays and travel times. *Health Care management science*, 11(3):262–274, 2008.

[32] E. U. Jacobson, N. T. Argon, and S. Ziya. Priority assignment in emergency response. *Operations Research*, 60(4):813–832, 2012.

[33] T. James, K. Glazebrook, and K. Lin. Developing effective service policies for multiclass queues with abandonment: asymptotic optimality and approximate policy improvement. *INFORMS Journal on Computing*, 28(2):251–264, 2016.

[34] J. L. Jenkins, M. L. McCarthy, L. M. Sauer, G. B. Green, S. Stuart, T. L. Thomas, and E. B. Hsu. Mass-casualty triage: time for an evidence-based approach. *Prehospital and Disaster Medicine*, 23(1):3–8, 2008.

[35] S. Jin, S. Jeong, J. Kim, and K. Kim. A logistics model for the transport of disaster victims with various injuries and survival probabilities. *Annals of Operations Research*, 230(1):17–33, 2015.

[36] J. Joseph, K. Mistovich, J. H. Keith, and Q. Brent. *Prehospital Emergency Care Plus New MMND Mybradylab.* PRENTICE HALL, 2013.

[37] A. Jotshi, Q. Gong, and R. Batta. Dispatching and routing of emergency vehicles in disaster mitigation using data fusion. *Socio-Economic Planning Sciences*, 43(1):1–24, 2009.

[38] B. Kamali, D. Bish, and R. Glick. Optimal service order for mass-casualty incident response. *European Journal of Operational Research*, 261(1):355–367, 2017.

[39] A. R. H. Kevin Sullivan and D. A. Fahrenthold. Harvey leaving record rainfall, at least 22 deaths behind in houston. 8 2017.

[40] G. LaFree and L. Dugan. Global Terrorism Database (GTD). `https://www.start.umd.edu/research-projects/global-terrorism-database-gtd`, 2018. (Accessed on 11/25/2018).

[41] 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, et al. Mass casualty triage: an evaluation of the data and development of a proposed national guideline. *Disaster Medicine and Public Health Preparedness*, 2(S1):S25–S34, 2008.

[42] D. Li and K. D. Glazebrook. An approximate dynamic programing approach to the development of heuristics for the scheduling of impatient jobs in a clearing system. *Naval Research Logistics (NRL)*, 57(3):225–236, 2010.

[43] S. A. Lippman. Applying a new device in the optimization of exponential queuing systems. *Operations Research*, 23(4):687–710, 1975.

[44] Z. Lu-Ping, J. M. Rodriguez-Llanes, W. Qi, B. van den Oever, L. Westman, M. Albela, P. Liang, C. Gao, Z. De-Sheng, M. Hughes, et al. Multiple injuries after earthquakes: a retrospective analysis on 1,871 injured patients from the 2008 wenchuan earthquake. *Critical Care*, 16(3):R87, 2012.

[45] M. S. Maxwell, M. Restrepo, S. G. Henderson, and H. Topaloglu. Approximate dynamic programming for ambulance redeployment. *INFORMS Journal on Computing*, 22(2):266–281, 2010.

[46] L. A. McLay and M. E. Mayorga. A dispatching model for server-to-customer systems that balances efficiency and equity. *Manufacturing & Service Operations Management*, 15(2):205–220, 2013.

[47] A. F. Mills. A simple yet effective decision support policy for mass-casualty triage. *European Journal of Operational Research*, 253(3):734–745, 2016.

[48] A. F. Mills, N. T. Argon, and S. Ziya. Resource-based patient prioritization in mass-casualty incidents. *Manufacturing & Service Operations Management*, 15(3):361–377, 2013.

[49] A. F. Mills, N. T. Argon, and S. Ziya. Dynamic distribution of patients to medical facilities in the aftermath of a disaster. *Operations Research*, 66(3):716–732, 2018.

[50] A. F. Mills, N. T. Argon, S. Ziya, B. Hiestand, and J. Winslow. Restart: A novel framework for resource-based triage in mass-casualty events. *Journal of Special Operations Medicine*, 14(1):30–39, 2014.

[51] T. Mizumoto, W. Sun, K. Yasumoto, and M. Ito. Transportation scheduling method for patients in MCI using electronic triage tag. In *Proceedings of International Conference on eHealth, Telemedicine, and Social Medicine (eTELEMED)*, pages 156–163. Citeseer, 2011.

[52] D. M. Navin, W. J. Sacco, and G. McGill. Application of a new resource-constrained triage method to military-age victims. *Military Medicine*, 174(12):1247–1255, 2009.

[53] U. D. of Health and H. Services. START adult triage algorithm. radiation emergency medical management(remm). `http://www.remm.nlm.gov/startadult.htm`. (Accessed on 11/25/2018).

[54] W. H. Organization. Mass casualty management systems: strategies and guidelines for building health sector capacity, 2007.

[55] J. G. Oxley. *Matroid theory*, volume 3. Oxford University Press, USA, 2006.

[56] M. Pinedo. Stochastic scheduling with release dates and due dates. *Operations Research*, 31(3):559–572, 1983.

[57] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., 1994.

[58] P. P. Repoussis, D. C. Paraskevopoulos, A. Vazacopoulos, and N. Hupert. Optimizing emergency preparedness and resource utilization in mass-casualty incidents. *European Journal of Operational Research*, 255(2):531–544, 2016.

[59] A. J. Rettke, M. J. Robbins, and B. J. Lunday. Approximate dynamic programming for the dispatch of military medical evacuation assets. *European Journal of Operational Research*, 254(3):824–839, 2016.

[60] M. Reuter-Oppermann, P. L. van den Berg, and J. L. Vile. Logistics for emergency medical service systems. *Health Systems*, 6(3):187–208, 2017.

[61] S. M. Ross. A sequential scheduling problem with impatient jobs. *Naval Research Logistics (NRL)*, 62(8):659–663, 2015.

[62] D. M. Roth. Tropical cyclone point maxima. Available at `https://www.wpc.ncep.noaa.gov/tropical/rain/tcmaxima.html`, 10 2017. [Accessed on April 10, 2019].

[63] W. J. Sacco, D. M. Navin, K. E. Fiedler, I. Waddell, K. Robert, W. B. Long, and R. F. Buckman. Precise formulation and evidence-based application of resource-constrained triage. *Academic Emergency Medicine*, 12(8):759–770, 2005.

[64] W. J. Sacco, D. M. Navin, R. K. Waddell, K. E. Fiedler, W. B. Long, R. F. Buckman Jr, et al. A new resource-constrained triage method applied to victims of penetrating injury. *Journal of Trauma and Acute Care Surgery*, 63(2):316–325, 2007.

[65] K. Shin and T. Lee. Emergency medical service resource allocation in a mass casualty incident by integrating patient prioritization and hospital selection problems. *Working Paper*, 2019.

[66] H. Siddique. South Asia floods kill 1,200 and shut 1.8 million children out of school. *The Guardian*, 2017.

[67] T. Stocker. *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2014.

[68] Z. Sun, N. T. Argon, and S. Ziya. Patient triage and prioritization under austere conditions. *Management Science*, 64(10):4471–4489, 2017.

[69] I. Sung and T. Lee. Optimal allocation of emergency medical resources in a mass casualty incident: Patient prioritization by column generation. *European Journal of Operational Research*, 252(2):623–634, 2016.

[70] G. Super, S. Groth, R. Hook, et al. START: Simple Triage and Rapid Treatment plan. *Newport Beach, CA: Hoag Memorial Presbyterian Hospital*, 199, 1994.

[71] J. Tate. Storm summary number 15 for tropical storm harvey rainfall and wind. Available at https://www.wpc.ncep.noaa.gov/winter_storm_summaries/storm19/stormsum_15.html, 8 2017. [Accessed on April 10, 2019].

[72] W. Weidler. Hurricane Harvey Overview. Available at https://www.preparingtexas.org/Resources/documents/TDEM%20Training/2018%20Conference%20Presentations/Hurricane%20Harvey%20Texas%20Strong.pdf.