12-2022

Models and algorithms for trauma network design.

Sagarkumar Dhirubhai Hirpara
University of Louisville

MODELS AND ALGORITHMS FOR TRAUMA NETWORK DESIGN

By

Sagarkumar Dhirubhai Hirpara
B.E., Gujarat Technological University, 2015
M.S., Wright State University, 2019

A Dissertation
Submitted to the Faculty of the
J. B. Speed School of Engineering of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy in Industrial Engineering

Department of Industrial Engineering
University of Louisville
Louisville, Kentucky

December 2022

MODELS AND ALGORITHMS FOR TRAUMA NETWORK DESIGN

By

Sagarkumar Dhirubhai Hirpara
B.E., Gujarat Technological University, 2015
M.S., Wright State University, 2019

A Dissertation Submitted on

November 18, 2022

to the following Dissertation Committee:

____

Dissertation Director, Pratik Parikh, Ph.D.

____

Faisal Aqlan, Ph.D.

____

Monica Gentili, Ph.D.

____

Nan Kong, Ph.D.

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my life-long advisor, Dr. Pratik Parikh, for his excellent guidance, great support, wisdom, encouragement, and immense knowledge he has given me over last 5 years. I have learned a so much from him as a researcher and also a person.

I would like to thank Dr. Nan Kong for sharing his unique perspective, guidance, and support. In addition, I would also to like to acknowledge my committee members, Dr. Faisal Aqlan and Dr. Monica Gentili for their encouragement and suggestions.

I would also like to thank the faculty and staff of the Industrial Engineering department for their generosity and support. I would also like to extend my sincere gratitude to everyone in the Data Analytics and Optimization Laboratory for the numerous way they have helped me to finish this research.

Finally, I am grateful to my family and friends for the constant love and support. This work would have surely not be possible without all of you.

ABSTRACT

MODELS AND ALGORITHMS FOR TRAUMA NETWORK DESIGN

Sagarkumar Dhirubhai Hirpara

November 18, 2022

Trauma continues to be the leading cause of death and disability in the US for people aged 44 and under, making it a major public health problem. The geographical maldistribution of Trauma Centers (TCs), and the resulting higher access time to the nearest TC, has been shown to impact trauma patient safety and increase disability or mortality. State governments often design a trauma network to provide prompt and definitive care to their citizens. However, this process is mainly manual and experience-based and often leads to a suboptimal network in terms of patient safety and resource utilization.

This dissertation fills important voids in this domain and adds much-needed realism to develop insights that trauma decision-makers can use to design their trauma network. In this dissertation, we develop multiple optimization-based trauma network design approaches focusing minimizing mistriages and, in some cases, ensuring equity in care among regions. To mimic trauma care in practice, several realistic features are considered in our approach, which include the consideration of: (i) both severely and non-severely injured trauma patients and associated mistriages, (ii) intermediate trauma centers (ITCs)

along with major trauma centers (MTCs), (iii) three dominant criteria for destination determination, and (iv) mistriages in on-scene clinical assessment of injuries.

Our *first* contribution (Chapter 2) proposes the Trauma Center Location Problem (TCLP) that determines the optimal number and location of major trauma centers (MTCs) to improve patient safety. The bi-objective optimization model for TCLP explicitly considers both types of patients (severe and non-severe) and associated mistriages (specifically, system-related under- and over-triages) as a surrogate for patient safety. These mistriages are estimated using our proposed notional tasking algorithm that attempts to mimic the EMS on-scene decision of destination hospital and transportation mode. We develop a heuristic based on Particle Swarm Optimization framework to efficiently solve realistic problem sizes. We illustrate our approach using 2012 data from the state of OH and show that an optimized network for the state could achieve 31.5% improvement in patient safety compared to the 2012 network with the addition of just one MTC; redistribution of the 21 MTCs in the 2012 network led to a 30.4% improvement.

Our *second* contribution (Chapter 3) introduces a Nested Trauma Network Design Problem (NTNDP), which is a nested multi-level, multi-customer, multi-transportation, multi-criteria, capacitated model. The NTNDP model has a bi-objective of maximizing the weighted sum of equity and effectiveness in patient safety. The proposed model includes intermediate trauma centers (TCs) that have been established in many US states to serve as feeder centers to major TCs. The model also incorporates three criteria used by EMS for destination determination; i.e., patient/family choice, closest facility, and protocol. Our proposed '3-phase' approach efficiently solves the resulting MIP model by first solving a relaxed version of the model, then a Constraint Satisfaction Problem, and a modified version of the original optimization problem (if needed). A comprehensive experimental study is conducted to determine the sensitivity of the solutions to various system parameters. A case study is presented using 2019 data from the state of OH that shows more than 30% improvement in the patient safety objective.

In our *third* contribution (Chapter 4), we introduce Trauma Network Design Problem considering Assessment-related Mistriages (TNDP-AM), where we explicitly consider mistriages in on-scene assessment of patient injuries by the EMS. The TNDP-AM model determines the number and location of major trauma centers to maximize patient safety. We model assessment-related mistriages using the Bernoulli random variable and propose a Simheuristic approach that integrates Monte Carlo Simulation with a genetic algorithm (GA) to solve the problem efficiently. Our findings indicate that the trauma network is susceptible to assessment-related mistriages; specifically, higher mistriages in assessing severe patients may lead to a 799% decrease in patient safety and potential clustering of MTCs near high trauma incidence rates.

There are several implications of our findings to practice. State trauma decision-makers can use our approaches to not only better manage limited financial resources, but also understand the impact of changes in operational parameters on network performance. The design of training programs for EMS providers to build standardization in decision-making is another advantage.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Trauma is a body wound caused by sudden physical injury likely from a motor-vehicle crash, gunshot, fall, or violence and requires immediate medical attention. Traumatic injury is a major public health problem worldwide, with 4.4 million deaths yearly, nearly 8% of all deaths (WHO, 2021). In the US, it continues to be the #1 cause of death, disability, and morbidity for individuals under the age of 44 (#3 across all ages), with almost 200,000 deaths annually (CDC, 2022). The Centre for Disease Control and Prevention (CDC) estimated the economic cost of injuries in 2019 as $4.2 trillion, including $327 billion in medical care, $69 billion in work loss, and $3.8 trillion in value of statistical life and quality of life losses (CDC, 2021).

To ensure continuum of care for trauma patients, state governments often establish an integrated and coordinated trauma care system. There are three major phases of such a system: (i) prehospital, (ii) acute care, and (iii) rehabilitation. These are elaborated below.

1.1 Elements of a Trauma Care System

The ‘prehospital’ phase is a critical link between the occurrence of injury and the care provided at a hospital with the goal of ensuring prompt and definitive care. This phase is triggered by a 911 call seeking help for an injured victim (patient). The emergency medical dispatcher coordinates with various Emergency Medical Services (EMS) and dispatches appropriate EMS responders to the incident location (also known as the scene). EMS paramedics have two key decisions to make, in sequence: (i) determine the severity of the injury and (ii) determine the appropriate destination hospital and transportation mode.

The 'acute care' phase is initiated as soon as this patient arrives at the hospital's emergency department. The trauma response team follows a set protocol of treating the patient and transferring them to either to another hospital that offers appropriate medical services or the current hospital's operating room (or even) or intensive care according to the patient's condition.

The 'rehabilitation phase' is initiated after patient has been discharged from the hospital (inpatient care) and has been deemed to follow additional care at a rehabilitation facility (or even at home under the supervision of a provider).

The focus of this research is on the 'prehospital' phase, specifically, the impact of different on-scene operational decisions and strategic decision of locating trauma centers on patient safety.

1.2 Role of On-Scene Decision-Making

Once the EMS arrives at the scene, subsequent decisions become vital for patient safety. As pointed out earlier, on-scene decision-making practice involves two components: (i) injury assessment and (ii) destination determination. For assessment, EMS providers check the patient's vitals and perform various diagnostic tests to determine the underlying severity of the injury. On-scene injury assessment can be challenging and time-sensitive; therefore, national and state organizations have proposed guidelines to assist EMS paramedics during this process; these guidelines are referred to as Field Triage Guidelines (FTG). A correct injury assessment (severe or non-severe) is vital as assessed severity during this step becomes the basis for the destination determination decision.

Based on the assessed severity of the patient, the EMS paramedics will then evaluate access to a variety of trauma and non-trauma hospitals in the vicinity. The access time may different based on ground or air transportation; air transport is limited in many cases, though. The ACS FTG does suggest the type of hospital based on the injury severity; some states also have similar guidelines for the EMS. However, not every time is the guideline followed by the EMS. Other criteria such as patient or family choice or closest facility to the scene have been reported in the literature, sometimes upwards of 60% of the times collectively. That is, based on the assessed injury and the chosen guideline or criterion, the network of trauma centers can influence what hospital is eventually selected by the EMS providers.

1.3 Role of Trauma Centers

Trauma centers are specialized hospitals equipped and operated to provide a designated level of care for patients suffering traumatic injuries. The American College of Surgeons (ACS) has verified and categorized trauma centers based on their level of care, from Level I (L1) to Level V (L5). Levels I and II are equipped to provide definitive care for patients suffering from major traumatic injuries (severely injured) and are referred to as major trauma centers (MTCs). Major trauma centers are capable of providing care for every aspect of injury, from prevention to rehabilitation. The lower-level trauma centers

(Levels III - V) are intermediate facilities (ITCs) that only provide a subset of services provided by MTCs, only part of the day, and serve as centers for initial care, resuscitation, and transfer to major trauma centers. All other hospitals are referred to as non-trauma centers (NTCs), which are the ideal destination for non-severely injured trauma patients.

1.4 Motivation and Research Questions

Due to the time-sensitive nature of traumatic injuries, timely access to a trauma center is a vital factor in patient outcome (Branas et al., 2013; Jansen et al., 2015). The survival of severely injured patients improves by 25% if they receive care at a major trauma center relative to the care delivered at an NTC (MacKenzie et al., 2006). However, according to CDC, nearly 45 million Americans have no access to advanced trauma centers within 60 minutes (known as a golden hour). The main reason for this is the geographic maldistribution of MTCs. Such maldistribution impacts the time to reach a trauma center from the scene, forcing EMS paramedics to transport a patient to an inappropriate hospital based on their injury severity, which is referred to as a mistriage. Such mistriages have negative implications on patient safety and can lead to a higher likelihood of an adverse outcome such as disability, morbidity, and even mortality or unnecessary higher medical bills and financial burden on the trauma system.

ACS Committee on Trauma (ACS COT) has developed a Needs-Based Assessment of Trauma System (NBATS) tool to determine the required number of MTCs in a given geographical area, also known as the trauma service area (TSA). However, NBATS is limited in how it evaluates the need for the number of MTCs in the region; it also does not suggest the location of the MTCs and their impact on patient safety (ACS-NBATS, 2015).

Few studies have attempted to employ optimization-based approaches to design a trauma network; however, fundamental questions are yet to be addressed that form the basis of this research, which are summarized as follows:

Q1. What is the optimal network of MTCs that maximizes patient safety?

Q2. How sensitive is the network to changes in system parameters?

Q3. How do intermediate TCs (ITCs) support patient safety?

Q4. What effect do destination determination criteria have on the MTC/ITC network?

Q5. What is the impact of focusing on equity of patient safety on a trauma network's performance?

Q6. How sensitive is the MTC/ITC network to the distribution of trauma patients?

Q7. How do mistriages during on-scene injury assessment (operational decision) impact patient safety and the network of major trauma centers (strategic decision)?

Q8. Do mistriages have a higher impact on patients with moderate or those with severe injuries?

We address research Q1-Q2 in Chapter 2, Q3-Q6 in Chapter 3, and Q7-Q8 in Chapter 4. These are described below.

1.5 Research Contributions

In the following, we present the structure of this research and briefly outline the contributions of each chapter:

1.5.1 Trauma Center Location Problem (TCLP) – Chapter 2

In this chapter, we introduce the TCLP that determines the optimal number and location of major TCs to improve patient safety. To address Q1, our proposed bi-objective optimization model accounts for the two types of mistriages as a surrogate for patient safety. A mistriage is classified as a mismatch in the injury severity of a patient and the destination hospital type. We propose a notional tasking algorithm that mimics EMS's on-scene decision-making and determines the destination hospital and transportation type for each patient based on their injury severity. We propose a heuristic based on the Particle Swarm Optimization framework to derive efficiently near-optimal solutions for realistic problem sizes.

To illustrate the use of our approach, we use 2012 data from the state of Ohio to analyze the network's sensitivity to changes in system parameters (Q2). We observe that the solutions are sensitive to the choice of weights for two mistriages in the objective, volume requirements at a MTC, and the two thresholds used to mimic EMS decisions. The optimized network for the state of Ohio using our approach results in over 31.5% improvement in patient safety with only 1 additional MTC; redistribution of the existing 21 TCs led to a 30.4% improvement.

1.5.2 Nested Trauma Network Design Problem (NTNDP) – Chapter 3

This chapter extends the model developed in Chapter 2 and incorporates several essential dynamics when it comes to trauma patient care. To address Q3, we include intermediate trauma centers (ITCs) that are set up in many states in the US to serve as feeder centers to major TCs to improve patient safety. Next, we incorporate three criteria to determine the destination hospital to capture the dynamics of EMS's on-scene decision-making (Q4). Most state trauma systems are divided into subareas known as regions or districts to oversight trauma care within that region. Because of the existence of such regions, it becomes critical to consider equity in patient safety among regions when designing a state-wide trauma network (Q5).

To address Q3, Q4 and Q5, we propose a Nested Trauma Network Design Problem (NTNDP), which can be characterized as a nested multi-level, multi-customer, multi-transportation, multi-criteria, capacitated model with the bi-objective of maximizing the weighted sum of equity and effectiveness in patient safety . In the proposed model, multi-choice refers to the inclusion of all 3 dominant criteria for destination determination. While equity quantifies the similarity in patient safety across regions, effectiveness quantifies overall patient safety in the state.

To solve this model, we propose a '3-phase' solution approach that first solves a relaxed version of the model, then solves a Constraint Satisfaction Problem, and a modified version of the original optimization problem (if needed), all using a commercial solver. To address Q6, we consider a collection of counties in an existing midwestern US state and refer to it as a TSA. Using the data collected from several state trauma reports, we generated three different distributions of the patients in the TSA.

Results indicate that substantial improvement in patient safety can be achieved by using only protocol criteria for destination determination as suggested by the ACS. We observe that clustered distribution of patients improves patient safety compared to other distributions. While equity among regions is essential, our results indicate that an emphasis on only equity (and ignoring effectiveness) in a network may lead to a decline in overall patient safety. Finally, we also illustrate our approach using real data from the state of Ohio and delineate opportunities to improve performance by 30%.

1.5.3 Trauma Network Design Problem Considering Assessment Mistriages (TNDP-AM) - Chapter 4

While literature suggests mistriages could occur during the on-scene injury assessment phase, prior research and our earlier models assume 100% certainties (no mistriages) in designing the trauma network. To address this limitation (Q7), we propose a stochastic nested multi-level, multi-transportation capacitated model that explicitly considers mistriages in injury assessment and determines the number and locations of trauma centers to maximize patient safety. Because injury assessment is a binary classification problem (severe or non-severe), we model assessment-related mistriages in the ISS 9-15 (moderate injuries) and ISS>15 (severe injuries) via a Bernoulli random variable. To solve the model efficiently, we propose a Simheuristic approach that integrates Monte Carlo Simulation with a genetic algorithm (GA). We incorporate a feasibility algorithm in the proposed GA to convert infeasible solutions during offspring generation into feasible ones.

Our findings suggest that solutions are sensitive to mistriages in assessing severe patients (ISS>15) and may lead to the clustering of MTCs near high trauma incidence rates. The trauma network is also sensitive to mistriages in assessing non-severe (ISS 9-15) and the resulting network tends to have a fewer and dispersed distribution of MTCs.

1.6 Dissertation Outline

The remainder of this dissertation is organized as follows. Chapter 2 introduces the trauma center location problem. The nested trauma network design problem is summarized in Chapter 3. Chapter 4 proposes a trauma network design problem considering assessment errors. Finally, Chapter 5 summarizes this research and offers guidance for future research in this area.

CHAPTER 2

TRAUMA CENTER LOCATION PROBLEM (TCLP)

2.1 Introduction

Trauma is a body wound caused by sudden physical injury likely from a motor-vehicle crash, gunshot, fall, or violence and requires immediate medical attention (Cho et al., 2014). It is the #1 cause of death, disability, and morbidity for those under the age of 44 in the United States, resulting in almost 200,000 deaths and an economic burden of over $670 billion annually (ACS, 2016).

The hospitals equipped and operated to provide a designated level of care for patients suffering from major traumatic injuries are referred to as trauma centers, TCs (Cho et al., 2014). The American College of Surgeons (ACS) has verified and categorized TCs based on their level of care, from Level I (L1) to Level V (L5). Both L1 and L2 are designated major trauma centers with access to high-quality medical and nursing care, and highly sophisticated surgical and diagnostic equipment. They are required to have 24/7 in-house coverage and prompt availability in surgical specialties such as orthopedic, neurology, radiology, and even burn. On the other hand, the lower level of trauma centers (L3-L5) are intermediate facilities that only provide a subset of these services, only part of the day, and serve as centers for initial care, resuscitation, and transfer to L1/L2 centers (TCs). Because L1/L2 centers are destinations for appropriate care of severely injured trauma patients, we refer to them as trauma centers (TCs) in this study; all other lower-level trauma facilities and community hospitals are referred to as non-trauma centers (NTCs), which are ideal destinations for the non-severely injured trauma patients.

2.1.1 On-field EMS decision making

When a trauma event occurs, the subsequent Emergency Medical Service (EMS) decision making process involves two components; (i) injury assessment and (ii) destination determination. In (i), the EMS providers focuses on the extent of the injury using various diagnostic tests and underlying clinical factors to determine if the injury is severe or not. In (ii), the providers use this injury severity level and the network of hospitals nearby to determine which hospital is reachable in a certain timeframe and using what transportation mode (ground or air). Both components of the EMS decision-making are vital for the appropriate triage of the patient. An error during any step of EMS decision-making results in the mistriage of the patient. In (i), incorrect classification of the injury type (severe or non-severe) results in the 'clinical mistriage.' For instance, classifying a severely injured patient as non-severe (during the diagnosis on the scene) and subsequently transporting to NTC. While in (ii), if a patient is not transported to the right hospital based on injury severity due to any reason, then it results in 'system-related mistriage.' For example, transporting a severely injured patient to NTC is a type of 'system-related mistriage.' We included the modifier 'system-related' because these mistriages are due to system-related parameters such as network of hospitals and transportation resources that impact the determination of the hospital type (TC vs. NTC).

2.1.2 State-of-the art in trauma care

While a number of approaches have been proposed for injury assessment (Rotondo et al., 2014; Parikh et al., 2017), the impact of the underlying network on destination determination has recently received significant attention. Because trauma is a time-sensitive condition, timely access to an L1/L2 TC is one of the key determinants of outcome in a trauma care system (Branas et al., 2013; Jansen et al., 2015). If a severely injured trauma patient is able to receive care at a L1 trauma center, then his/her survival improves by 25% relative to the care delivered at an NTC (MacKenzie et al., 2006).

However, according to the Centers for Disease Control and Prevention, “there is no access to an advanced trauma center for nearly 45 million Americans within the golden hour (60 minutes)” (ACS, 2016). The reason for this is the geographic maldistribution of TCs in the U.S.; in 2010, nearly 9 states had a clustered pattern, 22 had a dispersed pattern, and 10 had a random pattern of TC distribution in the U.S. (Brown et al., 2016). Figure 1 shows the distribution of nearly 520 L1/L2 TCs in the U.S. with a coverage of 90.8% of the total

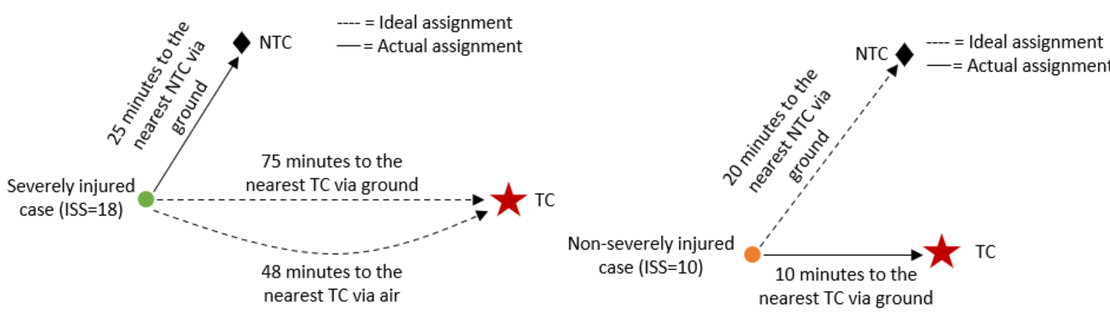
Figure 1. Network of L1/L2 TCs in U.S. Dark dots=TCs, dark shade = 60-minute coverage

population in 60 minutes across 30.38% land via ambulance and helicopter; for 45 minutes coverage, the coverage drops substantially to 76.72% population and 14.09% land (Branas et al., 2005; Carr & Branas, 2010).

2.1.3 System-related mistriages

The geographical maldistribution of TCs affects the time to reach a TC from the incidence location (i.e., field) by the EMS provider, and subsequently result in either system-related under-triage (srUT) or system-related over-triage (srOT). A lack of a TC within a prespecified time (per clinical recommendations, usually 45 minutes upon EMS arrival) from the field can *compel* the EMS providers to take a severely injured patient to a nearby NTC, which is referred to as system-related under-triage (srUT). Figure 2(a) illustrates of the case of a severely injured patient transported to a nearby NTC because the nearest TC is not accessible within prespecified time (45 minutes) via ground and air. Note that in case of severe injuries, it is vital to transport such patients to the nearest TC, and not just to any TC which meets the prespecified time (Chen et. al., 2019; Gauss et al., 2019).

Similarly, an excess (or cluster per Brown et al., 2016) of TCs in the vicinity of a field can *induce* the EMS providers to take a less severely injured patient to one of those



(a) srUT (b) srOT

Figure 2. An example of system-related mistriages

TCs (instead of an NTC), which is referred to system-related over-triage (srOT), the other form of mistriage (Newgard et al., 2016). An example of srOT is shown in Figure 2(b).

Note that an appropriate clinical triage (severely injured patient identified as such) can still result in system-related under-triage because a TC is too far away, and this patient, therefore, has to be taken to a local NTC. In that sense, the EMS decision around 'destination determination' is similar to a binary classification problem with four possible outcomes; true positive (severely injured patient is taken to a TC), true negative (less severely injured patient is taken to an NTC), false positive (a less severely patient is taken to a TC, leading to srOT), and false negative (a severely injured patient is taken to an NTC, leading to srUT). The srUT rate is then calculated as (1-sensitivity), while srOT rate is calculated as (1-specificity).

Both srUT and srOT have negative implications on patient safety. srUT can cause a delay in providing definitive care and increase the likelihood of an adverse outcome such as disability, morbidity, and even mortality (Rotondo et al., 2014). In contrast, srOT can cause overcrowding at emergency departments (Lerner, 2006), unnecessary trauma activation requiring trauma providers (physicians and nurses) to suspend their care of admitted trauma patients in an operating room and/or trauma inpatient unit to attend the arriving trauma patient (who does not have major trauma injuries), and loss of other salvageable lives in mass casualty trauma (Frykberg, 2002; Armstrong et al., 2008).

While the ACS has developed a guideline, Needs Based Assessment of Trauma Systems (NBATS), which suggests the number of TCs in a region using a score derived based on trauma providers' experiences, it does not suggest the locations of these TCs and cannot evaluate the impact of these TCs on srUT and srOT rates. A few studies have emerged that attempt to use optimization-based approaches (see Section 2), but they do not account for srOT and provide insights on the effect of changes in the system parameters (e.g., weights for srUT and srOT rates, required volume at TC, and thresholds used to mimic EMS decision making) on the optimal network of TCs.

2.1.4 Focus of this work

Our work contributes to this field by addressing the questions posed to us by our collaborating trauma decision-makers and researchers, but cannot be done so using existing approaches: (i) *What is the optimal network of TCs that minimizes the weighted sum of mistriages (i.e., srUT and srOT)? and* (ii) *How sensitive is the network to changes in system parameters?* To address these questions, we propose the Trauma Center Location Problem (TCLP) of determining the optimal number and locations of TCs in order to minimize the Weighted Sum of Mistriages (WSM) and present a bi-objective optimization model.

The key contributions of our approach are as follows.

- *First*, we present a bi-objective optimization model for TCLP that determines which hospitals, among candidate locations, should be TC or NTC such that the weighted sum of srUT and srOT rates are minimized. Essentially, our model optimizes the network's performance in terms of patient safety. This model extends the multi-facility and multi-customer location models by incorporating individual customer characteristics and individualized network-dependent allocation, along with multi-transportation modes.
- *Second*, the patient safety surrogates (i.e., srUT and srOT rates) are estimated based on actual incidences; incidences are typically used in the Trauma literature to estimate srUT and srOT as the population may not always be a good surrogate (Røislien et al.,

2018). This is done through our proposed high-fidelity modeling of the mistriages via a notional tasking algorithm that emulates the ‘destination determination’ part (subsequent to the injury assessment part) of the EMS decision making process. We consider a variety of factors including the network of TCs (x_j), thresholds (α and β; see Appendix A for details), the severity of the injury (S_i), and ambulance and helicopter parameters. We also use estimated driving times (using Google Distance Matrix API) and air times (using the Haversine formula) from the field to all the candidate hospital locations.

- *Third*, we propose a heuristic using binary particle swarm optimization to efficiently solve the proposed MIP model for the TCLP for real world instances. The complexity of the resulting mixed integer programming model limited the use of state-of-the-art optimization solvers for realistic problem size (1000s of cases in a network of over 150 hospitals).
- *Fourth*, we evaluate the sensitivity of our solutions to trauma volume, choice of weights that dictate the emphasis on srUT vs. srOT rates, and threshold values for srUT and srOT estimation. In so doing, we provide quantitative guidance to state trauma policy makers on appropriate choices of these parameters and their impact on patient safety across the state. For our experiments, we use a representative sample of 6,002 de-identified trauma patient data from 2012 available from the US state of OH. We illustrate the use of our approach through a case study based on the actual network of this US state where we derive a ‘greenfield’ design and also a ‘redistribution’ of the TCs existing in 2012.

Our findings suggest that there is a direct relationship between the number of TCs in the region and the corresponding srUT and srOT rates. That is, as the number of TCs increases, for severely injured patients the access to these TCs becomes easier, which can lower the srUT rate. However, a larger number of TCs in the vicinity can prompt the EMS providers to transport less severely injured patients to these TCs, leading to a higher srOT rate. While the number and location of TCs are sensitive to the choice of weights that dictate the contribution of srUT and srOT rates in the objective function, they are also sensitive to the volume requirements and the threshold values. The application of our approach on the real 2012 trauma network in OH demonstrated over 31.5% decrease in the weighted objective (51.8% decrease in srUT rate and 1% increase in srOT rate) with only one additional TC. Redistribution of the 21 TCs led to a 30.4% decrease in the weighted objective (46.6% decrease in srUT rate and 4.95% decrease in srOT rate). Essentially, our approach not only provides a benchmark to evaluate an existing trauma network in the state, but can also be used to redistribute TCs (within a region or the entire state) to unearth latent benefits in terms of patient safety.

The rest of the paper is organized as follows. Section 2 summarizes relevant literature, Section 3 presents the mathematical model that involves the estimation of srUT and srOT rates based on the mathematical programming-based formalization of a notional tasking algorithm (that approximates the EMS decision making process). Our proposed Binary PSO is detailed in Section 4 and insights based on our experimental study with a real dataset are presented in Section 5. Section 6 presents a case study where we use our approach to identify greenfield and redistribution networks for OH. Finally, Section 7 summarizes our work and offers guidance in future research in this area.

2.2 Literature Review

The literature on healthcare facility location is vast and includes locating long-term health care facilities (Cardoso et al., 2015), blood bank locations (Çetin & Sarul, 2009), organ-transplant centers (Caruso & Daniele, 2018), tuberculosis testing laboratories (Saveh-Shemshaki et al., 2012), and mobile healthcare units (Doerner et al., 2007). See Reuter-Oppermann et al. (2017), Ahmadi-Javid et al. (2017), and Gunes et al. (2019) for a comprehensive review of healthcare facility location models. These models vary in their objectives, may it be cost-based or patient safety-based. Several cost-based models have been proposed; e.g., location-allocation of organ-transplant centers (Zahiri et al., 2014), design of medical service (Shishebori & Babadi, 2015), and health centers for traumatic brain injury (Côté et al., 2007; Syam & Côté, 2010). Because the focus of our work is on patient safety, we now summarize key literature below.

Access to a facility has often been used as a surrogate for patient safety; for instance, (i) minimizing the total distance or time traveled across all constituents and (ii) maximizing the demand coverage within a fixed assess time. Objective (i) has been used to improve access to healthcare facilities (Cocking et al., 2012); e.g., optimizing the locations of organ transplant centers (Beliën et al., 2013), location and dispatching decisions for an ambulance system (Schmid, 2012; Toro-Díaz et al., 2013), and shelter location in humanitarian logistics (Bayram et al., 2015; Chen et al., 2013). Similarly, objective (ii) has been preferred in general healthcare facility planning (Kim & Kim, 2013; Shariff et al., 2012); e.g., optimizing the location of ambulances (Ingolfsson et al., 2008), distribution centers in a relief network (Balcik & Beamon, 2008), and emergency response facility during an earthquake (Salman & Yücel, 2015), as well as the relocation of ambulance stations (Cheng et al., 2011).

A few approaches have been proposed in the IE/OR literature for multi-facility and multi-customer problems. Marianov and Taborga (2001) presented a hierarchical p-covering type model to locate public health centers providing non-vital services in the presence of competing private centers to maximize low-income coverage. Yasenovskiy and Hodgson (2007) proposed a hierarchical location-allocation model that allows for by-passing to maximize patron's benefits. Teixeira and Antunes (2008) presented a hierarchical location model with two different types of assignment constraints: closest-assignment constraint and path-assignment constraint. Recently, Nasrabadi et. al. (2020) proposed a bi-hierarchy multi-service capacitated facility location-allocation problem with the bi-objective of minimizing total weighted travel time, and the fixed and operating cost of facilities. These studies, however, do not account for the time-sensitive nature of the assignment and only consider ground transport mode.

Patient safety has been an important criterion in trauma facility location literature. Branas et al., (2000) propose a linear programming model, namely the Trauma Resource Allocation Model for Ambulance and Hospitals (TRAMAH), to simultaneously locate trauma centers and air ambulance with an objective of maximizing coverage of severely injured patients using Maryland as a test region. TRAMAH, first of its kind, considers Rand-McNally TripMaker Version 1.0 to calculate the shortest driving time and Euclidean distance for air time and is solved using CPLEX Version 1.2. The model, however, uses a proxy for incident location, lacking geographical granularity and does not account for less severely injured patients. Cho et al. (2014) present a model that simultaneously locates trauma centers and medical helicopters with the objective of maximizing the expected number of patients transported to a TC within 60 minutes. The authors not only incorporate busy fraction for medical helicopters, but also develop the Shifting Quadratic Envelopes algorithm to optimize the problem. However, the model only considers severely injured patients (ISS>15), employed Euclidean distance between the demand region and each TC, and did not consider the aspect of mistriages that occur for both severely and less severely injured patients.

Jansen et al. (2015) propose a novel data-driven approach with a bi-objective of minimizing the total access time and the number of exceptions or system-related UT (srUT) for Scotland. The authors extend the model formulation in Handing et al. (2016) and solve the extended formulation with a multi-fidelity surrogate-management strategy via NSGA-II. They demonstrate the viability of their approach using real data from the state of Colorado's trauma system (Jansen et al., 2018). In contrast, the model is computationally complex requiring considerable processing time and also fails explicitly in considering the over-triage cases, an important factor of a patient-safety metric. The ACS Committee on Trauma suggested tool, Needs-Based Assessment of Trauma System (NBATS), helps determine the required number of TCs in a specified geographical region by allocating points based on population, transport time, community support, where are severely injured patients transported (TCs and NTCs), and the total number of TCs (ACS-NBATS, 2015). However, the tool does not determine the location of the TCs.

Our review of the above literature reveals the following gaps. *First*, the derivation of OT rates, based on injury score and its on-field operational decision-making process, has never been explicitly considered and accounted in the optimization models. *Second*, none of the prior approaches consider the fact that the determination of medically-appropriate time to access a suitable hospital (TC or NTC) varies by the type and volume of the injuries. For a severely injured patient, the proposed transportation times to the TC are as low as 30 and as high as 60 minutes (depending on the region/state), but for a less-severely injured patient, there is no such reasonable transport time to the NTC proposed in the literature. *Third*, the sensitivity of the 'access' and 'bypass' thresholds for a patient to reach their designated level of care, used for determining the srUT and srOT rates, has not been explored. *Finally*, we know of no literature that jointly considers the metrics of mistriages (i.e., srUT and srOT) to determine the optimal number and location of TCs.

To fill the gap as mentioned above, we propose a bi-objective trauma facility location optimization model to determine the optimal number and location of trauma centers with the aim of minimizing the weighted sum of mistriages. The key feature of our model is the inclusion of patient level decision-making related to destination selection, which is in turn based on patient's severity of injuries and estimated drive times to each candidate location (TC or NTC). Our proposed notional tasking algorithm helps to estimate the resulting srUT and srOT rates. Several practical insights are presented based on the sensitivity analysis conducted by varying minimum trauma case volume, weights of mistriages, and threshold values for the srUT and srOT rates. We now present our proposed model.

2.3 A Bi-Objective Model for TCLP

We define the Trauma Center Location Problem (TCLP) as determining the optimal location of TCs that minimizes the weighted sum of mistriages (srUT and srOT) in the entire trauma care network. The model assumes that a geographically defined area, typically known in the trauma literature as the Trauma Service Area (TSA), is known. This defined region could be a county, a region in the state, or the state itself.

Before we present the model, it is important to effectively capture the EMS decision making around destination determination. Based on the existing trauma literature (Jansen et al., 2018) and our discussions with EMS providers in our region, this process requires both clinical and resource considerations. To mimic this decision-making process, we propose two thresholds: (i) 'access' and (ii) 'bypass.' Here the 'access' threshold is a clinically-driven value that prespecifies the time to reach a hospital for a severely injured patient; this time is specified by the American College of Surgeons or state regulations. On the other hand, the 'bypass' threshold is a resource-driven value that prespecifies the maximum additional minutes (compared to a nearby TC) the EMS can dedicate towards a non-severely injured patient in order to transport them to an NTC.

Further, in line with the existing trauma literature, we use Injury Severity Score (ISS) as a surrogate for the severity of injuries on the field; ISS is a post-hoc metric evaluated after the patient arrives at the hospital. For a severely injured patient (ISS>15), the EMS providers often first check if a TC (the appropriate hospital) is accessible within the 'access' threshold time. If yes, then the patient is transported to that TC, resulting in a system-related appropriate triage positive ($srAT^{P}$). If no, then they check if an air ambulance can be called in to transport the patient to the nearest TC ($srAT^{P}$ via air). However, if the sum of the inbound-to-field, loading, and transport-to-TC times for the air ambulance is higher than the 'access' threshold, then the EMS would most likely transport the patient to a nearby NTC, resulting in a srUT.

Table 1. Confusion matrix

		Injury Severity Score (ISS)	
		ISS>15	*ISS≤15*
Destination	*To TC*	System-related Appropriate-triage ($srAT^P$)	System-related Over-triage (srOT)
	To NTC	System-related Under-triage (srUT)	System-related Appropriate-triage ($srAT^N$)

In contrast, the case of a srOT is a bit more complicated. A TC may be located close to the trauma incidence site compared to an NTC. In this case, if for a less severely injured patient (with ISS ≤ 15) the additional time (beyond the time to TC) to reach an NTC (the appropriate hospital for this patient) is within the 'bypass' threshold, then the EMS will likely take the patient to the NTC; this would be deemed as system-related appropriate triage negative ($srAT^N$). Otherwise, the EMS would likely take the patient to the nearby TC; this would be deemed as srOT. Anecdotally, such situations may arise due to EMS perception of a nearby TC's reputation to be higher (i.e., the bigger the hospital the better the care), patient/family choice, insurance situation, and even negotiated contracts between the EMS and TC.

Both srUT and srOT are estimated based on, and as indicated earlier, the EMS decision making process for 'destination determination' which is similar to a binary classification problem. Accordingly, we can generate a confusion matrix with $srAT^P$ (true positive), $srAT^N$ (true negative), srUT (type-1 error), or srOT (type-2 error); see Table 1. The notional tasking algorithm provides a means to classify each patient into these 4 categories in the confusion matrix; as explained above (see Appendix A for examples); corresponding analytical expressions are in the optimization model below. If there are multiple patients at the incidence site (say, during a multi-vehicle crash), then each patient will be evaluated individually (as suggested by the EMS providers and specified in the data).

The srUT rate is then calculated as (1-sensitivity), where the true positive value is the count of total system-related appropriate triages (via ground or air), and the false negative value or type-1 error is the total number of system-related under-triage cases for incidents with ISS>15 and for a given configuration. Similarly, the srOT rate is calculated as (1-specificity), where the true negative value is the count of total system-related appropriate triages, and the false positive value or type-2 error is the total number of system-related over-triage cases for incidents with ISS≤15 and for a given configuration. The two rates can be determined by srUT = 1 – sensitivity = $1 - \left(\frac{srAT^P}{srAT^P + srUT}\right)$ and srOT = 1 – specificity = $1 - \left(\frac{srAT^N}{srAT^N + srOT}\right)$ (Newgard et al., 2016).

Given this background, we now present our model under the following assumptions:

- The candidate locations for the TCs and NTCs are known and finite.
- Injury Severity Score (ISS) is used as a surrogate to estimate a patient's injury severity at the field.
- While ground ambulance services are available without constraints, the availability of air ambulance was restricted to 6.6% based on historical available data.
- In line with existing trauma literature, air ambulance is only allowed for severely injured patients

Tables 2 and 3 summarize the parameters and decision variables, respectively, used in our model.

Table 2. Parameters in the model

Notation	*Definition*
i	Index for trauma incidence (case); i = 1, 2, …, I
j	IndexFigure 29: Objective and # MTCs for different % of arOT in ISS 9-15 grouphreshold for srUT (in minutes)
β	'Bypass' time threshold for sFigure 13. TSA with counties, and region; grey filled areas are urban counties; '+'represents candidate locationsrOT in ISS 9-15 groupm its base to field (in minutes)
T_{load}	Loading time of a patient to an air ambulance (in minutes)
Z	Maximum allowable air ambulance use
V^{min}, V^{max}	Minimum and maximum volume of a severely-injured patient if TC is located at j
C	Maximum number of allowable TCs in the TSA
ω_1 and ω_2	Weights for the srUT and srOT terms in the objective function; $\omega_1 + \omega_2 = 1$
TG_{ij}	Travel time from the location of case i to any candidate location j via ground
TA_{ij}	Travel time from the location of case i to any candidate location j via air
SG_{ij}	$\{l \in J: TG_{ij} < TG_{il}\}$, $i \in I$, $j \in J$, that is the subset of candidate locations with higher time from case i than candidate location j via ground
SA_{ij}	$\{l \in J: TA_{ij} < TA_{il}\}$, $i \in I$, $j \in J$, that is the subset of candidate locations with higher time from case i than candidate location j via air
M	Big number

Table 3. Decision variables in the model

Notation	*Definition*
x_j	1, if a candidate location j is designated to be a TC; 0, otherwise
z_{ij}^1	1, if location j is marked as TC and is the nearest TC for case i via ground; 0, otherwise
z_{ij}^0	1, if location j is marked as NTC and is the nearest NTC for case i via ground; 0, otherwise
y_{ij}^1	1, if case i is transported to location j via ground transport; 0, otherwise (i.e., if j is a TC, then case i is $srAT^P$ and if j is an NTC, then case i is $srAT^N$)
y_{ij}^2	1, if a severely injured case i is transported to location j (that is marked as TC) via air transport; 0, otherwise (i.e., this case is considered $srAT^P$ via air)
y_{ij}^3	1, if case i is transported to location j that is marked as TC via ground transport; 0, otherwise (i.e., this case is considered *srUT*)

Minimize: $\omega_1 \left(1 - \frac{\sum_{i:S_i=1}\sum_j (y_{ij}^1 + y_{ij}^2)}{\sum_i S_i}\right) + \omega_2 \left(1 - \frac{\sum_{i:S_i=0}\sum_j y_{ij}^1}{\sum_i (1-S_i)}\right)$

Subject to:

Determining the nearest TC via ground

$z_{ij}^1 \leq x_j; \forall\, i \in I, \forall\, j \in J$ (1)

$\sum_j z_{ij}^1 = 1; \forall\, i \in I$ (2)

$x_j + \sum_{l \in SG_{ij}} z_{il}^1 \leq 1; \forall\, i \in I, \forall\, j \in J$ (3)

Determining the nearest NTC via ground

$z_{ij}^0 \leq (1 - x_j); \forall\, i \in I: S_i = 0, \forall\, j \in J$ (4)

$\sum_j z_{ij}^0 = 1; \forall\, i \in I: S_i = 0$ (5)

$(1 - x_j) + \sum_{l \in SG_{ij}} z_{il}^0 \leq 1; \forall\, i \in I: S_i = 0, \forall\, j \in J$ (6)

Each severely injured case is assigned to only one category (to TC via ground, to TC via air, or srUT)

$\sum_j (y_{ij}^1 + y_{ij}^2 + y_{ij}^3) = 1; \forall\, i \in I: S_i = 1$ (7)

Assign severely injured cases to nearest TC that is within 'access' time threshold via ground

$y_{ij}^1 = 0; \forall\, i \in I: S_i = 1, \forall\, j \in J, TG_{ij} > \alpha$ (8)

$y_{ij}^1 = z_{ij}^1; \forall\, i \in I: S_i = 1, \forall\, j \in J, TG_{ij} \leq \alpha$ (9)

Assign severely injured cases to nearest TC that is within 'access' time threshold via air

$y_{ij}^2 = 0; \forall\, i \in I: S_i = 1, \forall\, j \in J, TA_{ij} + T_{in} + T_{load} > \alpha$ (10)

$x_j + \sum_{l \in SA_{ij}} y_{il}^2 \leq 1; \forall\, i \in I: S_i = 1, \forall\, j \in J$ (11)

$\sum_i \sum_j y_{ij}^2 \leq Z$ (12)

Assign severely injured cases to nearest TC located outside 'access' time threshold via ground (transfer srUT cases to TC from NTC)

$y_{ij}^3 \leq z_{ij}^1; \forall\, i \in I: S_i = 1, \forall\, j \in J, TG_{ij} > \alpha$ (13)

Assign non-severely injured cases to nearest NTC if 'bypass' time criteria met

$\sum_j \left(z_{ij}^0\, TG_{ij}\right) - \sum_j \left(z_{ij}^1\, TG_{ij}\right) - \beta \leq M\left(1 - \sum_j y_{ij}^1\right); \forall\, i \in I: S_i = 0$ (14)

$y_{ij}^1 \leq z_{ij}^0; \forall\, i \in I: S_i = 0, \forall\, j \in J$ (15)

Allowable number of TCs, and their minimum and maximum volume

$\sum_j x_j \leq \mathrm{C}$ (16)

$\sum_{i:\,S_i=1}(y_{ij}^1 + y_{ij}^2 + y_{ij}^3) \leq x_j\, V^{max};\ \forall\, j\ \in\ J$ (17)

$\sum_{i:\,S_i=1}(y_{ij}^1 + y_{ij}^2 + y_{ij}^3) \geq x_j\, V^{min};\ \forall\, j\ \in\ J$ (18)

Bounds on decision variables

$x_j, z_{ij}^1, z_{ij}^0, y_{ij}^1, y_{ij}^2, y_{ij}^3\ \in \{0, 1\};\ \forall\, i\ \in\ I, \forall\, j\ \in\ J$ (19)

The objective of the TCLP is to minimize the weighted sum of total srUT and srOT rates (referred to as WSM from now on) for the TSA. In the above objective function, the first term in bracket represents srUT rate = 1 – sensitivity = $1 - \frac{appropriately\ triaged\ to\ a\ TC\ cases}{Cases\ with\ ISS>15} = 1 - \frac{\sum_{i:S_i=1}\sum_j (y_{ij}^1 + y_{ij}^2)}{\sum_i S_i}$ and the second term represents srOT rate = 1 – specificity = $1 - \frac{appropriately\ triaged\ to\ a\ NTC\ cases}{Cases\ with\ ISS\leq 15} = 1 - \frac{\sum_{i:S_i=0}\sum_j y_{ij}^1}{\sum_i (1-S_i)}$.

Note that a severely injured case *i* is classified as srUT if it is not accessible to any TC (within the ‘access’ time threshold) via air or ground ($\sum_j (y_{ij}^1 + y_{ij}^2)\ = 0$). On the other hand, a non-severely injured case *i* is classified as srOT if the difference between nearest NTC and TC time (via ground) exceeds the ‘bypass’ time threshold ($\sum_j y_{ij}^1\ = 0$); i.e., this case would likely be transported to a TC via ground as the NTC (correct hospital) is much further from the nearest TC(which mimics the practice among EMS). The values of srUT and srOT rates in the above objective function are weighted by multipliers ω_1 and ω_2, respectively.

Constraints (1)-(3) determine the nearest TC. While Constraints (1) ensure that candidate location *j* must be a TC to be considered as the nearest TC, Constraints (2) ensure that for every case *i*, only one TC should be considered as the nearest TC. For any pair of case *i* and candidate location *j*, if candidate location *j* is marked as TC, then Constraints (3) rule out the assignment of case *i* to candidate location(s) *l* that are located further (in terms of time) than *j* (the nearest TC for case *i*). Constraints (4)-(6) serve the same purpose as (1)-(3), respectively, for the nearest NTC via ground.

Constraints (7) ensure that each severely injured case is either assigned to a TC within the 'access' time threshold via air or ground (resulting in $srAT^P$) or transferred to a TC located outside of the 'access' time threshold after the case *i* has been stabilized at a nearby NTC (resulting in srUT). Constraints (8) rule out the assignment of severely injured cases to candidate locations that are not accessible within 'access' threshold via ground. Constraints (9) enforce the assignment to the nearest TC when a nearest TC exists within the 'access' threshold for a severely injured case *i*.

Constraints (7), (10), and (11) assign the remaining severely injured patients (unassigned via ground) to the TC via air if the total time to the TC is within the 'access' threshold. Constraints (10) rule out an assignment of severely injured cases to candidate locations that require total transport time more than the 'access' threshold via air. Constraints (11) rule out an assignment of severely injured cases to further located candidate location(s) via air if a given candidate location *j* marked as TC. Constraint (12) makes sure that the air transport usage does not exceed their availability; $Z = \lfloor \mu \sum_i SI_i \rfloor$, where μ = availability of air ambulance; $0 \le \mu \le 1$.

As mentioned earlier, every srUT case (originally transported from the scene to an NTC) is eventually transferred to a TC to receive definitive care. Constraints (13) capture such transferred srUT cases to the nearest TC (considered from the incidence location) for eventual volume estimation at this TC. Here we assume that the reason a severely injured patient is transported to a NTC (srUT case) is because there is no nearby TC (say TC-1) within 'access' threshold from the incidence. Our analysis of real data from a US midwestern state (see Section 5.1) indicated that the ratio of NTC to TC was 6.57 in 2012; i.e., there are a lot more NTCs than TCs. That is, there is a fairly low likelihood that another TC (say TC-2) is closer to the NTC than TC-1 (which was the closest from the incidence, but outside of the 'access' threshold). We use this low likelihood as the basis of assigning the srUT case to a TC that was closer to the incidence (i.e., TC-1, which was already identified as part of the constraints), instead of adding new constraints to locate a nearby TC from the NTC.

Further, an NTC is not designed to provide definitive care for severely injured patients. A srUT patient is only resuscitated/stabilized at an NTC before an eventual transfer to a TC. This would typically happen within 24 hours of arrival to the NTC. Because of that, NTCs do not have any capacity requirements associated with treating severely injured patients, and hence we do not need such constraints on NTC.

For each non-severely injured case *i*, Constraints (14) rule out the assignment of non-severely injured case *i* to an NTC if the 'bypass' threshold criterion is not met for that case; this case is marked as srOT. That is, it captures the situation when the nearest TC is located closer to the incidence site than the nearest NTC. Given that we have already categorized such a case as srOT, we do not need to explicitly assign srOT to a TC as they are not counted towards trauma volume; these are non-severely injured cases and are often discharged from the ED of a TC (without admission to the inpatient trauma unit). For non-severely injured cases where the 'bypass' time threshold is met, Constraints (15) assign those cases to the nearest NTC and mark them as $srAT^N$.

Constraint (16) ensures that the total number of TCs must be less than or equal to the maximum allowable TC. Constraints (17) and (18) calculate the volume of severely injured cases at each candidate location *j* that is designated as a TC location and ensures that the volume is within the allowable bound. As mentioned before, besides the $srAT^P$, srUT cases are also counted in TC volume. The

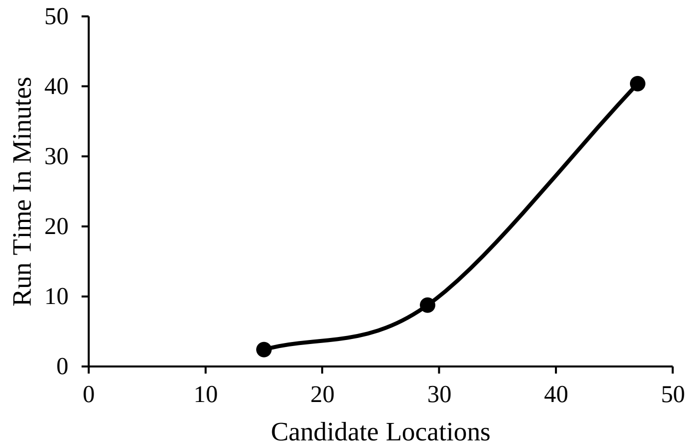


Figure 3. Number of candidate locations vs runtime using commercial solver

minimum bound essentially reflects the recommendations from the American College of Surgeons to ensure the financial viability of a TC; each TC must be able to offset the high cost of trauma readiness (physician, staff, equipment, and infrastructure).

Clearly, TCLP is a specific case of the discrete multi-facility location optimization problem with specific focus on patient-level safety measures. Such problem is combinatorial in nature and has been shown to be NP-hard (Daskin, 2013). TCLP exhibits the same characteristic where the decision to open TC or NTC at each of the *n* candidate locations. For *n*=159, this results in $2^{159} = 7.3\text{x}10^{47}$ solutions. A commercial solver such as CPLEX 12.10 and Gurobi were not able to obtain an optimal solution to our original problem due to the large problem size and resulting out-of-memory issues. We noticed that runtime increased exponentially with an increase in the number of candidate locations (x_j). For problem sizes with more than 47 candidate locations, we encountered out-of-memory issues with commercial solvers; see Figure 3.

Considering this limitation of solving TCLP exactly, we explored the use of a heuristic-based approach via Particle Swarm Optimization (PSO) to derive near-optimal solutions within a reasonable amount of time. We now discuss our proposed PSO algorithm.

2.4 Binary Particle Swarm Optimization

PSO is a nature-inspired population-based metaheuristic algorithm that optimizes continuous nonlinear function (Kennedy James & Eberhard Russell, 1995). The approach mimics the social behavior of birds flocking and fish schooling. It is easy to implement, makes fewer assumptions about the problem, is flexible and robust, and can be applied in a parallel manner. It has been implemented in a wide range of research areas such as facility location (Yapicioglu et al., 2007; Latha et al., 2013), network design (Chia-Feng Juang, 2004; Izquierdo et al., 2008), and scheduling (Liu et al., 2007; Liao et al., 2007).

The algorithm starts with a randomly distributed set of particles (potential solutions). With mathematical operators, the algorithm tries to progress to a solution with quality measure (fitness function). As the swarm of particles searches over time, they tend to fly towards better search regions, resulting in the convergence to a global optimum solution (Clerc & Kennedy, 2002). Each particle keeps track of its position which associates with the best solution it has achieved so far, known as particle best (*pbest*). On the other hand, global best (*gbest*) keeps track of the overall best value obtained thus far by any particle in the swarm.

For example, the *i*th particle is represented as $x_i = (x_{i1}, x_{i2}, \ldots, x_{id})$ in a d-dimensional search space. The previous best position of the *i*th particle is represented as $pbest_i = (pbest_{i1},$

$pbest_{i2}, \ldots, pbest_{id}$). The location of the best particle in the swarm is designated as $gbest$= ($gbest_1$, $gbest_2$, ..., $gbest_d$). The rate of position change (velocity) for the particle is represented as v_i= (v_{i1}, v_{i2}, ..., v_{id}). The velocity v_{id} and particle x_{id} used to update the d^{th} dimension of the i^{th} particle for the t^{th} iteration are given by:

$$x_{id}^{t} = x_{id}^{t-1} + v_{id}^{t}, \tag{20}$$

$$v_{id}^{t} = K(v_{id}^{t-1} + c_1 r_1 (pbest_{id} - x_{id}^{t-1}) + c_2 r_2 (gbest_d - x_{id}^{t-1})), \tag{21}$$

where c_1 and c_2 are acceleration constants; c_1 = 2.05 (Clerc & Kennedy, 2002), while c_2 is initially set to $c_1/5$ and gradually increase by 0.41 for every 25 iterations to allow particles to move slowly toward the global best solution. Further, r_1 and r_2 are two uniformly distributed random numbers in [0,1]. Constriction coefficient, K, aids in the convergence of the particle swarm algorithm; K=0.7298 (Clerc & Kennedy, 2002). The particle velocity given in equation (21) is composed of three primary parts: velocity from the previous iterations, cognitive or selfish influence (which uses the particle's personal best to improve the individual particle), and social influence (which represents alliance among the particle in the swarm using global best).

Recall that the decision variables in the TCLP are binary. We, therefore, use the binary version of the PSO, referred to as the BPSO (Kennedy & Eberhart, 1997). Accordingly, each particle represents its position in binary values, and the velocity of a particle is defined as the probability that might change the particle to either zero or one. The behavior and meaning of the velocity clamping and the inertia weight in the BPSO differ considerably from the real-valued PSO (Khanesar et al., 2007). However, the velocity update equation (21) remains unchanged, except that now the positions are binary and particle update equation (20) is replaced by:

$$if\ (rand() < S(v_{id})), then\ x_{id} = 1; else\ x_{id} = 0, \quad (22)$$

where function $S(v)$ is a sigmoid limiting transformation function, $S(v_{id}) = 1/(1 + e^{-v_{id}})$, and *rand*()~Uniform [0,1].

The likelihood of a change in a bit-value depends on $S(v)$. Furthermore, the probability that a bit will be 1 equals $S(v_{id})$, and that a bit will be 0 equals 1 - $S(v_{id})$ (Kennedy & Eberhart, 1997). The high-level structure of the PSO is as follows:

```
Initialize a population of particle with positions and velocities
 Do
     For each particle:
         Evaluate fitness function using the notional tasking algorithm
         Evaluate constraints
         If feasible:
            If the fitness value is better than the particle best:
               Set the current solution as particle best
            If the fitness value is better than the global best:
               Set the current solution as global best
         Else:
             Reject the solution
     End
     For each particle:
         Update the particle velocity
         Update the particle position
     End
 Until the termination criterion is met
```

In our proposed BPSO, we consider a swarm of 40 initial feasible particles, each representing a network of TCs, with the following representation: H = {0, 1, 0, 1, 1, 0, ..., 0, 1}; where 1 represents TC and 0 represents NTC, and $|H|$ represents the total number of existing hospitals. As the optimization model aims to minimize the objective function, the value given to an infeasible solution is set much higher. Hence, keeping them out of the loop. Equation (21) and (22) are applied to updating the velocity and particle, respectively.

We used *R* to implement our proposed BPSO and the notional tasking algorithm on a computer with 2 nodes, each node had 12 cores and each core had 2 threads (i.e., a total of 48 parallel processing options). Each core had a 3 gigahertz processor. The total RAM across all 12 cores was 256 GB. We also implemented parallel processing in *R* to allow for faster evaluation of each particle, which helped reduce the computation time to about 8 hours. Preliminary experiments suggested that 40 particles sufficiently balanced solution quality and solution time. Further, we implemented a dynamic change in the value of acceleration constant c2, which gradually increased the attraction to the global best (compared to personal best). This allowed the particles sufficient time to explore the search space around their personal best instead of speedy attraction to the global best position. We used two termination criteria based on preliminary experiments: maximum iterations (set to 1,000) and less than 0.1% improvement in the global best solution within 100 iterations.

2.5 Experimental Setting

To generate managerial insights, we conducted a series of experiments using a sample dataset made available for Ohio state by the Ohio Department of Public Safety (ODPS). The Wright State University's Institutional Review Board approved this study as not involving human subjects (IRB#06027). We first summarize the characteristics of this dataset and then present our insights.

2.5.1 Data collection

We considered the 2012 network of hospital locations (TCs and NTCs) made available to us by the ODPS. The 2012 data corresponded to a network of a total of 159 hospital sites; 21 TCs and 138 NTCs. We were able to obtain the (latitude, longitude) information for each of these sites. We were also able to derive a sample of 6,002 de-identified trauma incidences from the data provided by the ODPS for that year. This sample was about 1/11th of the typical number of trauma incidences occurring in the state; 67,542 cases in 2018 (Ohio Department of Public Safety, 2019); the correlation of a number of cases in each county between them was 0.84 suggesting that the 2012 data sample is a good enough representation of the spatial distribution of incidences. Figure 4 illustrates the heat map of 6,002 incidents, and the location of TCs and NTCs in 2012.

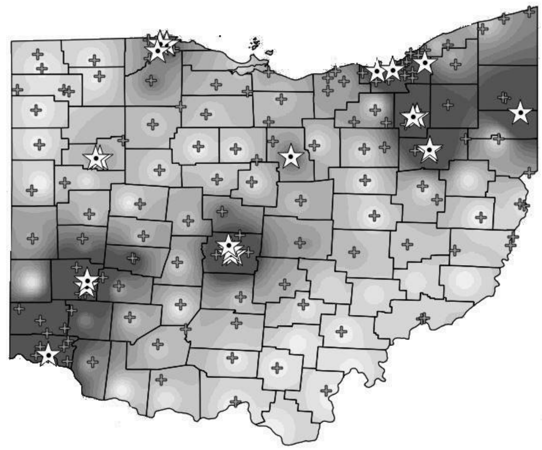


Figure 4. Trauma Care in OH for 8 regions; stars indicate TCs and crosses indicate NTCs. Darker shades of grey indicate higher values of incidences

We used the Google Distance Matrix API to estimate drive time based on the quickest route from each incident to a hospital site. We used the Haversine formula for the corresponding air travel time (assuming the helicopter speed of 120 mph). The resulting time matrix for ground and air (each 159×6002 in size) served as a look-up table to the notional tasking algorithm in order to estimate srUT and srOT rates. Because helipad locations were not available, we let the time from helicopter depot to the field be 10 minutes; in a similar vein, the loading of the patient was set to 5 minutes. Aggregating these two with the airtime from field to the nearest TC calculated using the Haversine formula, we estimated the total air transport time.

Table 4. Summary of the parameters, levels, and values in the sensitivity analysis

Parameter	Levels	Values
Weights (ω_1, ω_2)	5	(1,0), (0.8,0.2), (0.6,0.4), (0.4,0.6), (0.2,0.8)
V^{min}	4	0, 11, 22, 33
Access threshold (α)	3	15, 30, 45 minutes
Bypass threshold (β)	3	-10, 0, 10 minutes

2.5.2 Experimental study and insights

Based on preliminary experiments, we noticed that the solutions were sensitive to four key factors, which are summarized in Table 4 along with their levels and values. Further, the American College of Surgeons recommends having at least 240 trauma cases per year for a TC to be viable; i.e., V^{min} = 240 cases with severe injuries estimated as ISS>15 (Rotondo et al., 2014). Hence, to correspond with the sample of 6,002, we scaled V^{min} to 22 (=$\lceil 240/11 \rceil$).

We set our 'base case' with V^{min}=22 and α=45, and β=0 to mimic the current protocols in most states in the US. We set (ω_1, ω_2) = (0.8, 0.2) to allow for more focus on patient safety; again, attempting to mimic how state governments try to locate TCs. We set V^{max} = 91 (equivalent to 1,000 cases) as the upper bound on a TC volume and C = 159 in all our experiments. Maximum air ambulance usage for severely injured patients was bounded; i.e., Z= 61 to match with sample 2012 data. Below we summarize the results and insights from the sensitivity analysis.

Insight 1: A higher emphasis on reducing the srUT rate means a corresponding increase in the number of TCs, but this can lead to a higher srOT rate.

The selection of the weights plays a vital role in determining the optimal number and location of TCs. We varied both weights (ω_1, ω_2) between 0 and 1 in steps of 0.2 such that $\omega_1 + \omega_2 = 1$. Note that when $\omega_1 >> \omega_2$, the emphasis is towards reducing the srUT rate (likely resulting in more TCs); while for $\omega_1 << \omega_2$, the emphasis is towards reducing the srOT rate (likely resulting in less TCs).

Figure 5 represents the trend in srUT and srOT rates, and WSM value over the weights. The figure shows that as ω_1 decreased the srUT rate increased and as ω_2 increased the srOT rate decreased, resulting in a drop in the number of TCs. Although a solution with (1.0, 0.0) may be attractive in terms of the lowest WSM, it comes at a cost. First, the corresponding network has the highest number of TCs, which puts a financial burden on the state and the hospital system. Second, a higher corresponding srOT rate (0.14) means

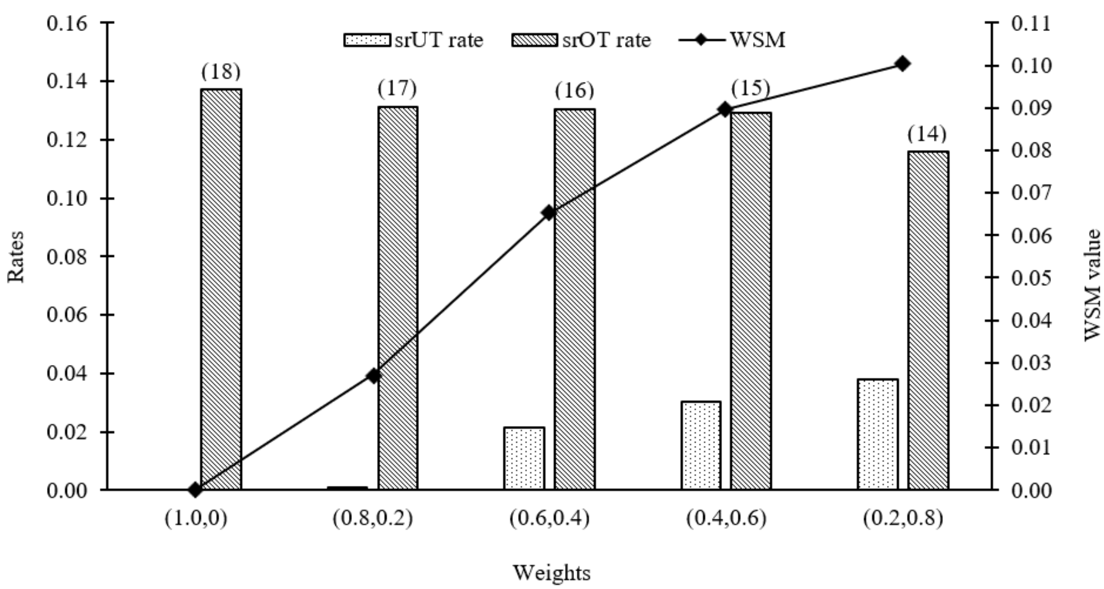


Figure 5. Representation of the srUT rate, srOT rate and WSM value over the weights; TCs in parenthesis for each column

a higher number of less severely injured patients at a TC, which is much more expensive than treating such patients at an NTC. Because such costs are difficult to estimate, we expect that this analysis will allow the trauma decision makers to make an informed judgement on the most appropriate network suitable for their region. From what we have learnt first-hand from trauma network designers, srUT is given a much higher emphasis compared to srOT. We would expect trauma decision makers to use our tools and start with a high ω_1 and then gradually lower it until a tolerable srUT is achieved to effectively trade-off srUT and srOT.

Insight 2: An increase in the minimum required volume of severely injured patients at a TC reduces the number of TCs in the network, but substantially increases srUT.

We varied V^{min} between 0 and 33 in increments of 11 to evaluate the impact of the minimum trauma volume on the TC network.; As mentioned earlier, the 240 cases (22 in our scaled down version) is a suggestion by the ACS based on empirical evidences, and,

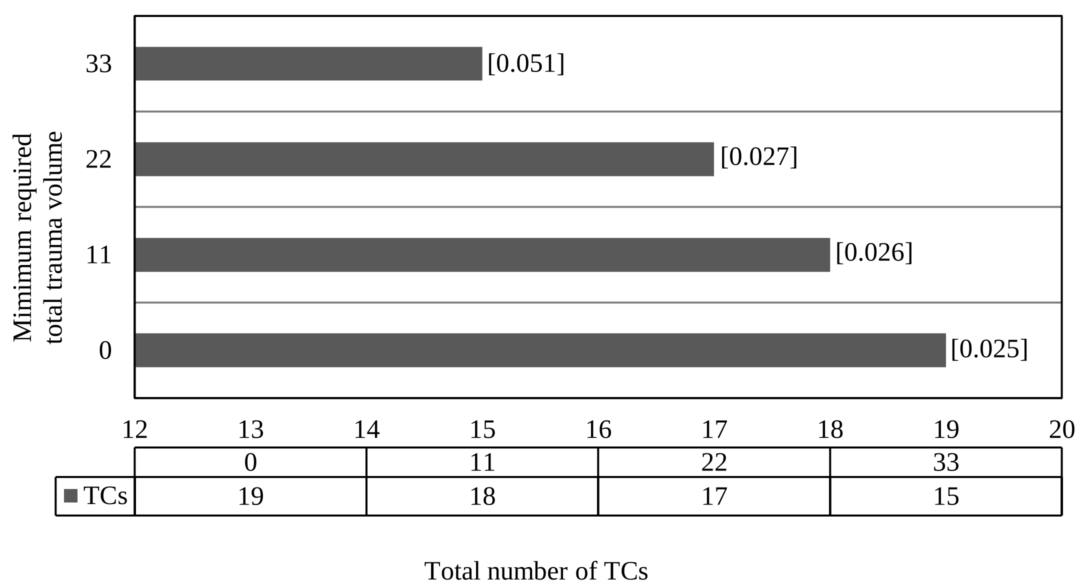


Figure 6. Representation of V^{min} against the total number of TCs; WSM value in the []

therefore, this sensitivity analysis provided a much-needed quantitative evaluation of the impact of changes in this value on the TC network and resulting srUT and srOT rates. Our results suggest that as minimum total trauma volume requirement at TC increased, the WSM value also increased. For a smaller value of the V^{min}, the network tends to have more TCs in order to minimize the srUT rate; recall, we used ω_1=0.8 for srUT (base case). This is intuitive as an increase in the number of TCs would likely allow more severely injured patients to reach a TC, which results in a decrease in the srUT rate. However, it also means that less severely injured patients may now be transported to a TC (as there is likely a TC as close to the field as an NTC) resulting in an increase in the srOT rate. However, as the V^{min} increased, the number of TCs decreased in order to satisfy the V^{min} constraint. As a result, the srUT rate and the WSM value both increased. Figure 6 illustrates this trend.

Essentially, a lower volume requirement can result in higher TCs and better patient safety. The implication of this is that the trauma decision-maker must appropriately set the minimum volume requirement as a TC with a low volume may not be financially viable.

Insight 3: An increase in the 'access' threshold reduces the number of TCs.

For this analysis, we considered the 'access' threshold (α) at 15, 30, and 45 minutes and a constant 'bypass' threshold of 0 minutes. Figure 7 illustrates the trend in the srUT and srOT rates, the WSM, and the number of TCs. Note that as the 'access' threshold (α) increased, the WSM (objective function) decreased. This is intuitive as, for the same network, an increase in α would mean that there is more allowable time for the EMS to transport a severely injured patient to a TC further away from the field (as compared to lower values of α). This means that the corresponding network will need fewer TCs to

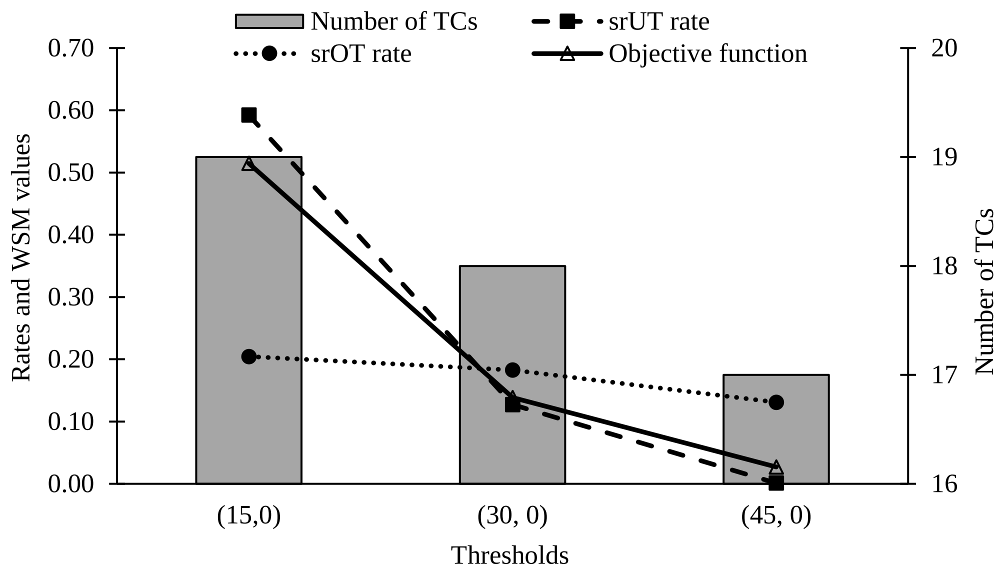


Figure 7. Representation of trend in srUT rate, srOT rate, objective function, and number of TCs

achieve lower levels of srUT rate. Fewer TCs also means a lower srOT rate. As both srUT and srOT rates decrease, the WSM will also experience a drop with an increase in α.

Insight 4: An increase in the 'bypass' threshold has a slight impact on the number of TCs.

For this analysis, we considered the 'bypass' threshold (β) at -10, 0, and 10 minutes and a constant 'access' threshold of 45 minutes. Table 5 summarizes the corresponding number of TCs and the resulting srUT, srOT, and WSM. Notice the increase in the number of TCs is only marginal. The reason is that as the 'bypass' threshold increases, the EMS providers now have more opportunities to skip the nearby TC and reach the appropriate NTC for a less severely injured patient. This, in turn, means that even if the number of TCs increases marginally (as seen in Table 5), the NTCs are still reachable from the incidence location, resulting in a reduction in srOT. Note that because of higher access threshold and

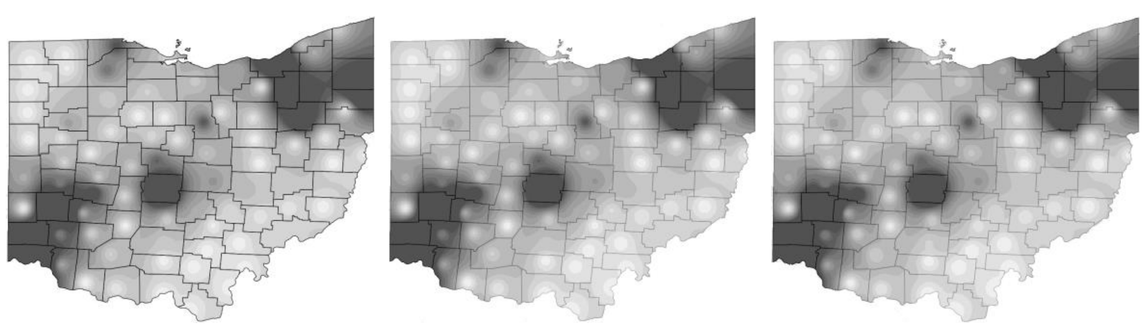
Table 5. Impact of 'bypass' threshold on srOT and the number of TCs

Thresholds	#TC	srOT	srUT	WSM
(45, -10)	16	0.492	0.002	0.100
(45, 0)	17	0.131	0.001	0.027
(45, 10)	18	0.018	0.000	0.004

reasonable number of TCs, srUT rates are fairly low and their effect on the WSM is negligible. Any further increase from 18 TCs, for the case of (45, 10), led to an increase in the srOT rate (as more TCs means an increased likelihood of srOT cases) causing WSM to increase.

2.5.3 Performance of the derived network with respect to unseen demand

As mentioned earlier, there is a significant cost associated with upgrading an NTC to a TC. Consequently, it is important to evaluate if the PSO-derived, near-optimal, TC network based on historical data would perform reasonably well with respect to the unseen, future demand. To do this, we used the Train-Test approach. Accordingly, we apportioned the 2012 data (6,002 cases) into Train (4,002) and Test (2,000) datasets, approximately a 2/3:1/3 split. To ensure that the spatial distribution of trauma incidences in each of these



Full Data (6002 Incidences) Train Data (4002 Incidences) Test Data (2000 Incidences)

Figure 8. Heatmap of incidences (darker area indicate higher values of incidences)

two datasets is similar to the Full dataset, we conducted the apportionment at the county level. The GIS-generated heatmaps in Figure 8 indicate that the apportionment was reasonable. We ran our proposed solution approach on the 'base case' (i.e., access threshold = 45 min, bypass threshold = 0 min) and adjusted other parameters corresponding to the reduced train and test sets.

The following are the key observations from this analysis; see Table 6 for a summary:

- The number of TCs found through the Train data set is identical to that found when using the full data set; WSMs are also similar.
- Test WSM on the same network (obtained through Train data) is very similar to Train WSM.

This analysis provides evidence that a near-optimal network obtained using 2012 (Full) data will work reasonably well with respect to unseen, future demand.

Table 6. Comparison of performance of network for full data and train data, and performance of test data for the network obtained through train data

	Full Data (6002 incidences)	**Train Data (4002 incidences)**	**Test Data (2000 incidences)**
WSM	0.0271	0.0291	0.0299
TCs	17	17	-

2.6 Case Study Based on OH's Trauma Network

We now illustrate how we used our proposed approach using the 2012 data from OH to derive (i) an optimal network (greenfield problem) and (ii) an optimal redistribution of existing TCs within that network (redistribution problem). Due to limited data fields in this data set, we used the overall UT and OT rates (that could also include clinical mistriages) for this case study; these were UT rate = 0.20 and OT rate = 0.515 (which we use as srUT and srOT rate, respectively, in our discussion below). Because we observed variations in EMS practice on ‘access (α)’ and ‘bypass (β)’ thresholds in the state, and in order to conduct a fair comparison, we treated both thresholds as meta-parameters that encompass the existing variations in EMS-practice when it came to ‘destination determination.’ Subsequently, we empirically derived α = 30 minutes and β = -9 minutes ensuring that the resulting performance of the network met the 2012 srUT and srOT rates (0.20 and 0.515, respectively). Note that due to limited data fields, it was difficult for us to tease out the clinical mistriages; we, therefore, used these values of 0.2 and 0.515 as surrogate estimates for srUT and srOT rates, respectively.

Our analysis of the 2012 trauma network is shown in Figure 4, which shows the distribution of the 21 TCs in the state. These TCs are generally located in areas with higher population density, resulting in a clustered pattern (also alluded in Brown et al., 2016); the resulting WSM at ω_1=0.8 and ω_2=0.2 was 0.270. Not surprisingly, Regions 7 and 8 with no TCs experienced the highest srUT rate (=1.00) and a zero srOT rate; in contrast, Regions 2 and 5 yielded a much lower srUT rate (0.078 and 0.084), but higher srOT rates of 0.527 and 0.772, respectively. On the other hand, Region 1 with 5 TCs still produced an unusually high srUT rate of 0.47, largely because of the clustering of 3 out of 5 TCs in a single urban area (Toledo), which result in high access times for incidences that occur outside of Toledo.

2.6.1 Greenfield design of OH's trauma network

Table 7. Comparison of 2012 network and optimized greenfield network

Region	# of TCs		srUT rate		srOT rate	
	2012 allocation	*TCLP allocation*	2012 allocation	*TCLP allocation*	2012 allocation	*TCLP allocation*
1	5	*2*	0.470	*0.313*	0.410	*0.252*
2	3	*3*	0.078	*0.000*	0.527	*0.525*
3	2	*4*	0.227	*0.061*	0.553	*0.668*
4	4	*4*	0.184	*0.143*	0.576	*0.588*
5	6	*5*	0.084	*0.036*	0.772	*0.515*
6	1	*3*	0.174	*0.062*	0.302	*0.553*
7	0	*0*	1.000	*0.786*	0.000	*0.229*
8	0	*1*	1.000	*0.278*	0.000	*0.324*
Overall	**21**	***22***	**0.206**	***0.099***	**0.525**	***0.530***

To optimize the network, we used identical system parameters: (ω_1, ω_2)=(0.8, 0.2), α=30, β=-9, V^{max} = 91, C=159; we set V^{min}=22 to meet the ACS guidelines. The best solution obtained by BPSO (with 40 particles) resulted in 22 TCs with WSM=0.185 (a 31.5% decrease from the 2012 estimate of 0.270). This optimized network reduced the srUT rate by 51.9% (i.e., 0.099 vs. 0.206 in 2012), with srOT rate increased by around 1% (i.e., 0.530 vs. 0.525 in 2012). Evaluation of the results depicted a rather dispersed pattern of TCs across the state (see Table 7). Specifically, Regions 7 and 8 (with TC in Region 8 near to the boundary of the Region 7) now experienced a lower srUT rate of 0.786 and 0.278, respectively. But the counter effect is that because of a TC in the region or near to the boundary of the region, the srOT rates increased in both Regions 7 and 8 (i.e., 0.229 and 0.324, respectively). Alternatively, a reduction from 5 TCs to 2 TCs in Region 1 resulted in the srUT rate dropping to 0.313 (compared to 0.47 in 2012) with a significant decrease in the srOT rate (0.252 compared to 0.41 in 2012). That is, while the state of OH may have a nearly optimal number of TCs, their suboptimal distribution leads to high WSM.

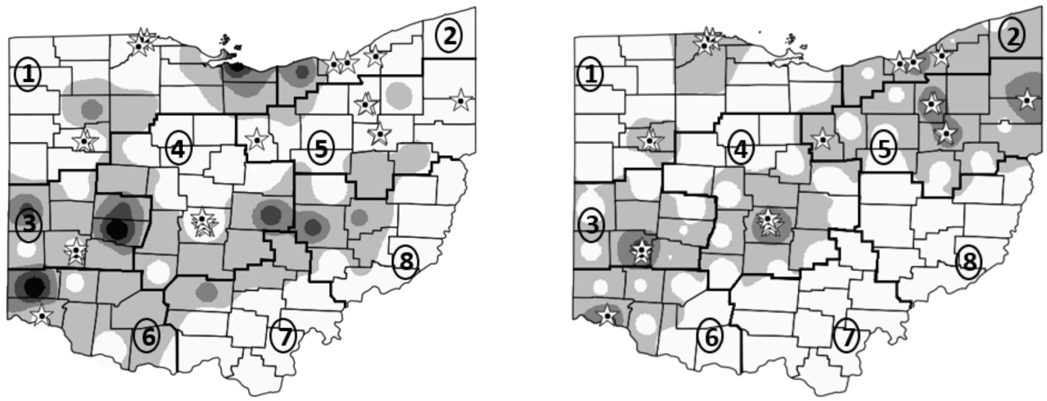
2.6.2 Redistribution of 21 TCs in OH

Table 8. Comparison of 2012 network and redistributed network

Region	**# of TCs**		**srUT rate**		**srOT rate**	
	2012 allocation	*TCLP allocation*	2012 allocation	*TCLP allocation*	2012 allocation	*TCLP allocation*
1	5	*2*	0.470	*0.253*	0.410	*0.223*
2	3	*4*	0.078	*0.022*	0.527	*0.341*
3	2	*4*	0.227	*0.055*	0.553	*0.668*
4	4	*2*	0.184	*0.156*	0.576	*0.540*
5	6	*5*	0.084	*0.080*	0.772	*0.504*
6	1	*3*	0.174	*0.056*	0.302	*0.553*
7	0	*0*	1.000	*0.786*	0.000	*0.229*
8	0	*1*	1.000	*0.389*	0.000	*0.324*
Overall	21	*21*	0.206	*0.110*	0.525	*0.499*

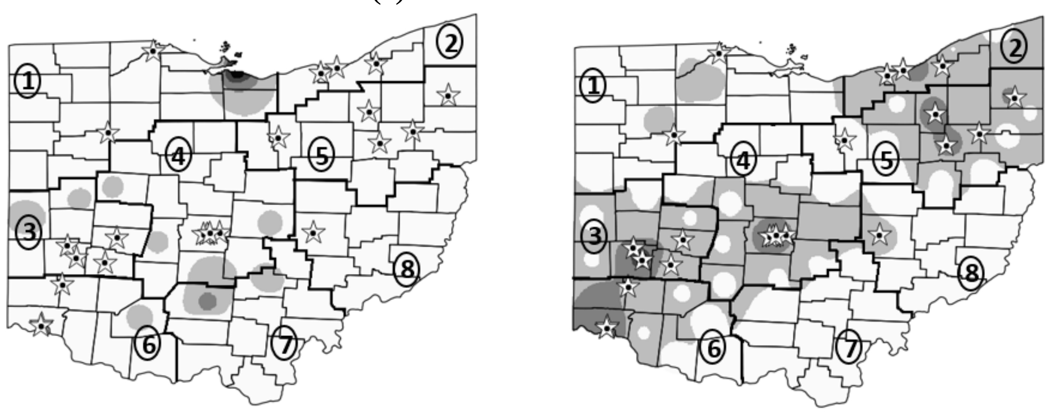
If a 'greenfield' design may not be possible, then could a redistribution of the 21 TCs within the state reduce the mistriages rate? To answer this question, we set C=21 in Constraint (16) of the TCLP model and kept the rest of the parameters identical to Section 6.1. Figure 9 illustrates the differences in the heat maps for srUT and srOT across the 3 networks (2012, greenfield, and redistributed).

The results were quite interesting; the 21 TCs were distributed quite differently across the state (see Table 8 region-wise comparison). This redistribution likely allowed more trauma patients to access a TC within the 'access' threshold (via ground or air). This is evident from a substantial drop in the srUT rate (by 46.6% to 0.11); the srOT rate also decreased by 4.95% to 0.499; WSM reduced to 0.188 compared to 0.270 (a 30.4% decrease).



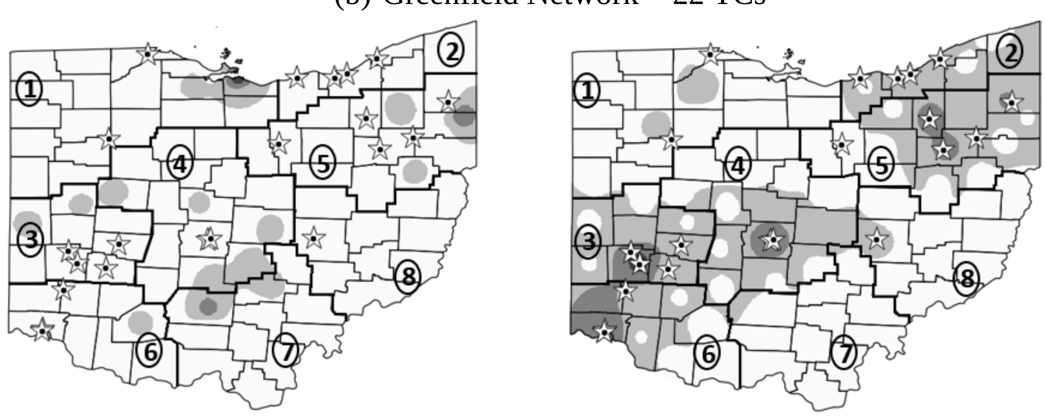
srUT = 0.206 srOT = 0.525

(a) 2012 Network = 21 TCs



srUT = 0.099 srOT = 0.530

(b) Greenfield Network = 22 TCs



srUT = 0.110 srOT = 0.499

(c) Redistributed Network = 21 TCs

Figure 9. Heat maps of mistriages.

(Note: Darker shades indicates higher values; Stars represents TCs)

The above two illustrations of our approach (i.e., greenfield and redistribution) using an actual state-wide network of all hospitals (TCs and NTCs) not only demonstrate that opportunities exist in the state to substantially improve patient safety, but also that our proposed approach is able to unearth those by specifying better networks.

2.7 Summary

Timely access of severely injured trauma victims to trauma centers can improve survival by 25%. Given the limitations of existing approaches in locating trauma facilities to address patient safety, we proposed the Trauma Center Location Problem (TCLP). The TCLP is to determine the optimal number and location of TCs in order to minimize the weighted sum of mistriages (srUT and srOT). This problem is an extension of multi-facility and multi-customer location models, which incorporates individual customer characteristics and individualized network-dependent allocation, along with multi-transportation modes.

We introduced an optimization model for TCLP that explicitly models patient safety via srUT and srOT rates, both estimated using our proposed notional tasking algorithm based on the standing guidelines in the trauma literature. To efficiently solve the resulting model, we proposed a Binary Particle Swarm Optimization (BPSO) approach and illustrated its use on 2012 data for the state of Ohio. The Train-Test approach provided further validity to our approach.

The key insights from our study include the following:

- While an increase in the number of TCs can reduce srUT, it can increase srOT; setting an appropriate emphasis on their reduction (via weights) in the objective function is critical.
- There is an inverse relationship between TC volume requirement and the number of TCs in the network. As a minimum volume requirement increases, some of the TCs need to downgrade to NTC's because of infeasibility due to low volume. A downgrade of TC increases the srUT rate that eventually decreases patient safety.
- While requiring EMS to transport severely-injured patients to the nearest TC is desirable (reflected by a lower ‘access’ threshold), this can only be achieved through an increase in the number of TCs in the network (with the corresponding effect indicated in (i) above).
- The illustration of our approach using real data from OH suggested that state has the nearly optimal solution in terms of the number of TC but significantly suboptimal objective value. A state can achieve up to 51.9% reduction in srUT at almost the same srOT rates can be realized with 1 additional TC; redistributing the same 21 TCs can still achieve the high reduction in srUT (46.6%) along with a 4.95% reduction in srOT.

We believe our proposed approach is effective and efficient in helping state trauma decision makers not only evaluate their current system, but also optimize it (either as a greenfield or redistribution problem). They can also conduct ‘what-if’ analysis by fixing certain TCs in their current locations and allowing the optimization approach to find the locations of other TCs in the state. This latter approach can be of particular interest to those states where a mass reallocation of TCs is not possible; instead, they are seeking a gradual change over a period of time, or evaluating the viability of a proposal by a healthcare system to upgrade an NTC to a TC or downgrade an existing TC to an NTC.

CHAPTER 3

NESTED TRAUMA NETWORK DESIGN CONSIDERING EQUITY AND EFFECTIVENESS IN PATIENT SAFETY

3.1 Introduction

In the US, trauma is the leading cause of death for individuals aged 44 and under (#3 across all ages), resulting in almost 200,000 deaths and an economic burden of over $670 billion annually (ACS, 2016; CDC, 2022). Trauma is a serious public health problem with significant social and economic costs. A trauma care system in a state (or a region within a state) is often established in an attempt to provide prompt and definitive care to trauma patients. Timely access to a trauma center (TC) is one of the key determinants of patient outcomes (Branas et al., 2013; Jansen et al., 2015).

3.1.1 Types of trauma centers

The American College of Surgeons (ACS) verifies TCs as Levels I-V based on the presence of the type of trauma resources and their availability (American Trauma Society, 2022). ACS-verified Levels I and II are referred to as major trauma centers (MTCs) and capable of providing definitive care for patients suffering from major traumatic injuries (i.e., severely injured patients). MTCs are equipped with highly sophisticated surgical and diagnostic equipment, with 24/7 surgeon availability, to provide high-quality medical and nursing care. While timely access to a LI (MTC) improves survival of severely injured patients by 25% relative to care delivered at a non-trauma center (MacKenzie et al., 2006).

According to the Centers for Disease Control and Prevention, “there is no access to an advanced trauma center for nearly 45 million Americans within the golden hour (60 minutes)” (ACS, 2016). The reason for this is the geographic maldistribution of MTCs in the U.S.; in 2010, 9 states had a clustered pattern, 22 had a dispersed pattern, and 10 had a random pattern (Brown et al., 2016). Further, there is a significant cost associated with building and operating MTCs, and it can be financially challenging to open an MTC in rural areas due to concerns of sufficient patient volume.

To circumvent this problem, Levels III-V TCs are set up to serve as feeder centers to MTCs for communities that do not have timely access to MTCs; we refer to such TCs as intermediate TCs (ITCs). ITCs provide a subset of services offered by LI/LII MTCs, but only during part of the day, and serve as centers for initial care, resuscitation, and subsequent transfer to major trauma centers (MTCs). It has been shown that an inclusion of ITCs in underserved counties decreases trauma-related mortality rates due to improved survival of transferred severely injured patients after stabilizing at those ITCs (Barringer et al., 2006; Tinkoff et al., 2007). After stabilizing, a patient is eventually transferred to an MTC as ITCs are not capable of providing definitive care to severely injured patients. All other hospitals are referred to as non-trauma centers (NTCs), which are the ideal destination for non-severely injured trauma patients.

3.1.2 On-field decisions and trauma triage

The majority of trauma deaths occur in the pre-hospital environment or within 4 hours of the trauma event (ACEP, 1987). The pre-hospital trauma triage is designed to transport the right patient to the right hospital at the right time. The emergency medical service (EMS) is crucial in providing initial care to the injured patient and accurate pre-hospital triage. EMS providers' on-field decision making practice involves two components; (i) injury assessment (how severe the injuries are) and (ii) destination determination (which hospital to select and how to transport). An error in making any of these decisions can lead to pre-hospital mistriage.

Note that, besides mortality, mistriage has been used in the trauma literature as a surrogate for patient safety as it often increases the risk of short/long disability caused due to delay in provision of definitive care (Jansen et al., 2015; Hirpara et al., 2022, Parikh et al., 2022). Consequently, considering (i), an error in accurately assessing the injury type (severe or non-severe) can lead to 'clinical mistriage.' Similarly, for (ii), an error in determining the most suitable hospital type (trauma center or not) can lead to 'system-related mistriage.'

3.1.3 System-related mistriages

We define three types of system-related mistriages (as surrogates for patient safety). A situation when a severely injured patient is taken to an NTC because of a lack of access to an MTC experiences is referred to as 'system-related under-triage (srUT).' Further, in a trauma network with MTCs and ITCs, if a severely injured patient, who ideally should be transported to an MTC, is first transported to an ITC due to lack of access to MTC, then we refer to that as ‘system-related under-triage stabilized (srUT^{s}).’ We use the modifier ‘stabilized’ because an ITC has the ability to provide prompt assessment, resuscitation, limited surgery, intensive care and stabilization of injured patients and emergency operations, compared to an NTC (in which case we would have referred this patient as srUT). In contrast, an excess (or cluster per Brown et al., 2016) of MTCs and ITCs in the vicinity of an incidence location (also known as scene) could induce EMS to transport a less severely injured patient to such hospitals, which we refer to as ‘system-related over triage (srOT).’

Generally, srUT (and srUT^{S}) and srOT have negative implications on patient safety. A srUT increases the likelihood of an adverse outcome such as disability, morbidity, and even mortality due to delay in receiving definitive care (Rotondo et al., 2014). In contrast, a srOT indirectly impacts patient safety by causing overcrowding at emergency departments (Lerner, 2006), unnecessary trauma activation resulting in additional charges to the patient, and loss of salvageable lives in mass casualty trauma (Frykberg, 2002; Armstrong et al., 2008).

3.1.4 Trauma network’s influence on destination determination

It is during the destination determination phase when the network of MTCs and ITCs is critical. Table 9 shows three destination determination criteria used by EMS providers at the incidence location, the decision makers, and how the network of MTC/ITC impacts the corresponding decision. Clearly, the network of MTC/ITC influences the

Table 9. Influence of the trauma network on different destination determination criteria

Dest. det. criteria	**Decision maker**	**Influence of the MTC/ITC network**
Protocol	EMS paramedics	For severe injuries, take to an MTC (ideally) or ITC (if no MTC available)
Patient choice	Patient or family of a patient	Choices tend to favor MTC or ITC based on perception of different hospital types in the vicinity, past experience, and access time
Closest facility	EMS paramedics	Take to nearest hospital (even if NTC) during extreme weather condition or road closure

selection of an appropriate hospital for prompt and definitive care, eventually reducing mistriages and improving patient care.

Although trauma literature alludes to the importance of network of MTC and ITC and implications on mistriages (a key patient safety metric), there key questions are yet to be addressed, which form the basis of our research.

3.1.5 Focus of this work

This paper focus on the strategic decision of jointly determining the number and location of MTCs and ITCs to improve patient safety. We address the following questions:

- How do ITCs support patient safety?
- What effect does destination determination criteria have on the MTC/ITC network?
- How sensitive is the MTC/ITC network to the distribution of trauma patients?
- What is the impact of focusing on equity of patient safety on the trauma network's performance?

The key contributions of our research are as follows. First, we propose a Nested Trauma Network Design Problem (NTNDP), which is a nested multi-level, multi-customer, multi-choice, multi-transportation capacitated model with a bi-objective of maximizing equity and effectiveness in patient safety. Multi-choice refers to the inclusion of all 3 dominant criteria for destination determination (see Table 9). While 'equity' quantifies the level of similarity in patient safety across regions in a geographical area (portion of a state or the state), 'effectiveness' quantifies overall patient safety (see Section 3.1 for details). Second, we propose a three-step approach to efficiently solve the proposed MIP model. This approach is able to find a near-optimal solution in a reasonable amount of time for instances of realistic problem sizes. Finally, to test our approach, we generate several datasets with different distributions of trauma patients using information available from the trauma system of Ohio, a midwestern US state. We also evaluate the sensitivity of the solution to variations in proportion attributed to the 3 destination determination criteria, weights associated with equity and effectiveness, and different distributions of patients. Finally, we illustrate the use of our approach for real data from a midwestern US state (i.e., the state of Ohio).

Our experiments suggest that destination determination criteria impact a trauma system's design and performance. While ACS and many state trauma agencies recommend using 'protocol' as the primary destination determination criteria, increased use of 'patient choice' criteria (often practiced in reality) results in more ITCs in suburban and rural zones; the corresponding mistriages are also high. Further, for the same number of patients, dispersed distribution of patients results in a 21.8% decrease in the trauma network performance (i.e., causes high mistriages) even with almost 3 times of ITCs in the network compared to cluster distribution. Further, if only equity among regions was emphasized (compared to effectiveness), the performance of the resulting network declines by over 8%

given the limitations inherent in the equity objective. Using real data from OH for 2019, we demonstrate that the state could achieve a 31.2% and 33.1% reduction in mistriages by using our approach to redistribute and optimize their trauma network.

In the following sections, we first review the existing literature in Section 2. Our proposed optimization model for NTNDP and the solution approach are discussed in Sections 3 and 4, respectively. Next, we discuss our experimental study in Section 5 and illustrate the use of our approach on a real network in Section 6. Finally, in Section 7, we summarize our key findings and discuss avenues for further research.

3.2 Literature Review

Several approaches to address a variety of healthcare facility location problems have been proposed; e.g., primary health centers (Günes et al., 2014), long-term care centers (Cardoso et al., 2015; Intrevado et al., 2019), preventive healthcare facilities (Zhang et al., 2009; Zhang et al., 2010), ambulance location and/or relocation (Reuter-Oppermann et al., 2017; Vanbuuren et al., 2018), among others. For a comprehensive review, see Reuter-Oppermann et al. (2017), Ahmadi-Javid et al. (2017), and Gunes et al. (2019).

Because our work focuses on patient safety, our review suggests that two types of surrogate metrics for patient safety have been widely used in the literature; (i) minimizing total distance or travel time across all constituents (Cocking et al., 2012; Schmid, 2012; Beliën et al., 2013; Toro-Díaz et al., 2013; Chen et al., 2013; Bayram et al., 2015) and (ii) maximizing demand coverage within a fixed access time (Ingolfsson et al., 2008; Balcik & Beamon, 2008; Lim et al., 2011; Shariff et al., 2012; Kim & Kim, 2013; Salman & Yücel, 2015).

In terms of patient safety in trauma network design, Branas et al. (2000) proposed a model (known as TRAMAH) to simultaneously locate major trauma centers and air ambulances to maximize coverage of severely injured patients. Cho et al. (2014) also presented a model to simultaneously find major trauma centers and medical helicopters to maximize the expected number of patients transported to an MTC within 60 minutes. The authors incorporated busy fraction of medical helicopters in their model and developed the Shifting Quadratic Envelopes algorithm to optimize the problem. Lee et al. (2018) extended this model to a multiperiod location model by introducing an additional decision on when to locate trauma centers and air ambulances over a planning horizon. Considering additional complexity, the authors proposed a solution approach that iteratively updates helicopters' availability using the previous step of optimization result. However, these approaches do not account for non-severely injured patients (who affect srOT) and intermediate trauma centers (which can improve access in rural areas).

Jansen et al. (2015) proposed a novel data-driven approach to locate MTCs and ITCs with the bi-objective of minimizing the total access time and the number of exceptions or srUT for Scotland. The same authors developed a multi-fidelity surrogate-management strategy to reduce the computation time for real-world data-driven optimization problems (Wang et al., 2016). They demonstrated the viability of their approach using real data from the state of Colorado's trauma system (Jansen et al., 2018). While this model considered ITCs, it failed to account for non-severely injured patients and various destination determination criteria.

To support decision making around trauma networks, the ACS Committee on Trauma (ACS COT) developed the Needs-Based Assessment of Trauma System (NBATS) tool (ACS-NBATS, 2015). NBATS uses six criteria to suggest the required number of MTCs in a given geographical area, also known as the trauma service area (TSA); population, median travel times, lead agency support, an existing number of major trauma centers, and where severely injured patients are transported (MTCs and NTCs). However, NBATS does not determine the location of the MTCs. To address this gap, Parikh et al. (2022) proposed a model for a Performance-based Assessment of Trauma System (PBATS) to find the minimum number and location of MTCs by keeping system-related under-triage (srUT) and over-triage (srOT) rates within a prespecified limit. Recently, Hirpara et al. (2022) proposed a bi-objective model for trauma center location problem (TCLP) to determine the number and location of MTCs and NTCs in order to minimize the weighted sum of srUT and srOT rates. They demonstrated their approach through a case study based on the existing network of a US state with focus on ‘greenfield’ design and ‘redistribution’ of existing MTCs. While both these recent works consider both types of patients and associated mistriages, they do not explicitly consider ITCs (a critical trauma facility for a viable trauma system) and various destination determination criteria (that affect mistriages).

In terms of destination determination criterion, prior trauma location models have only considered a single criterion, often mimicking the ACS-suggested protocol. However, multiple criteria have been observed in practice besides this protocol, with patient choice and closest facility being dominant (Newgard et al., 2011, Newgard et al., 2013). Patient choice has been studied in many IE/OR journals to determine destination location in an optimization framework. Zhang et al. (2012) studied the impact of client choice behavior on the preventive care facility network configuration. The authors presented two alternative models; (i) probabilistic-choice model based on the multinomial logit (MNL) model, where a client may patronize each facility with a certain probability based on the attractiveness of the facilities, and (ii) optimal-choice model, where each client will go to the most attractive facility. Zhang & Atkins (2019) presented several models for designing a network of walk-in medical facilities. For a choice model, they considered travel time, attractiveness, and waiting time at the facility to calculate the utility of receiving care at a given facility. Further, they also considered deterministic patient choice, where a patient chooses the facility with the highest utility to receive care. Closest facility criteria have been considered in Cardoso et al. (2015), Mestre et al. (2015), and Nasrabadi et al. (2020).

Our review of the literature suggests the following gaps:

- All prior trauma system design approaches failed to explicitly consider multiple destination determination criteria alluded in medical literature and followed in practice.
- None of the prior research considered both types of patients (severe and non-severe), along with consideration of intermediate trauma centers.
- Further, equity in safety among regions, along with effectiveness, have not been considered jointly in the trauma literature (further elaborated in Section 3).

To fill the above gaps, we propose a nested multi-level, multi-customer, multi-destination determination criteria and multi-transportation bi-objective (equity and effectiveness) capacitated model. Our proposed NTNDP model not only accounts for both types of patients (severely and non-severely injured) and associated mistriages, but also explicitly considers several other factors that affect system performance; i.e., ITCs, three criteria for destination determination, and equity and effectiveness in patient safety. We now present our proposed model.

3.3 A Bi-Objective Model for NTNDP

Our generic model is developed for a Trauma Service Area (TSA); a geographical area comprising a collection of counties in a state, the state itself, or even collection of states, similar to the definition in NBATS tool. Further, this TSA is divided into subareas known as regions or districts, which have the oversight to providing trauma care within that region. Because of the existence of such regions within a TSA, it becomes critical to consider the equity of patient safety among regions when designing a trauma network.

A variety of equity measures have been proposed in the literature when allocating public resources; e.g., minimax, variance, range, sum of absolute deviations, sum of absolute deviation from desire standard, squared coefficient of variation, and Gini index (Burkey et al., 2012; Lejeune et al., 2013; Smith et al., 2013; Chanta et al., 2014; Wang et al., 2015; Ares et al., 2016; Enayati et al., 2019). However, little consensus exists concerning which equity measure researchers should employ (Stone, 1997; McLay & Mayorga, 2013). Based on our interactions with trauma collaborators, their general focus is to improve patient safety in the worst-performing region (among all regions) of the TSA. Therefore, we use the minimax equity measure as it intrinsically focuses on improving the performance of the worst one.

However, any equity measure as a standalone objective often results in undesirable, sometimes meaningless, solutions (Burkey et al., 2012; Smith et al., 2013; Enayati et al., 2019). For instance, minimax cannot distinguish between two networks with identical worst performing regions; however, one solution could have better performance in other regions than the other solution. In some situations, if a higher aggregated network performance can be achieved with a slightly less equity among individual regions, then it may be a preferred network for the decision makers. Considering both these factors, recent literature has proposed ‘effectiveness’ as a supporting metric, alongside equity (Burkey et al., 2012; Smith et al., 2013; Enayati et al., 2019). We, therefore, use both equity and effectiveness as objective terms in the proposed model. That is, the NTNDP is to determine the optimal number and location of MTCs and ITCs to maximize the weighted sum of equity in patient safety (among regions) and effectiveness (across the TSA).

Recall that patients with traumatic injuries can be classified into two categories; (i) severely injured patients with life-threatening injuries and (ii) non-severely injured with other trauma injuries. In line with the existing trauma literature, we use Injury Severity Score (ISS) as a surrogate to estimate the severity of injury at the incidence location. We also define two thresholds: ‘access’ threshold as a clinically-driven time (specified in trauma literature) to reach a hospital (MTC, ideally) and ‘bypass’ threshold as a resource-driven value that specifies the maximum additional minutes (compared to a nearby MTC/ITC) that EMS can dedicate to transport them to an NTC (ideal hospital).

Before we present the model, we first present some preliminaries around destination determination and triage classification.

3.3.1 Triage classification

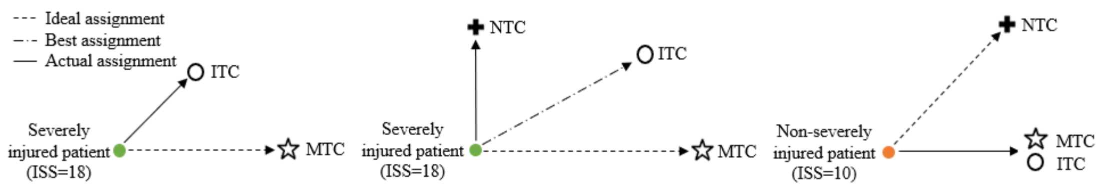
Table 10 classifies the triage types based on injury severity and destination hospital type. Irrespective of the destination determination criteria, if a patient is transported to the

Table 10. Classification of triage type based on injury severity and destination hospital type

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Injury Severity Score (ISS)</td></tr>
<tr><td>ISS>15 (severely injured)</td><td>ISS≤15 (non-severely injured)</td></tr>
<tr><td rowspan="3">Destination hospital type</td><td>MTC</td><td>System-related appropriate-triage (srAT^{P})</td><td rowspan="2">System-related over-triage (srOT)</td></tr>
<tr><td>ITC</td><td>System-related under-trigae stabilized (srUT^{S})</td></tr>
<tr><td>NTC</td><td>System-related under-triage (srUT)</td><td>System-related appropriate-triage (srAT^{N})</td></tr>
</table>

ideal hospital type based on their injury severity, then it is deemed as appropriate triage; i.e., severely injured transported to MTC is classified as srATP and non-severely injured transported to NTC is classified as srATN. Mismatch in injury severity and destination hospital type results in mistriage (see Figure 10); recall that ISS>15 is considered a severely injured patient.

As mentioned earlier, delay in definitive care for severely injured patients (i.e., srUTS or srUT) increases the likelihood of an adverse outcome due to the life-threatening nature of those injuries. We combine both mistriages associated with severely injured patients and refer to it as 'system-related aggregated under-triage (srAU).' It defines as a weighted sum of srUT and srUTS. In contrast, mistriage of the non-severely injured patients indirectly impacts patient safety and is relatively non-serious. Therefore, we consider mistriages of severely injured patients (i.e., srUT and srUT^{S}, aggregated as srAU in the model) as the primary patient safety metric, while mistriage of non-severely injured patients (i.e., srOT) as a secondary patient safety metric.



(a) System-related under-triage stabilized ($srUT^s$)

(b) System-related under-triage (srUT)

(c) System-related over-triage (srOT)

Figure 10. Mistriages based on severity of injury and destination hospital type

3.3.2 Destination determination

Recall that in Section 2, we mentioned that we incorporate three dominant destination determination criteria that EMS use at the incidence location; protocol, patient choice, and closest facility.

3.3.2.1 Protocol

The protocol criterion is essentially the Notional Tasking Algorithm (NTA) that attempts to mimic the EMS decision making process at the incidence location, as proposed by the American College of Surgeons (ACS). It considers clinical and resource factors for destination determination. In this paper, we extend the NTA used in Hirpara et al. (2022) to consider ITCs for severely injured patients (see Figure 11). The NTA follows an ordered priority list based on patient's injury severity, the thresholds, and the vicinity of MTCs, ITCs, and NTCs.

<u>3.2.1.1 Severely injured patient</u>: The top priority is to assign a severely injured patient to any MTC (ideal hospital) within the 'access' threshold via ground; if this occurs, we refer to it as system-related appropriate triage positive ($srAT^P$). If no MTC is accessible via ground, then the second priority is assigning them to an MTC that is accessible via air

ambulance (if available); this is also considered as $srAT^P$. For air ambulance transport, the NTA considers inbound-to- incidence location, loading, and transport-to-MTC times and compares it against the 'access' threshold; if below, then such transport is feasible.

If one of the first two priorities satisfy, then the third and fourth priorities are to assign them to an accessible ITC (not ideal, but better equipped than NTC) via ground and air. Because the ideal trauma hospital (i.e., MTC) is not chosen, we consider such a patient as system-related under-triage stabilized ($srUT^S$). The modifier 'stabilized' is used because ITCs are often capable to stabilize a severely injured patient. If all of the above are infeasible, then, as the last option for EMS, the patient is assumed to be transported to a nearby NTC; this results in system-related under-triage (srUT).

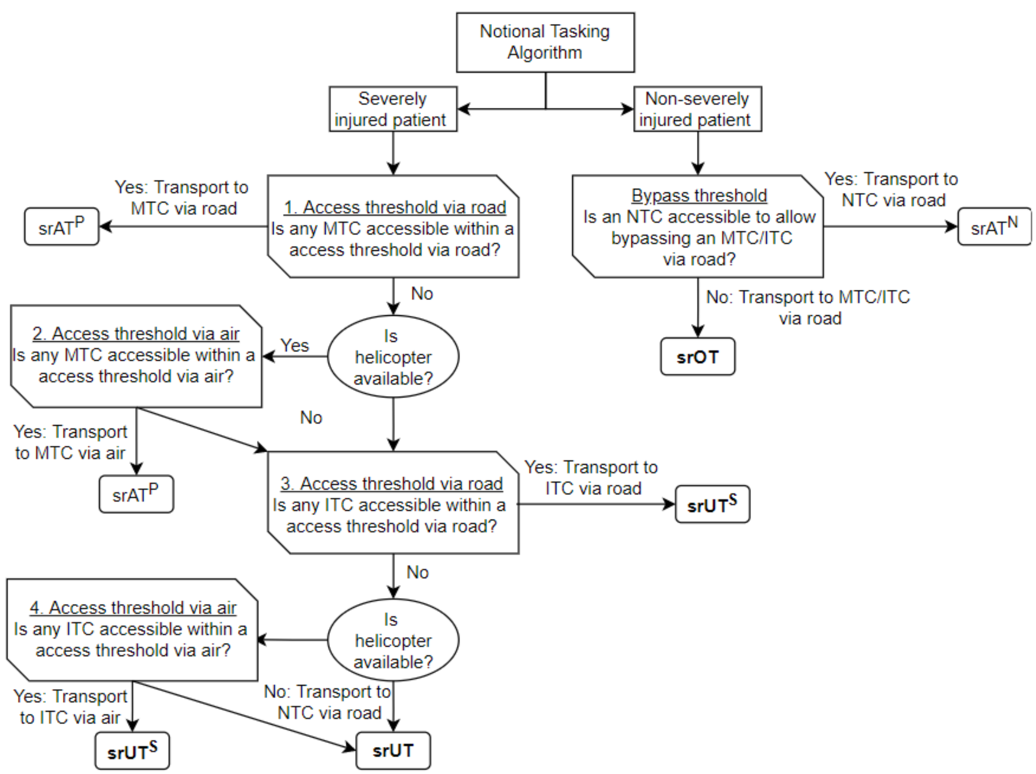


Figure 11. Notional Tasking algorithm

3.2.1.2 Non-severely injured patient: For such ISS≤15 patients, the 'bypass' threshold is a resource-driven value that specifies the maximum additional minutes (compared to a nearby MTC/ITC) EMS can dedicate to transport the patient to an NTC (ideal hospital). For example, suppose the additional time to reach an NTC beyond the time to nearest MTC/ITC (say, 10 minutes) is within the 'bypass' threshold (say 15 minutes), then, in practice, the EMS is often likely to take the patient to that NTC. We refer to this type of a situation as system-related appropriate triage negative ($srAT^N$). Otherwise, the EMS would likely take the patient to the nearby MTC/ITC (due to longer drive or other operational criteria) resulting in system-related under-triage (srOT).

3.3.2.2 Patient choice

Anecdotal evidence and discussions with EMS suggest that patients often choose bigger hospitals over nearby hospitals due to their perception that the bigger the hospital, the better the care. However, travel time to the hospital also impacts their decision as they want to reach the hospital soon to avoid delay in receiving the care. In line with literature in the healthcare domain, we model patients' choices through a linear utility model (Zhang et al., 2012; Haase & Müller, 2015; Zhang & Atkins, 2019). Accordingly, linear function comprises two dominant components that impact patients' decision making; (i) the attractiveness of hospitals and (ii) ground travel time to those hospitals. The below equation calculates the utility of patient i receiving care at hospital j as a linear function of attractiveness of facility j (A_j) and ground travel time from the location of patient i to hospital j (TG_{ij}):

$$u_{ij} = \beta_1 A_j - \beta_2 \, TG_{ij}$$

Here, we let the attractiveness of a hospital for a patient (A_j) depend upon the hospital type (MTC, ITC, or NTC) and that it is identical for all patients. The coefficients β_1 and β_2 denote the sensitivity to the two components, respectively, and can be estimated empirically based on available data or existing literature. Each patient is assigned to a hospital that has the highest utility across all hospitals.

3.3.2.3 Closest facility

In case of extreme weather considerations, road closures, or other unforeseen circumstances, EMS providers tend to prioritize closest facility over protocol or patient choice, irrespective of patient's injury severity and closest hospital's type. We model this by assigning such a patient to the closest hospital from the incidence location.

3.3.3 Optimization model

With this background, we now present the model under the following assumptions:

- The candidate locations for the MTCs, ITCs, and NTCs are known and finite.
- The number of patients, their locations, and severity are deterministic and known.
- The destination determination criteria for each patient is preassigned based on the given %-allocation among the three criteria.
- All severely injured patients, if initially transported to an ITC or an NTC, will eventually be transferred via ground to the nearest MTC from the incidence location (to allow them access to definitive care); patients are categorized as srUT and $srUT^S$ accordingly because of delays in reaching MTC.

- The attractiveness of the facility to patients is given and depends only on the hospital's type.
- A severely injured patient can be assigned to any MTC/ITC accessible within the access threshold in protocol criteria.

Further, in keeping up with the existing literature and what was observed in the data we had access to, we make the following assumptions about transportation modes:

- Ground and air transport times are known and deterministic.
- Air ambulance is only allowed to transport severely injured patients to MTCs and ITCs in the protocol criteria.
- While ground ambulance services are available without constraints, the availability of air ambulances was restricted to 15% of total severely injured patients based on data from state trauma agencies reports.

Tables 11 and 12 summarize the parameters and decision variables, respectively, used in the model.

Table 11. Parameters in the model

Notation	*Definition*
I	Set of incidences for trauma patients, divided into the subsets • I^O; subset of patients assigned via protocol criteria; $i \in I^O \subseteq I$ • I^P; subset of patients assigned via patient choice criteria; $i \in I^P \subseteq I$ • I^C; subset of patients assigned via closest facility criteria; $i \in I^C \subseteq I$
J	Set of candidate locations (for MTC, ITC, and NTC); $j \in J$
K	Set of regions in the TSA; $k \in K$
L	Set of hospital type; $l \in L$; l = 1, 2, 3 represent MTC, ITC, and NTC, respectively
ω_1, ω_2	Weights for equity and effectiveness in the objective function; $\omega_1 + \omega_2 = 1$
γ, δ	Weight for srUT and $srUT^S$ patient
S_i	Injury severity of patient i; 1, if severely injured (ISS > 15); 0, otherwise
R_{ik}	Region indicator, 1 if patient i is from a region k; 0, otherwise
TG_{ij}, TA_{ij}	Travel time from patient i to any candidate location j via ground and air
SG_{ij}	Subset of set J corresponding to each i-j pair and includes all other locations $t \in J$ such that ground travel time from patient i to t is greater than from i to j (i.e., $t \in SG_{ij}$, if $TG_{ij} < TG_{it}$; $j,t \in J$)
α	'Access' time threshold to determine srUT (for protocol criteria only)
β	'Bypass' time threshold to determine srOT (for protocol criteria only)
T_{in}, T_{load}	Inbound time from base-to-incidence location and loading time of patient at the incidence location for an air ambulance
Z	Maximum allowable patients via air ambulance
A^l	Attractiveness of hospital level l
β_1, β_2	Coefficient for attractiveness and travel time in the utility function
$V_{MTC}^{min}, V_{MTC}^{max}$	Minimum and maximum allowable volume of a severely injured patient at MTC
$V_{ITC}^{min}, V_{ITC}^{max}$	Minimum and maximum allowable volume of a severely injured patient at ITC
Ψ	Minimum allowable ratio of number of ITCs to MTCs
OT^{max}	Maximum allowable overall over-triage patients
A_{ij}^G, A_{ij}^A	Accessibility of candidate location j from patient i within α via ground and air; 1, if candidate location j is accessible from patient i; 0, otherwise
ρ	Equivalent fraction of an MTC corresponding to an ITC
C	Maximum equivalent MTCs allowed in the network
M	Big number

Table 12. Decision variables in the model

Notation	*Definition*
x_j^l	1, if a candidate location *j* is designated to be level *l*; 0, otherwise
au_k, au^{max}	System-related aggregated under-triage in region *k*; $au^{max} = \max_k\{au_k\}$
y_{ij}^1	1, if patient *i* is transported via ground to location *j* (i.e., if *j* is an MTC, then patient *i* is $srAT^P$ and if *j* is an NTC, then patient *i* is $srAT^N$); 0, otherwise;
y_{ij}^2	1, if severely injured patient *i* ($i \in I^O \subseteq I$) is transported via air to location *j* that is marked as MTC (i.e., $srAT^P$ via air); 0, otherwise
y_{ij}^3	1, if severely injured patient *i* is transferred (from ITC or NTC) to location *j* that is marked as MTC (i.e., transferred srUT or $srUT^S$ patient); 0, otherwise
y_{ij}^4	1, if severely injured patient *i* is transported via ground to location *j* that is marked as ITC (i.e., $srUT^S$ via ground); 0, otherwise
y_{ij}^5	1, if severely injured patient *i* ($i \in I^O \subseteq I$) is transported via air to location *j* that marked as ITC (i.e., $srUT^S$ via air); 0, otherwise
$ne_{ij}^{MTC\_ITC}$	1, if location *j* is marked as MTC or ITC and is the nearest non-NTC via ground for patient *i* ($i \in I^O \subseteq I$); 0, otherwise
ne_{ij}^{NTC}	1, if location *j* is marked as NTC and is the nearest NTC via ground for patient *i* ($i \in I^O \subseteq I$); 0, otherwise
u_{ij}, u_i^{max}	Utility of patient *i* receiving care at hospital *j*; $u_i^{max} = \max_j\{u_{ij}\}$
n_{ij}	1, if candidate location *j* is the nearest hospital for patient *i* ($i \in I^C \subseteq I$) or if the highest utility for patient *i* ($i \in I^P \subseteq I$) occurs for a hospital *j*; 0, otherwise

Minimize: $\omega_1\, au^{max} + \omega_2 \frac{\sum_k au_k}{|K|}$

Subject to:

Calculation of region-wise and maximum aggregated under-triage

$$au_k = \gamma \sum_{i:S_i=1} R_{ik} \sum_j (y_{ij}^3 - y_{ij}^4 - y_{ij}^5) + \delta \sum_{i:S_i=1} R_{ik} \sum_j (y_{ij}^4 + y_{ij}^5)\,;\ \forall k \in K \quad (1)$$

$$au^{max} \geq au_k;\ \forall k \in K \quad (2)$$

Limit on number of MTCs, and their minimum and maximum volume

$$\sum_l x_j^l = 1;\ \forall j \in J \quad (3)$$

$$\sum_j x_j^1 + \rho \sum_j x_j^2 \leq C \quad (4)$$

$$x_j^1\, V_{MTC}^{min} \leq \sum_{i:\,S_i=1} (y_{ij}^1 + y_{ij}^2 + y_{ij}^3) \leq x_j^1\, V_{MTC}^{max};\ \forall j \in J \quad (5)$$

Allowable number of ITCs, and their minimum and maximum volume

$$\sum_j x_j^2 \geq \psi \sum_j x_j^1 \quad (6)$$

$$x_j^2\, V_{ITC}^{min} \leq \sum_{i:\, S_i=1}(y_{ij}^4 + y_{ij}^5) \leq x_j^2\, V_{ITC}^{max};\ \forall j \in J \tag{7}$$

Limit on state-wide srOT

$$\sum_i (1 - S_i) - \sum_{i:S_i=0} \sum_j y_{ij}^1 \ \leq\ OT^{max} \tag{8}$$

Assignment and triage classification using protocol criteria

$$\sum_j (y_{ij}^1 + y_{ij}^2 + y_{ij}^3) = 1; \forall i \in I^O : S_i = 1 \tag{9}$$

$$y_{ij}^1 = 0;\ \forall i \in I^O : S_i = 1, \forall j \in J, TG_{ij} > \alpha \tag{10}$$

$$y_{ij}^2 = 0;\ \forall i \in I^O : S_i = 1, \forall j \in J, TA_{ij} + T_{in} + T_{load} > \alpha \tag{11}$$

$$\sum_j A_{ij}^G x_j^1 \leq M\ (1 - \sum_j y_{ij}^2);\ \forall i \in I^O : S_i = 1 \tag{12}$$

$$x_j^1 + \sum_{t\, \in\, SG_{ij}} y_{it}^3 \leq 1;\ \forall i \in I^O : S_i = 1, \forall j \in J \tag{13}$$

$$\sum_j (y_{ij}^4 + y_{ij}^5) \leq \sum_j y_{ij}^3\, ;\ \forall i \in I^O : S_i = 1 \tag{14}$$

$$y_{ij}^4 = 0;\ \forall i \in I^O : S_i = 1, \forall j \in J, TG_{ij} > \alpha \tag{15}$$

$$\sum_j A_{ij}^G x_j^1 \leq M\ (1 - \sum_j y_{ij}^4);\ \forall i \in I^O : S_i = 1 \tag{16}$$

$$y_{ij}^5 = 0;\ \forall i \in I^O : S_i = 1, \forall j \in J, TA_{ij} + T_{in} + T_{load} > \alpha \tag{17}$$

$$\sum_j A_{ij}^G\, x_j^2 + \sum_j A_{ij}^G x_j^1 + \sum_j A_{ij}^A x_j^1 \leq M\ (1 - \sum_j y_{ij}^5);\ \forall i \in I^O : S_i = 1 \tag{18}$$

$$\sum_i \sum_j (y_{ij}^2 + y_{ij}^5) \leq\ Z \tag{19}$$

$$ne_{ij}^{NTC} \leq x_j^3;\ \forall i \in I^O : S_i = 0, \forall j \in J \tag{20}$$

$$\sum_j ne_{ij}^{NTC} = 1;\ \forall i \in I^O : S_i = 0 \tag{21}$$

$$x_j^3 + \sum_{t\, \in\, SG_{ij}} ne_{it}^{NTC} \leq 1;\ \forall i \in I^O : S_i = 0, \forall j \in J \tag{22}$$

$$ne_{ij}^{MTC\_ITC} \leq x_j^1 + x_j^2;\ \forall i \in I^O : S_i = 0, \forall j \in J \tag{23}$$

$$\sum_j ne_{ij}^{MTC\_ITC} = 1;\ \forall i \in I^O : S_i = 0 \tag{24}$$

$$x_j^1 + x_j^2 + \sum_{t\, \in\, SG_{ij}} ne_{it}^{MTC\_ITC} \leq 1;\ \forall i \in I^O : S_i = 0, \forall j \in J \tag{25}$$

$$\sum_j \left(ne_{ij}^{NTC}\, TG_{ij}\right) - \sum_j \left(ne_{ij}^{MTC_{ITC}}\, TG_{ij}\right) - \beta\ \leq\ M\ \left(1 - \sum_j y_{ij}^1\right); \forall i \in I^O : S_i = 0 \tag{26}$$

$$\sum_j y_{ij}^1 \leq 1;\ \forall i \in I^O : S_i = 0 \tag{27}$$

Utility calculation for an assignment using patient choice criteria

$$u_{ij} = \beta_1 \sum_l A^l x_j^l - \beta_2\, TG_{ij};\ \forall i \in I^P, \forall j \in J \tag{28}$$

$$u_i^{max} \geq u_{ij};\ \forall i \in I^P, \forall j \in J \tag{29}$$

$$(u_i^{max} - u_{ij}) - M\ (1 - n_{ij}) \leq 0;\ \forall i \in I^P, \forall j \in J \tag{30}$$

Closest hospital for patients assigning through closest facility criteria

$$1 + \sum_{t \in SG_{ij}} n_{it} \leq 1;\ \forall i \in I^C, \forall j \in J \tag{31}$$

$$\sum_j n_{ij} = 1;\ \forall i \in I^C \cup\ I^P \tag{32}$$

Assignment of patients through patient choice and closest facility criteria

$$\sum_j (y_{ij}^1 + y_{ij}^3) = 1;\ \forall i \in I^C \cup\ I^P: S_i = 1 \tag{33}$$

$$n_{ij} + x_j^1\ \geq\ 2\, y_{ij}^1;\ \forall i \in I^C \cup\ I^P: S_i = 1, \forall j \in J \tag{34}$$

$$x_j^1 + \sum_{t \in SG_{ij}} y_{it}^3 \leq 1;\ \forall i \in I^C \cup\ I^P: S_i = 1, \forall j \in J \tag{35}$$

$$\sum_j y_{ij}^4 \leq \sum_j y_{ij}^3\,; \forall i \in I^C \cup\ I^P: S_i = 1 \tag{36}$$

$$n_{ij} + x_j^2\ \geq\ 2\, y_{ij}^4;\ \forall i \in I^C \cup\ I^P: S_i = 1, \forall j \in J \tag{37}$$

$$n_{ij} + x_j^3\ \geq\ 2\, y_{ij}^1;\ \forall i \in I^C \cup\ I^P: S_i = 0, \forall j \in J \tag{38}$$

Bounds on decision variables

$$x_j^l \in \{0, 1\};\ \forall j \in J, \forall l \in L \tag{39}$$

$$au_k, au^{max} \geq\ 0;\ \forall k \in K \tag{40}$$

$$y_{ij}^1 \in \{0, 1\};\ \forall i \in I, \forall j \in J \tag{41}$$

$$y_{ij}^3, y_{ij}^4 \in \{0, 1\};\ \forall i \in I, \forall j \in J \tag{42}$$

$$y_{ij}^2, y_{ij}^5 \in \{0, 1\};\ \forall i \in I^O, \forall j \in J \tag{43}$$

$$ne_{ij}^{NTC}, ne_{ij}^{MTC\_ITC} \in \{0, 1\};\ \forall i \in I^O, \forall j \in J \tag{44}$$

$$u_{ij}, u_i^{max} \in \mathbb{R};\ \forall i \in I^P, \forall j \in J \tag{45}$$

$$n_{ij} \in \{0, 1\};\ \forall i \in I^C \cup\ I^P, \forall j \in J \tag{46}$$

The model minimizes a weighted sum of maximum srAU patients among regions (equity measure) and average srAU patients across regions (effectiveness measure). We use y_{ij} variables in the model to classify triage types and record destination hospitals for further volume calculation.

For each region, Constraints (1) calculate the total aggregated under-triage patients (srAU), which is a weighted sum of overall srUT (first term) and srUT^S (second term) in a region. Constraints (2) calculate maximum srAU patients among all regions. Constraints (3) ensure that each candidate location is designated as either MTC, ITC, or NTC. Constraint (4) ensures that the total number of MTCs and ITCs must be less than or equal to the maximum allowable equivalent MTCs (which allows the model to find the best combination of MTCs and ITCs considering budgetary constraints). Constraints (5) bound volume of severely injured patients (directly transported to MTC or transferred from ITC or NTC) at candidate location *j* if it is designated as MTC. Constraint (6) ensures that the ratio of ITCs to MTCs is within a prespecified value. Constraints (7) limit the volume of severely injured patients (transported to ITC via ground or air) at candidate location *j* if it is designated as ITC. Constraint (8) ensures that total TSA-wide srOT patients (difference of total non-severely injured patients and $srAT^N$) is within an allowable limit.

For patients assigned via protocol criteria, Constraints (9)-(27) assign them to hospitals and classify their triage types. Constraints (9) ensure that each severely injured patient is either initially transported to an MTC via ground or air, or eventually transferred to MTC from an ITC/NTC. Constraints (10) and (11) rule out an assignment of severely injured patient *i* to every inaccessible MTC via ground and air, respectively. Note that an MTC is considered not accessible via ground if ground travel time is higher than the 'access' threshold; it is not accessible via air if the total airtime (sum of inbound, loading, and air travel) is higher than the 'access' threshold. Constraints (12) rule out an assignment of severely injured patient *i* to all MTCs via air if any MTC is accessible via ground. That is, in an effort to preserve the limited air ambulance trips, a patient is only airlifted if no MTC is accessible via ground.

Constraints (13) capture the transfer of severely injured patients from an ITC or NTC to the nearest MTC to receive definitive care. Constraints (14) ensure that severely injured patient *i* is assigned to an ITC (via ground or air) if initially not assigned to any MTC ($\sum_j y_{ij}^3 = 1$). Constraints (15) rule out an assignment of severely injured patient *i* to ITCs not accessible via ground. However, if any MTC is accessible via ground, then Constraints (16) rule out assignment of severely injured patient *i* to all ITCs. Constraints (17) rule out an assignment of severely injured patient *i* to ITCs not accessible via air, while Constraints (18) rule out assignments to all ITCs if any ITC is accessible via ground or any MTC is accessible via ground or air. Constraints (16) and (18) ensure priority-based assignment of severely injured patients discussed in the section 3.2.1.1. Constraints (19) ensure that the air transport usage does not exceed their availability.

For each non-severely injured patient *i*, Constraints (20)-(22) determine the nearest NTC. Constraints (20) ensure that a candidate location *j* must be an NTC to be considered as the nearest NTC, Constraints (21) make sure that for a non-severely injured patient *i*, only one NTC should be considered as the nearest NTC. For any pair of patient *i* and candidate location *j*, if a candidate location *j* is marked as NTC, then Constraints (22) rule out the assignment of patient *i* to candidate location(s) *t* that are located further (in terms of time) than *j*. Constraints (23)-(25) serve the same purpose as (20)-(22), respectively, for the nearest non-NTC (MTC or ITC) via ground. For non-severely injured patient *i*, Constraints (26) rule out the assignment to all NTCs if the 'bypass' threshold criterion is not met; this patient is marked as srOT. Note that srOT occurs when MTC or ITC is closer than the nearest NTC. We do not need to explicitly assign srOT patients to an MTC or ITC as they are not counted towards trauma volume; these patients are often discharged from the ED of an MTC or ITC without admission to the inpatient trauma unit.

For patients assigned via patient choice criteria, Constraints (28)-(30) capture patients' choices using the utility model. For each patient i, Constraints (28) calculate utility of receiving care at each hospital (candidate location), Constraints (29) find maximum utility among all hospitals, while Constraints (30) record the hospital with the maximum utility. Constraints (31) and (32) find the closest facility for each patient assigned via closest facility criteria. For patients assigned via patient choice and closest facility criteria, variable n_{ij} capture patient choice and the closest facility, respectively. Constraints (32) ensure that each patient has only one closest facility and select one hospital with maximum utility for the closest facility and patient choice criteria, respectively.

For patients assigned through patient choice and closest facility criteria, Constraints (33)-(38) assign them to hospitals and classify their triage types. Constraints (33) ensure that each severely injured patient is initially assigned to MTC or eventually transferred to MTC from ITC/NTC. Constraints (34) assign and classify severely injured patient i as $\mathrm{srAT}^{\mathrm{P}}$ if the nearest hospital or patient's choice is MTC. Constraints (35) ensure that each severely injured patient is transferred to the nearest MTC after being initially transported from the incidence location to an ITC ($\mathrm{srUT}^{\mathrm{S}}$ patient) or an NTC (srUT patient). Constraints (36) ensure that severely injured patient i is assigned to an ITC if initially not assigned to any MTC; i.e., $\sum_j y_{ij}^3 = 1$. Constraints (37) classify a severely injured patient i as $\mathrm{srUT}^{\mathrm{S}}$ if the criterion of nearest hospital or patient choice results in ITC. Constraints (38) classify a non-severely injured patient i as $\mathrm{srUT}^{\mathrm{N}}$ if the criterion of nearest hospital or patient choice results in NTC. Constraints (39)-(46) define bound on decision variables.

Note that the NTNDP can be characterized as a hierarchical, discrete, multi-facility location problem. Such problems are combinatorial in nature and have been shown to be NP-hard (Daskin, 2011). For even 50 candidate hospital locations, there are $3^{50} = 7.18\text{x}10^{23}$ solutions. Our preliminary experiments suggested that commercial software such as CPLEX and Gurobi encountered out-of-memory issues for realistic problem instances that normally have 100+ locations and 1,000+ patients. We, therefore, explored a tailored '3-phase' approach to avoid such issues and find a near-optimal solution. We now discuss our proposed approach.

3.4 A 3-Phase Solution Approach

A primary goal of any trauma system is to provide prompt care to severely injured patients. Data from the state of OH indicated that severely injured patients made up about 15% of the total patients. The problem complexity can thus be reduced if relaxed the model to first focus on severely injured patients, and then the non-severely injured patients. Considering this, we propose a '3-phase' approach that systematically reduces the problem complexity into different phases to decrease the number of decision variables and constraints (see Figure 12).

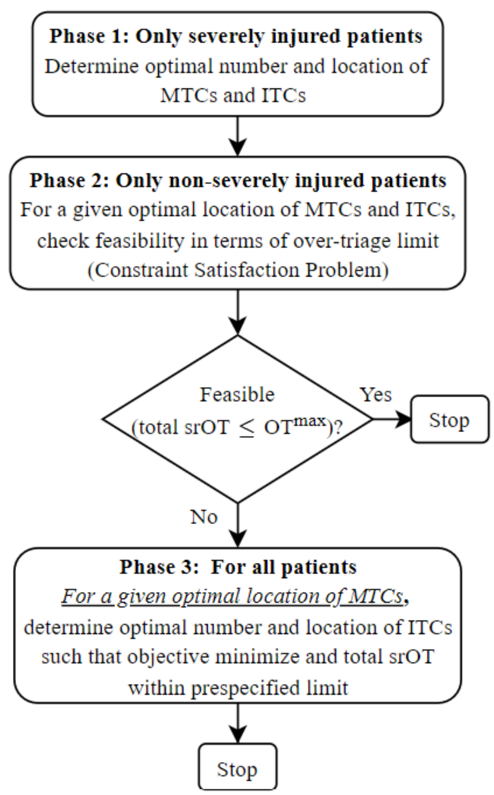


Figure 12. Flowchart of the '3-phase' approach

In Phase 1, we only consider severely injured patients ($S_i = 1$) and determine the optimal location of MTCs and ITCs based on the NTNDP model presented earlier. Essentially, we remove all decision variables and constraints related to non-severely patients. The Phase 1 problem is as follows:

Phase 1 problem

minimize: $\omega_1 \, au^{max} \; + \; \omega_2 \, \frac{\sum_k au_k}{|K|}$

s.t.

Constraints (1)-(7), (9)-(19), (28)-(37), (39)-(43), (45)-(46)

In Phase 2, we use the solution from Phase 1 and solve a Constraint Satisfaction Problem for non-severely injured patients. Essentially, we check the feasibility of the Phase 1 solution in terms of total over-triage patients (which are triggered by non-severely injured patients). Feasibility in Phase 2 means overall srOT patients are within the prespecified limit, and that the solution found in Phase 1 is optimal to the entire problem. However, infeasibility in Phase 2 indicates the need for changes in the solution from Phase 1 to keep total srOT patients within the limit, which then invokes Phase 3. As non-severely injured patients do not directly impact the objective of the NTNDP, the model for the Phase 2 can be defined by Constraints (8), (20)-(32), (38), (41), and (44)-(46).

Phase-3 problem

minimize: $\omega_1 \, au^{max} \; + \; \omega_2 \, \frac{\sum_k au_k}{|K|}$

s.t.

Constraints (1)-(46)

$$x_j^1 \; = \; 1; \; \forall j \in J' \qquad (47)$$

In Phase 3, we fix the location of MTCs (obtained from Phase 1) while considering both types of patients and solve the original model for NTNDP. Basically, for a given location of MTCs, we find the optimal location of ITCs to minimize the objective while keeping total srOT within a limit. Fixing MTCs is reasonable as the impact of MTCs on the objective is relatively higher than ITCs due to their capability of providing definitive care to severely injured patients. In the formulation for Phase 3, we add Constraints (47) to fix the location of MTCs obtained from Phase 1 (represented by set J' where $J' \in J$).

We used Gurobi solver on Dell I7-10700 CPU @2.90 GHz Desktop with 32GB RAM to find an optimal solution in each phase.

3.5 Computational Study

We now detail our experimental study starting with the test area generation (referred to as a TSA), sources of data collection, evaluation of the solution approach, sensitivity analysis, and insights.

3.5.1 TSA determination

We consider the collection of counties in an existing midwestern US state as TSA. In so doing, we can use the underlying transportation network to estimate actual ground transportation times from the incidence locations to the candidate hospitals. Figure 13 illustrates the TSA with 34 counties and 64 hospitals. In this TSA, 21 counties are rural (61.7%), in line with the % of rural counties in the US (i.e., 62%). All 64 hospitals in the TSA were considered as candidate locations for an MTC, ITC, or NTC. We also grouped counties to represent

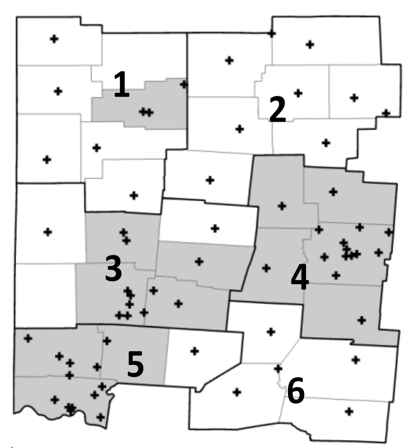


Figure 13. TSA with counties, and region; grey filled areas are urban counties; '+'represents candidate locations

regions similar to several state trauma agencies (MDHHS, 2022; TDSHS, 2022). In our chosen TSA, Regions 2 and 6 are entirely rural regions, while Region 4 is an entirely urban. Further, Region 1 is dominantly rural (with higher % of rural counties compared to urban counties), while Regions 3 and 5 are dominantly urban.

For all analyses, we used ArcGIS Pro 2.9.1 to calculate actual drive times to generate the ground time matrix (TG_{ij}) and the Haversine formula (assuming the helicopter speed of 120 mph) to generate the airtime matrix (TA_{ij}). Note that both these time matrices are pre-generated and serve as a look-up table during the solution process.

3.5.2 Performance of the 3-Phase approach

For performance evaluation of the '3-phase' approach, we considered 10 problem instances using this TSA, each with 5,000 patients (15.63% of severely injured patients), and between 15 and 60 candidate locations. We used 200 and 50 as a lower bound for the volume of severely injured patients at MTCs and ITCs, respectively. The limit for srOT patients is set as 50% of total non-severely injured patients. The attractiveness for MTCs, ITCs, and NTCs is set as 5, 4, and 1, respectively, and coefficients for attractiveness and ground travel time were set as 0.825 and 0.566. All other parameters are the same as the base case mentioned in section 5.4. We set the CPU-time limit as 12 hours for solving the Gurobi MIP solver.

Table 13 presents our computational experiments that compare the solution quality and runtime of the '3-phase' approach and 'Original model' (per Section 3.3) for several problem instances. The '% Difference' column represents the difference between the objective of the '3-phase' and 'Original model,' where positive value represents a '3-phase'

Table 13. Performance evaluation of the '3-phase' approach

Problem instance	*Candidate Locations*	Solution Quality			Runtime in Hours	
		Original model	*3-phase*	*% Difference*	*Original model*	*3-phase*
1	15	158.3	159.8	-0.93%	1.65	0.07
2	20	169.3	169.6	-0.22%	1.94	0.14
3	25	141.5	142.3	-0.53%	2.16	0.19
4	30	120.9	120.9	0%	3.29	0.28
5	35	122.5	122.5	0%	3.77	0.32
6	40	100.9	100.9	0%	6.70	0.4
7	45	113.1 (4.95)	112.9	0.17%	12	1.17
8	50	100.3 (6.17)	99.9	0.35%	12	1.53
9	55	Out of Memory	-	-	Out of Memory	1.68
10	60	Out of Memory	-	-	Out of Memory	5.2

approach outperformed the 'Original model.' The number in a bracket of the 'Original model' column of 'Solution Quality' represents the gap between best solution and lower bound when the solver reached the time limit.

These computational experiments verify that our '3-phase' approach can achieve high-quality solutions in a short amount of time; therefore, we used this approach for further experiments to generate insights.

3.5.3 Patient volume and sampling

We collected state-wide trauma data across various states from published annual reports (available on state trauma websites) and observed a substantial variation across these states. The patient volume varied between 11,000 and 72,000 per year, with 3.2 to 8.2 variation in number of trauma patients per thousand citizens. Additionally, patient volume at a county level was observed to be highly correlated with the population of that county. For the experimental study, our TSA attempts to mimic the trauma patient volume of a median US state; i.e., we used 5.2 as the average trauma patients per thousand citizens and median population of a US state as 4.5 million to arrive at 23,680 trauma patients. In line with the literature, we considered 15.63% of patients as severely injured and the rest as non-severely injured.

Through preliminary experiments, we also noticed that the computational time to reach a solution was prohibitively high when considering all 23,680 patients (65 hours). Instead of aggregating patients at the county or zip level (which would lose the granularity required for our problem), we adopted a sampling approach. We selected a representative sample among these 23,680 patients such that the underlying distribution of patients (Gini index) was highly correlated with the distribution of these 23,680 patients. All other parameters were appropriately scaled.

Figure 14 illustrates the solution quality and runtime comparison at various sampling rates. To balance quality and computational time, we selected 15% as the sample

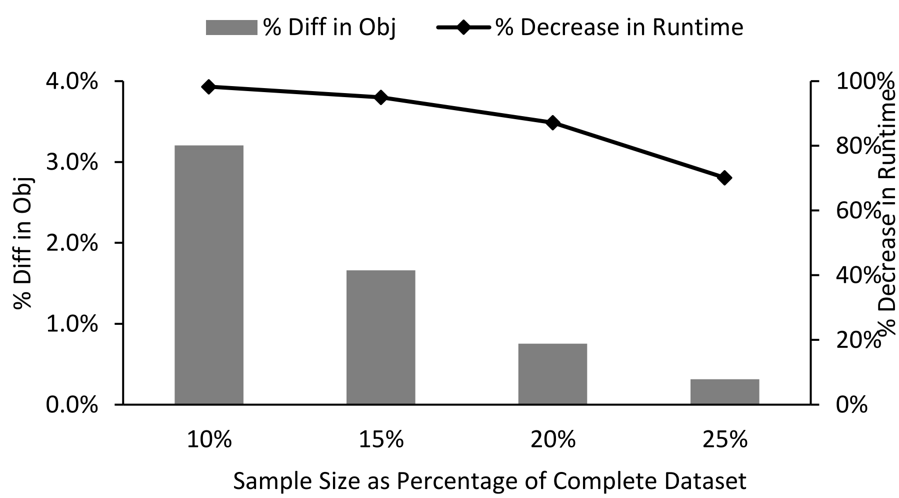


Figure 14. Solution quality and runtime for different sample size compared to complete data set

size; it reduced the time by 95% with about 1.6% difference in the solution compared to the problem being solved with complete data. Essentially, this sampling allowed us to solve problem instances (per Table 14) on average in 3.5 hours.

3.5.4 Experimental setting

During the preliminary experiments, we also noticed that the solutions appeared to be sensitive to three key factors. Table 14 summarizes these factors and their levels with bold entries in the last column indicating the base case. While ACS or state trauma agencies typically propose a protocol for destination determination based on the severity of injuries, only 40% of patients were assigned using protocol criteria according to literature and data from our collaborators. Patient choice (PC) is the second dominant criteria for destination determination, followed by assignment to the closest facility, which is inevitable. Therefore, we considered four scenarios to quantify the impact of assignment criteria on the performance and design of the network. We used the (40, 40, 20) combination as a base case based on our interactions with our trauma collaborators and the trauma literature; i.e., 40% of patients were assigned using protocol and patient choice (PC) criteria, while the remaining 20% used the closest facility criteria in all counties. Assignment criteria and injury severity are preassigned to each patient as part of the data preprocessing step.

Table 14. Summary of the parameters, levels, and values in the sensitivity analysis

Parameter	**Level**	**Values**
Percentage of assignment using protocol, patient choice and closest facility criteria	4	**(40, 40, 20)**, (60, 20,20), (80, 0,20), (100,0,0)
Distribution of trauma patients	3	Disperse (0.25), **Regular (0.5)**, Cluster (0.75)
Weights combination for equity and effectiveness	3	(0.1, 0.9), **(0.5, 0.5)**, (0.9, 0.1)

Distribution of patients in the TSA was distributed using 3 levels quantified through the Gini index, where 0 and 1 represent fully-dispersed and fully-clustered distributions. Accordingly, dispersed corresponded to Gini=0.25 (patients are less clustered and more homogeneously distributed), clustered corresponded to Gini=0.75 (patients are highly clustered around urban zones), and regular corresponded to Gini=0.5 (patients are moderately clustered around urban zones). While the dispersed distribution attempted to mimic states such as New Jersey, Delaware, and Vermont, the clustered distribution mimicked states such as Nevada, Texas, and Arizona.

Three weight combinations are used to evaluate the impact of emphasis on equity vs. effectiveness. We also use 1 and 0.5 as the γ and δ, respectively. Following ACS recommendation, we used 240 severely injured patients as a lower bound for MTCs and 60 for ITCs considering their limited resources (specialist surgeons, equipment and capacity). Additionally, per trauma literature, we used the upper bound on volume as 1,000 at both MTCs and ITCs, access time threshold as 30 minutes, bypass time threshold as 0 minutes, and C = 64 (total candidate locations). For air transport (via helicopter), we set the inbound time (time from the helicopter depot to the incidence location) as 10 minutes and the loading time as 5 minutes. Based on the range calculated from state trauma reports, we used 15% of severely injured patients as upper bound for helicopter use. We set 70% of total non-severely injured as the maximum allowable number of over-triage patients in the TSA.

For the utility model representing the patient choice, the attractiveness for MTC, ITC, and NTC is set as 5, 3, and 1, respectively, as a way to differentiate the relative perception of trauma centers among citizens. The coefficients for attractiveness and ground travel time were estimated as 0.1 and 0.05 using the optimization framework and data from the state of Ohio (see Appendix B for details).

3.5.5 Insights from the experiments

Below, we summarize key insights from our experimental study.

Insight 1: Destination determination criteria impacts patient safety; while 100% protocol usage improves it, increased use of patient choice lowers it.

As alluded earlier, the ACS and/or state trauma agencies prefer that EMS paramedics determine the destination of patients based on a protocol. Our results suggest that if all such determinations were done using this protocol (i.e., 100, 0, 0), we observed a 92.4% reduction in objective compared to the base case of (40, 40, 20) (i.e., 1.71 vs. 22.54) with fewer ITCs (1 vs. 9) and MTCs (14 vs. 15). That is, if patient choice and closest facility considerations were not part of the destination determination, the trauma network could be optimized and substantial performance benefits could be achieved.

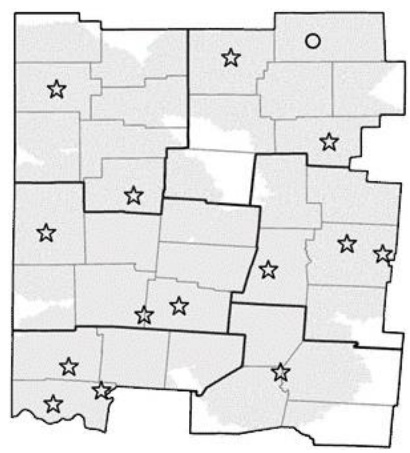


Figure 15. Dark areas indicate 30-minute access from the incidence location to at least one MTC or ITC; stars indicate MTCs and circle indicates ITC

In terms of the distribution of MTCs, we noticed a disperse distribution that accessed by most of the TSA by at least one MTC or ITC such that most severely injured patients had access to one of them within the access time, thus, reducing under-triage patients (see Figure 15). Further, a dispersed network of MTCs and ITCs also means a higher chance of having NTCs within

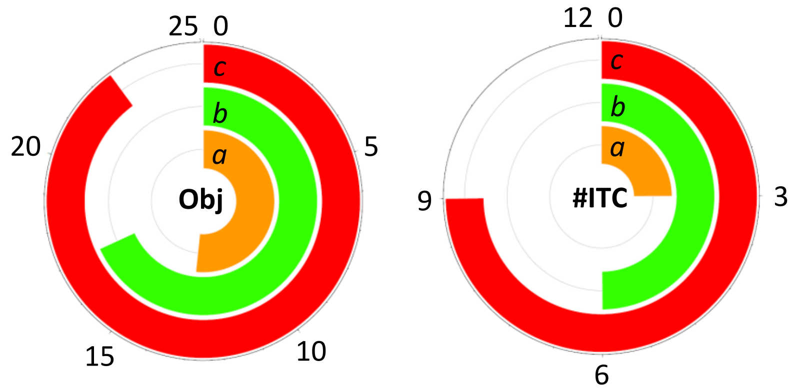
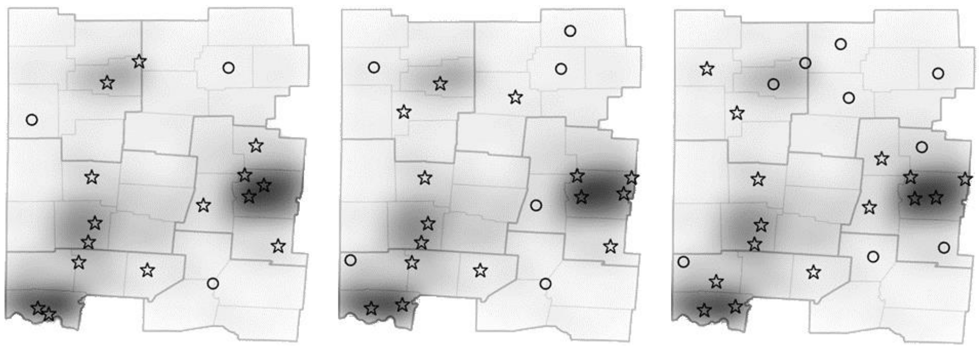


Figure 16. Objective and #ITCs for different destination assignment criteria scenarios; a= (80, 0, 20), b=(60, 20, 20), and c=(40, 40, 20), where each element represents protocol, patient choice, and closest facility, respectively

the bypass threshold for non-severely injured patients resulting in lower over-triage patients.

However, as shown in Figure 16, higher assignments through patient choice (PC) criterion require more ITCs (3 in *a* vs. 9 in *c*) and result in a 73.1% increase in objective (12.96 vs. 22.54).

To understand this further, imagine a network without ITC. Under the PC criterion, the absence of ITCs in the vicinity of the incidence location in a suburban or rural zone would leave a severely injured patient (or their family) to choose between a nearby NTC (say, at 5 min) and far located MTC (say, at 25 min). Considering a lower travel time, that patient will likely choose the NTC over MTC. This would result in that patient experience under-triage. While at least one MTC in that zone would mitigate such an under-triage, it may not be feasible due to MTC's minimum volume requirement. This is where an ITC could play a compromising role as it would likely induce the patient (or their family) to choose this ITC over an NTC, and eventually getting better care (see Figurer 17-c).



(a) (80, 0, 20) (b) (60, 20, 20) (c) (40, 40, 20)

Figure 17. Locations of MTCs and ITCs for different percentage of assignments; dense areas represent higher number of patients, stars represent MTCs, and circles represent ITCs

Insight 2: TSA with clustered distribution of patients appear to improve patient safety compared to other distributions.

To delineate different distributions of severely injured patients, we use D, R, and C to represent disperse, regular, and cluster with Gini indexes of 0.25, 0.5, and 0.75, respectively (see Figure 18). Our results indicate that as the patient distribution changes from disperse (D) to cluster (C), the overall objective decreases by 18.8% (22.42 vs. 18.5). The number of ITCs is almost three times (14 vs. 5) in the dispersed situation compared to cluster situation.

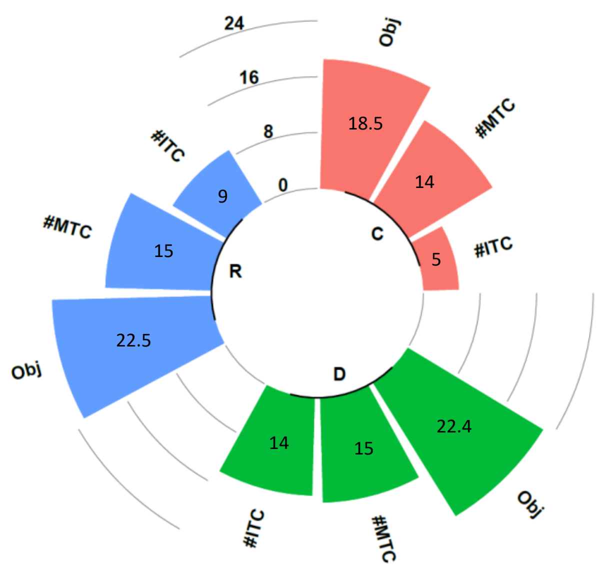
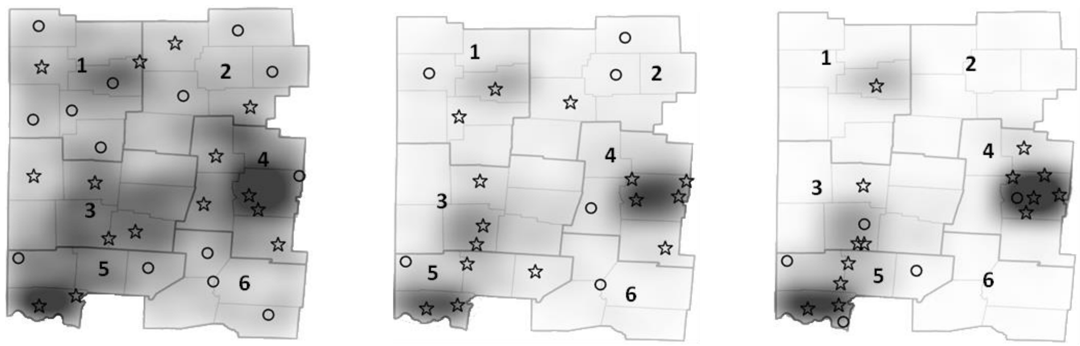


Figure 18. Objective, #MTC and #ITCs for disperse (D), regular (R) and cluster (C) distribution of patients

This is reasonable as clustered distribution increases



(a) Disperse distribution (D) (b) Regular distribution (R) (c) Cluster distribution (C)

Figure 19. Locations of MTCs and ITCs superimposed over heatmap of patient distribution

opportunities for treating more patients at the same MTCs and ITCs located around the clustered zones. However, the performance of cluster distribution highlights that better performance can be achieved with even fewer resources which is counterintuitive.

Moreover, we observed distinct location of MTCs and ITCs across three patient distribution scenarios (see Figure 19). In the cluster distribution, due to higher patients from dominantly urban regions (#3-#5), 13 out of 14 MTCs are located in those regions; however, in the dispersed situation, the MTCs are spread across the TSA. In addition, higher patients from suburban and rural zones in disperse distribution trigger the opening of ITCs in those zones as patients are still not enough to make an MTC feasible from a minimum volume perspective. As a result, 11 out of 14 ITCs are located in dominantly rural regions in dispersed distribution compared to zero in the case of cluster distribution. That is, the distribution of the patients tends to drive the number and location of MTCs and ITCs across the TSA.

Insight 3: An emphasis on equity in a network may lead to a decline in overall patient safety.

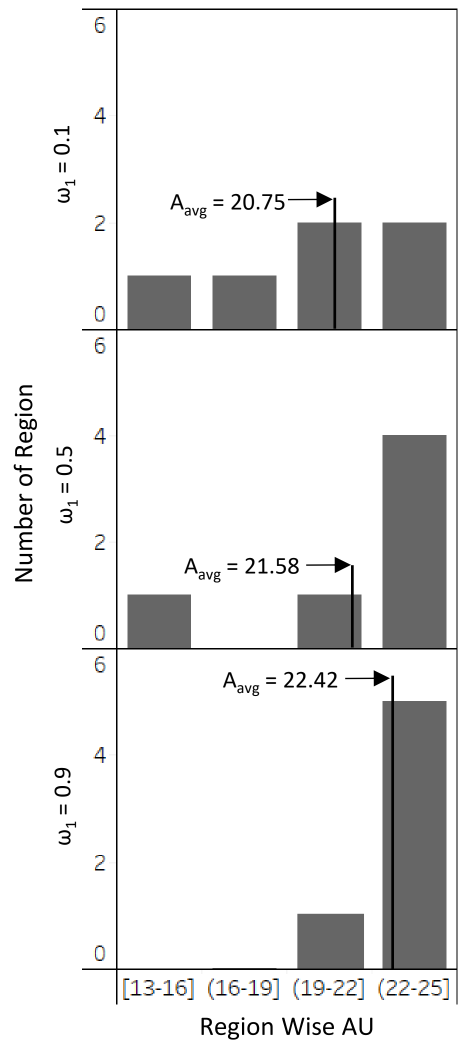


Figure 20. Region-wise AU (A_{avg}) for different values of ω_1; black line represents A_{avg} of the TSA

As expected, in an equitable network, most regions performed equally. This is evident from the histogram (depicting equity per region under ω_1 = 0.9) in Figure 20. Despite this, the performance of many regions is worse than the performance observed with lower values of ω_1 (a less equitable network). To quantify this, we used skewness of the distributions for each of the three ω_1 values. For the most equitable network (ω_1 = 0.9), the skewness was -2.4, while it reduced to -0.5 for the least equitable network (alternately, network with higher effectiveness, (ω_1 = 0.1). The corresponding TSA-wide AU (A_{avg}) increased by 8% (20.75 vs. 22.42) indicating an overall decline in the system performance. Note that an 8% increase is equivalent to an annual increase of 67 severely injured patients who will suffer aggregated under-triage (considering 23,680 data); they all could experience disabilities or mortality. The reason for this increase is due to the relocation of a few MTCs and ITCs to improve the performance of worse-performing regions. However, those relocations decrease AU of worse regions at the cost of a relatively higher increase of AU patients from the well-performing regions; the net effect is an overall increase of AU at the TSA level.

In a nutshell, results emphasize that the trauma decision maker should choose weights wisely as a higher focus on equity of patient safety can lead to higher under-triage patients, eventually increasing the likelihood of disability and mortality.

3.6 Case Study

To illustrate the practical benefit of the proposed approach, we considered the state of Ohio as a TSA and used actual data from the state for 2019. Among 71,971 trauma patients recorded in 2019, we received 17,757 de-identified patient records resulting after data linkage performed by the ODPS (Ohio Department of Public Safety). This data was further cleansed to remove missing data and unresolved addresses using ArcGIS. The resulting 11,313 patients in the cleansed dataset had a correlation of 0.99 with the 17,757 patients based on county-level case comparison, which indicated a similar spatial distribution of incidences between the original and cleansed datasets. This TSA consisted of a network of

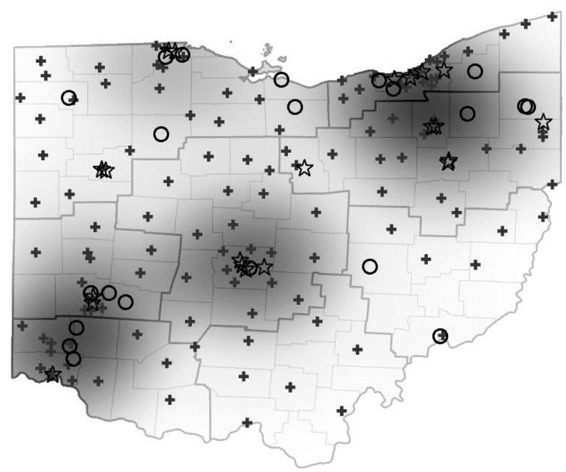


Figure 21. Trauma network in OH for 8 regions; star indicates MTCs, circle indicates ITCs and cross represents NTCs. Darker shades of grey indicate higher volume of incidences.

163 hospitals in 2019, which included 21 MTCs, 21 ITCs, and the remaining 121 NTCs. Figure 21 illustrates the heat map of 11,313 incidences and the location of hospitals.

In this data, destination determination through protocol, patient choice, and closest facility criteria were around 20%, 50%, and 30%, respectively. We empirically derived 'access'(α) *as* 25 minutes and 'bypass'(β) *as* -12 minutes such that the estimated srAU and srOT closely matched the observed values in the 20% of patients assigned through protocol criteria in the existing data. Similarly, patients assigned through patient choice criteria were used to estimate $\beta_1 = 0.1$ and $\beta_2 = 0.05$ through an optimization model shown in Appendix B. In line with the discussion in Section 5.3, we further sampled 3,552 patients (correlation of 0.999 with 11,313 data) to limit the computational burden; we scaled the MTC and ITC volume requirements accordingly. We set maximum allowable number of over-triage patients in the state as 67.31% of total non-severely injured patients (similar to observed in 11,313 data) and remaining parameters values are as used in Section 5.

Using $\omega_1 = 0.5$, we derived two optimal networks, one for the case when the number of effective MTCs is the same (Redistributed) and the other where this number is also optimally determined by the model (Greenfield). In addition, we also derived an optimal trauma network with all assignments through protocol criteria as recommended by ACS and/or state trauma agencies.

3.6.1 Existing vs. Redistributed vs. Greenfield trauma network

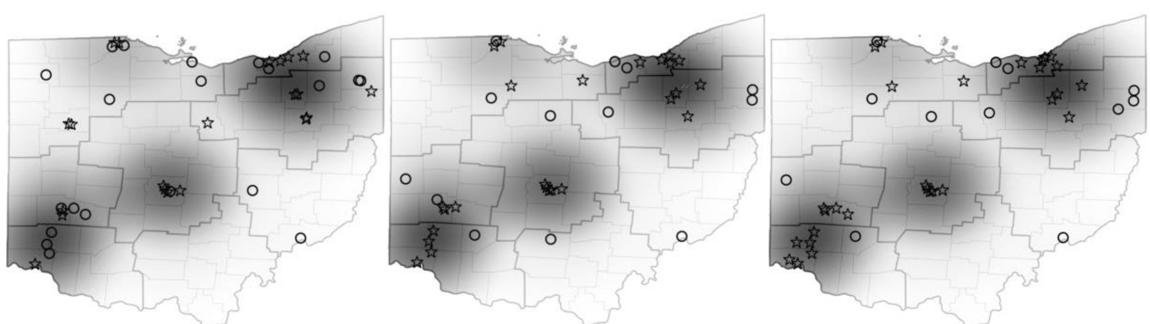
The Existing network had 31.5 effective MTCs (21 MTCs + 0.5*21 ITCs). Hence, for the Redistributed network, we set C = 31.5. However, for the Greenfield network, we

Table 15. Performance of Existing, Redistributed, and Greenfield network

Network	**Obj**	**Average srAU**	**Max srAU**	**# MTC**	**# ITC**	**Effect. MTC**
Existing	16.50	10	23	21	21	31.5
Redistributed	9.69	6.88	12.5	25	13	31.5
Greenfield	9.59	6.69	12.5	29	12	35

let C = 163 (all candidate locations) allowing the model to select as many MTCs and ITCs it needs to minimize the objective function. Table 15 summarizes the performance of the three networks, while Figure 22 illustrates the trauma network superimposed on heat maps of incidences.

The Gini index for the 2019 data is 0.751, representing a clustered distribution of patients. In the Existing network, many ITCs were observed in the clustered areas (darker areas in Figure 22); however, the Redistributed and Greenfield networks, we observed several MTCs in those clustered areas in line with Insight 2. Due to these additional MTCs, redistribution reduces the average srAU patients by 31.2% compared to the Existing network (10 vs. 6.88). The Greenfield network only did slightly better than the Redistributed network; the average srAU patients decreased by 33.1% (10 vs. 6.69) at the cost of an additional 3.5 effective MTCs. Further, in both the Redistributed and Greenfield



(a) Existing 2019 network (b) Redistributed network (c) Greenfield network

Figure 22. Comparison of trauma networks (darker shades indicate higher volume of incidences; stars indicate MTCs; circles indicate ITCs)

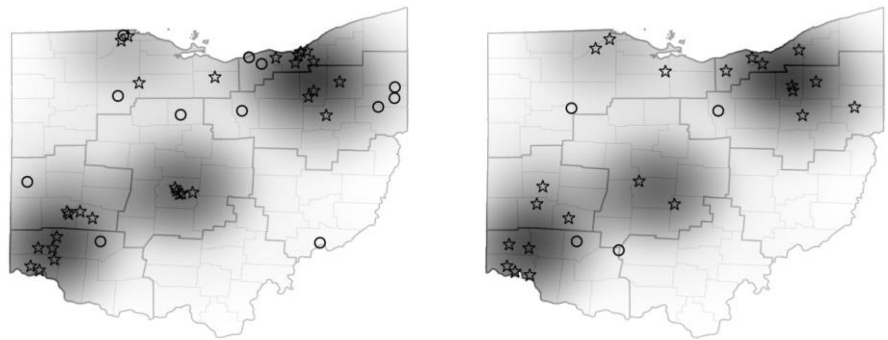
networks, ITCs were located in moderately dense areas (light grey areas in Figure 22) as highlighted in Insight 1.

Overall, our results indicate that for the same number of effective MTCs, the potential to improve patient safety is considerably high in the Redistributed approach. Further, the law of diminishing returns applies in the Greenfield network, where an increase from 31.5 to 35 effective MTCs does not yield a significant benefit in the performance. However, the Greenfield solution can enable benchmarking of existing, redistributed, or any other network that the state trauma decision makers may be considering.

3.6.2 Greenfield network with 100% protocol criteria

Figure 23 compares the Greenfield network with two destination determination allocations over 11,313 patients. We observed that a network with 100% protocol criteria results in a reduction in the objective by more than 50% (9.59 vs. 4.75) compared to a network with 20% protocol, 50% patient choice and 30% closest facility (similar to the Existing network). This observation is similar to Insight 1. We did observe that 100% protocol led to fewer number of effective MTCs (24 vs. 31.5; MTCs increase to 29 from 22, but ITCs decreased to 4 from 12).

Further, maximum srAU among all regions decreased by 52% (a 47.7% reduction in average srAU patients), which corresponds to 27 ({6.89-3.5}*8) fewer severely injured patients who would suffer a mistriage (srUT or $srUT^s$) in sample (3,552) data. Considering 71,971 patients reported in 2019 in OH, this would correspond to 547 fewer patients annually, which is substantial. Clearly, following ACS recommendation of using protocol as the primary criteria can lead to substantial benefits in patient safety; however, this will



a) 20% protocol, 50% patient choice, 30% closest facility (similar to Existing network)

b) 100% protocol

Figure 23. Optimal trauma network superimposed on heat maps of incidence

require the state to introduce new EMS training initiatives to promote protocol, while attempting to mitigate other reasons for destination determination.

3.7 Summary

Our research introduced the nested trauma network design problem (NTNDP) that determines the number and location of major, intermediate, and non-trauma centers for a trauma service area. The NTNDP minimizes a weighted sum of equity among regions and effectiveness across the TSA. Several practical considerations, compared to existing trauma literature, were incorporated in the NTNDP, such as multiple patient types, multiple choices for transportation, multiple destination determination criteria and multiple level of hospitals. Specific to the trauma literature, NTNDP generalizes TCLP by including intermediate trauma centers, a vital element of a trauma network that improves access to trauma care for regions that do not have access to a major trauma center. The inclusion of three dominant destination determination criteria ensures faithful representation of the on-scene EMS decision making process. In addition, consideration of both equity and effectiveness in the objective allow trauma decision makers to trade off these two contradicting measures in network design.

We modeled the NTNDP as a mixed-integer linear program and proposed a '3-Phase' solution approach to find near-optimal solutions in a reasonable amount of time. We generated a TSA using data from a state trauma system in the US and quantified the impact of system parameters on the performance of the trauma system. We illustrated the use of the proposed approach on 2019 data for the state of Ohio. The key findings from our study are as follows:

- If EMS providers could exclusively use 'protocol' as the criterion for destination determination at the incidence location, then a trauma network with low levels of under-triages can be realized with fewer MTCs and ITCs. Increased use of the 'patient choice' criterion could result in higher number of ITCs required in suburban and rural zones and increased under-triages.
- A clustered distribution of severely injured patients in the TSA appears to improve trauma system performance with fewer MTCs and ITCs.
- Solely focusing on equity of patient safety among regions as an objective function appears myopic; balancing it with effectiveness across the TSA appears to result in a better performing network.
- Illustration of our approach on real data from a midwestern US state indicated an over 30% improvement in patient safety; Greenfield network can enable benchmarking of existing, redistributed, and other networks. Importance of 100% use of protocol for destination determination was also verified.

These findings have several practical implications. Trauma decision makers can use our approach to comprehend the compromised role offered by ITCs on patient safety (via provision of intermediate care), especially in suburban and rural zones where MTCs are financially not viable (due to a low number of patients). Further, they can quantify the impact of various destination determination criteria used in practice on patient safety. They can, subsequently, design training programs for EMS providers that help them employ ‘protocol’ (which is the preferred approach suggested by ACS) during on-scene decision making

CHAPTER 4

NETWORK DESIGN PROBLEM CONSIDERING ASSESSMENT MISTRIAGES (TNDP-AM)

4.1 Trauma Injuries and Decision-making Processes

Traumatic injury continues to be a major public health problem, with worldwide 4.4 million deaths each year (nearly 8% of all deaths). It continues to be the #1 cause of death, disability, and morbidity for individuals aged 44 and under in the US (#3 across all ages) with almost 200,000 deaths. The corresponding economic burden is over $4.2 trillion annually (WHO, 2021; CDC, 2021; CDC, 2022).

In response to this mortality and economic burden, many state agencies have established integrated and coordinated trauma care system for their state. To ensure the continuum of trauma care, these systems have three major phases (prehospital, acute care, and rehabilitation). The 'prehospital' phase activates from the moment a call is made to 911 to seek help for the injured victim (patient). The emergency medical dispatcher coordinates with various Emergency Medical Services (EMS) nearby and dispatches appropriate EMS to the incidence location (also known as the scene). EMS paramedics stabilize the patient, assess their vital conditions with the aid of various diagnostic tests and medical protocols to determine injury severity, and the appropriate destination hospital and transportation mode for transport to that hospital.

Once a patient arrives at the hospital (the ‘acute care’ phase), the trauma response team assesses and treats their life-threatening injuries and transfers them to an operating room or intensive care according to their condition. Hospitals are verified by the American College of Surgeons (ACS) as major trauma centers (MTCs) if they are capable of providing definitive care for patients suffering severe injuries. Otherwise, they are deemed as non-trauma centers (NTCs), which are the ideal destination for non-severely injured trauma patients. In case the hospital is not capable of treating the patient's injury, the patient is further transferred to the appropriate hospital, known as a secondary transfer. After the patient stabilizes, they are moved to the general care unit before being discharged. Some patients may be sent to a rehabilitation program depending on the type of injury (i.e., the ‘rehabilitation’ phase).

It is the prehospital phase of the trauma care system that is critical to providing the right care to the right patient at the right time at the scene to improve chances of survival and avert life-long disabilities (Van Laarhoven et al., 2014; Van Rein et al., 2019; Van Rein et al., 2020). Three critical decisions impact the prehospital elements of the trauma care system: (i) EMS network and dispatch, (ii) on-scene injury assessment, and (iii) destination determination. There is a vast literature that focuses on (i); i.e., the location, relocation, and dispatch of EMS. For reviews on (i), see Brotcorne et al., 2003; Aringhieri et al., 2017; Bélanger et al., 2019.

4.1.1 On-scene decision-making and mistriages

Once the EMS has reached the incidence location, subsequent decisions (i.e., on-scene assessment and destination determination) become vital. On-scene assessment of an injured patient can be challenging and time-sensitive. To assist EMS paramedics during this process, national and state organizations have proposed guidelines, known as Field Triage Guidelines (FTG), using the best available evidence and expert consensus (Van Rein et al., 2018a; Lupton et al., 2022). An accurate on-scene injury assessment is crucial for timely patient care as assessed severity becomes the basis for the destination determination decision (right care at the right time). The latter depends upon the location of trauma centers in the vicinity and available transportation modes.

Mistriages in on-scene injury assessment often occur during both decision-making processes and can induce patient safety issues, such as short/long-term disability, morbidity, and even mortality. Mistriages (both under- and over-triage) have been preferred surrogate measures for patient safety in the healthcare literature (Jansen et al., 2015; Hirpara et al., 2022, Parikh et al., 2022). We use 'assessment-related' to refer to mistriages caused during injury assessment and 'system-related' to refer to mistriages during 'destination determination.' However, none of the prior research explicitly considered assessment-related mistriages and analyzed the impact of change in those mistriages on patient safety, which form the basis of our research. We first describe what these assessment-related mistriages and illustrate how they impact patient safety, before presenting our research questions.

4.1.2 Inaccuracy in injury assessment

Even though FTGs are developed by medical experts and updated periodically, they are not accurate in classifying the severity of a patient's injury on the scene. Several reasons for the inaccuracy have been reported in the literature based on field studies and retrospective data analysis; e.g., human error due to lack of training, internal injuries that are hard to assess, faulty diagnosis tests (Newgard et al., 2011a; Newgard et al., 2011b; Van Rein et al., 2020), and inability to capture regional factors (Parikh et al., 2017). Healthcare literature estimates 5 to 95% of injury assessment-related mistriages using various FTGs (Van Rein et al., 2018a; van Rein et al., 2018b; Gianola et al., 2021; Lupton et al., 2022). Statistical models have been proposed to reduce assessment-related mistriages at the scene (Newgard et al., 2013; Parikh et al., 2017; Van Rein et al., 2019; Larsson et al., 2021).

Given that the outcomes of this clinical assessment of injuries can be severe or non-severe, there are four possibilities; we use the modifier 'assessment-related' to represent these:

- Assessment-related appropriate triage positive ($arAT^P$): A patient with underlying *severe* injury *assessed severe* at the scene.
- Assessment-related appropriate triage negative ($arAT^N$): A patient with underlying *non-severe* injury *assessed non-severe* at the scene.
- Assessment-related under-triage (arUT): A patient with underlying *severe* injuries *assessed as non-severe* (Voskens et al., 2018; Shanahan et al., 2021; Lupton et al., 2022).

- Assessment-related over-triage (arOT): A patient with underlying *non-severe* injuries assessed as *severe*.

4.1.3 Effective triage

Considering the occurrences of mistriages at both on-scene decision-making processes (injury assessment and destination determination), the decision process may lead to 8 possible outcomes as shown in Figure 24. Note that we use the same definitions of 'system-related' mistriages (caused during destination determination) as suggested in prior literature (Hirpara et al., 2022; Parikh et al.; 2022); i.e., 'system-related' appropriate triage (assessed severely injured taken to MTC - $srAT^P$ and assessed non-severely to NTC - $srAT^N$), 'system-related' under-triage (assessed severely injured taken to NTC - srUT), and 'system-related' over-triage (assessed non-severely taken to MTC - srOT).

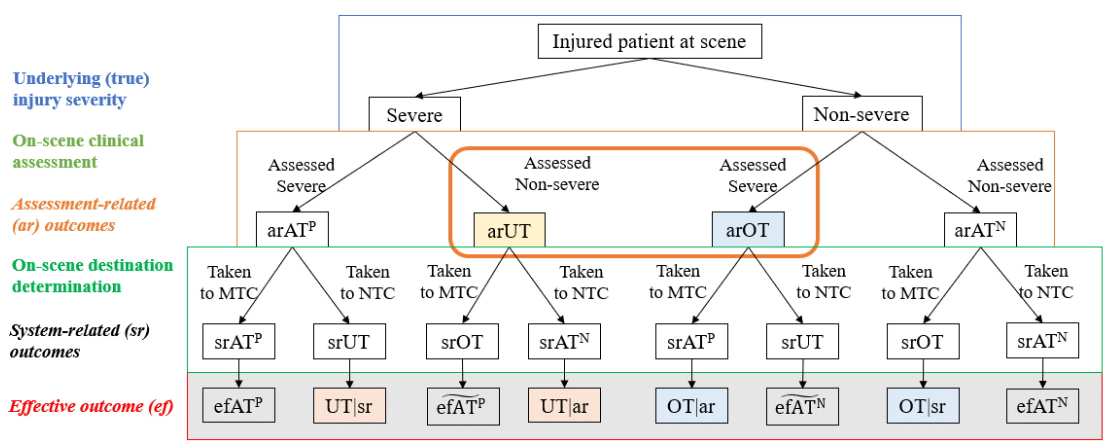


Figure 24. Outcomes based on mistriages in on-scene clinical assessment and choice of destination for a given underlying (true) injury severity

We aggregate these 8 outcomes into three groups based on whether or not the final hospital where the patient was transported to was the appropriate hospital to treat this patient's true underlying injury severity; we refer to them as *effective triage*:

- Effective appropriate triage (efAT): These include patients with *underlying severe* injuries transported to *MTCs* ($\mathrm{efAT}^{\mathrm{P}}$ and $\widetilde{\mathrm{efAT}^{\mathrm{P}}}$) and *underlying non-severe* injuries transported to *NTCs* ($\mathrm{efAT}^{\mathrm{N}}$ and $\widetilde{\mathrm{efAT}^{\mathrm{N}}}$). The '~' symbol represents the situation where a patient's injury was incorrectly assessed (resulting in either arUT or arOT), but was subsequently taken to an incorrect hospital (resulting in srOT or srUT, respectively), in turn resulting in the patient eventually making it to the correct hospital based on their true underlying injury.
- Effective under-triage (efUT): This group includes patients with *underlying severe* injuries transported to *NTCs* (not an ideal hospital) due to either assessment-related mistriage (UT|ar) or system-related mistriage (UT|sr). These patients could suffer from short/long-term disability, morbidity, and even mortality due to delay in the definitive care (Rotondo et al., 2014).
- Effective over-triage (efOT): Patients with *underlying non-severe* injuries transported to *MTCs* are referred to as efOT (OT|ar and OT|sr). Effective over-triage patients cause overcrowding at MTCs, higher medical bills due to unnecessary trauma activation, and occupy resources, which can delay care for other patients (Frykberg, 2002; Armstrong et al., 2008).

4.1.4 Research questions and contributions

Our focus in this research is on arUT and arOT mistriages (due to clinical assessment) and its eventual effect on efUT and efOT, and the network of trauma hospitals. It is worth noting that there is a void in the literature in exploring the impact of assessment-related mistriages on patient safety and/or network design. We address the following questions through our work.

- How do assessment-related mistriages during the on-scene injury assessment phase (an operational decision) impact the network of major trauma centers (a strategic decision)?
- How substantial are the effects of assessment-related mistriages in the ISS 9-15 group (moderate injuries) compared to the ISS>15 group (severe injuries) on patient safety?

The key contributions of our approach are as follows. First, to address the above questions, we propose a stochastic nested multi-level, multi-transportation capacitated model to maximize patient safety. Second, our proposed model explicitly considers mistriages in injury assessment by incorporating the uncertainty in the binary assessment decision (severe vs. non-severe) as a Bernoulli random variable with the probability of success referring to a patient being classified as having severe injuries. Third, we integrate Monte Carlo Simulation with a genetic algorithm (GA) to solve the proposed stochastic model efficiently. We incorporate a feasibility algorithm in the proposed GA to convert infeasible solutions during offspring generation into feasible ones. Finally, we generate a dataset using state-wide annual trauma registry data published across the US to evaluate the sensitivity of the solution to mistriages in injury assessment.

Our findings suggest that the network is sensitive to mistriages in the ISS >15 group (i.e., underlying severe injuries). The resulting network with high mistriages in this group increases efUT by 799% compared to the network without assessment mistriages. In addition, it may also lead to the clustering of MTCs near high trauma incidence rates. The trauma network is also sensitive to mistriages in the ISS 9-15 group, and the resulting network tends to have a fewer and dispersed distribution of MTCs.

In the following sections, we first summarize the relevant literature in Section 2. We introduce the problem and present the mathematical model in Section3. Next, in Section 4, we provide details of the proposed solution approach and insights based on an experimental study in Section 5. Finally, Section 6 summarizes our work and offers guidance for future research in this area.

4.2 Literature Review

Several studies have addressed a variety of healthcare facility location problems, such as the location of primary health centers (Günes et al., 2014), long-term care centers (Cardoso et al., 2015; Intrevado et al., 2019), preventive healthcare facilities (Zhang et al., 2009; Zhang et al., 2010), blood bank locations (Çetin & Sarul, 2009), organ-transplant centers (Caruso & Daniele, 2018) and ambulance location and/or relocation (Reuter-Oppermann et al., 2017; Vanbuuren et al., 2018). For a comprehensive review, see Reuter-Oppermann et al. (2017), Ahmadi-Javid et al. (2017), and Gunes et al. (2019).

In the area of trauma network design, approaches proposed have focused on maximizing patient safety due to the time-sensitive and life-threatening nature of traumatic injuries. Three types of surrogate metrics for patient safety have been used in this literature: (i) coverage, (ii) access time, and (iii) mistriages. Branas et al. (2000) proposed a model to maximize coverage of severely injured patients by simultaneously locating major trauma centers and air ambulances. Cho et al. (2014) further incorporated a busy fraction of medical helicopters in their model that simultaneously locates major trauma centers and medical helicopters to maximize the expected number of severe patients transported to an MTC within 60 minutes. Lee et al. (2018) extended this model and proposed a multiperiod location model that determines when and where to locate trauma centers and air ambulances over a planning horizon. However, all of these approaches assume 100% accuracy in on-scene injury assessment by the EMS and do not consider non-severely injured patients.

In terms of considering access time and mistriage as bi-objective, Jansen et al. (2015) proposed a data-driven bi-objective approach to minimize the total access time and the number of exceptions for severely injured patients by locating major and intermediate trauma centers. Jansen et al. (2018) further designed Colorado's trauma network with the same objective using a multi-fidelity surrogate-management strategy developed by Wang et al. (2016) for offline data-driven multi-objective optimization problems that reduce the computation time. Both models used the triage protocol proposed by ACS to determine the injury severity and clinical need (different levels of trauma care); however, they do not explicitly account for injury assessment-related mistriages. In addition, these approaches do not consider non-severely injured patients and associated mistriages, and also restrict the network design to only downgrading an existing MTC.

To support decision making around trauma networks, the ACS Committee on Trauma (ACS COT) developed the Needs-Based Assessment of Trauma System (NBATS) tool that determines the required number of major trauma centers using six criteria (ACS-NBATS, 2015). However, the NBTAS tool does not determine the location of those major trauma centers and the network's performance. Parikh et al. (2022) proposed a model for a Performance-based Assessment of Trauma System (PBATS) that determine the minimum number and location of MTCs by keeping system-related under-triage (srUT) and over-triage (srOT) rates within a prespecified limit. Hirpara et al. (2022) proposed a bi-objective model that maximizes patient safety by optimally locating major trauma centers. Patient safety is measured as the weighted sum of srUT and srOT rates; mistriages occur during the destination determination phase. These works consider both types of patients; however, they do not explicitly account for on-scene injury assessment-related mistriages.

Literature has also focused on estimating the accuracy of field triage guidelines (FTGs) proposed by national and state agencies worldwide using retrospective and prospective studies. The ACS developed a field triage guideline known as the National Field Triage Decision Scheme (FTDS) more than three decades ago and has periodically revised it (ACS COT, 1986; ACS COT, 2006; Sasser et al., 2011). Newgard et al. (2011b) estimated the performance of 2006 FTDS through a retrospective study using data from 7 regions of the Western US from 2006 through 2008 and estimated 14.2% and 31.3% arUT and arOT rates, respectively. Newgard et al. (2016) sought a prospective validation of 2006 FTDS through a study of 1 year in 7 counties in 2 states, and estimated 33.8% and 12.2% arUT and arOT rates, respectively. Parikh et al. (2017) assessed the performance of the 2011 FTDS and Ohio Prehospital Trauma Triage Decision Tree (OPTTDT) through a retrospective analysis of 5 years of data from the state of Ohio. They observed 9.09% and 9.57% arUT and 87.45% and 86.99% arOT rates for FTDS and OPTTDT, respectively. An analysis of the Dutch Field Triage protocol using a retrospective study of 2 years of data from central Netherlands resulted in 63.8% arUT and 7.4% arOT rates. See Van Rein et al. (2018a), van Rein et al. (2018b), Gianola et al. (2021), and Lupton et al. (2022) for comprehensive reviews of the performance of various FTGs developed worldwide.

Considering the lower accuracy rates of the FTGs, several studies have proposed statistical models to improve assessment quality using various medical characteristics known at the scene. Parikh et al. (2017) developed a multivariate logistic regression-based prediction model using 5-year data from the state of Ohio. They observed considerably better injury severity estimates through their proposed model compared to 2011 FTDS (arUT = 1.93% versus 9.03%; arOT = 66.42% versus 87.52%). Van Rein et al. (2019) developed a bivariable logistic regression model based on prehospital predictors associated with injury using the data from the central Netherlands region. The developed model uses only 8 prehospital parameters (vs. an average of 40 in FTGs) to estimate injury severity. Their model resulted in an arUT rate of 11.2% and an arOT of 50.0%, which is 10.4% and 19.4% lower than the Dutch Field Triage Protocol for the same data.

Our review of existing trauma literature suggests the following gaps:

- Prior work assumes 100% accuracy in on-scene clinical injury assessment by EMS and does not explicitly account for mistriage rates in assessments; this limits our understanding of how these mistriage rates impact patient safety.
- While a number of alternate models for on-scene injury assessment have been proposed, none have attempted to quantify the impact of their quality on the trauma network.

This work fills the above gaps by proposing a stochastic nested trauma network design model to determine the optimal number and location of trauma centers in order to maximize patient safety. The proposed model considers both severe and non-severe patients, and explicitly accounts for on-scene injury assessment-related mistriages. We now present details of our approach.

4.3 Optimization Model

Trauma care systems are often designed for a specific Trauma Service Area (TSA), which is a geographical area comprising a collection of counties in a state, the state itself, or even a collection of states. Typically, a TSA is further divided into subareas known as regions (or districts), which oversee trauma care within that region. The proposed generic model explicitly accounts for assessment-related mistriages and determines the optimal number and location of MTCs within a TSA in order to maximize overall patient safety (quantified in terms of effective under-triage). As mentioned earlier, due to severe life-threatening injuries, delay in definitive care for under-triage patients increases the likelihood of adverse outcomes; therefore, we consider effective under-triages (efUT) as the primary patient safety metric while effective over-triages (efOT) as a secondary metric.

Although a true underlying injury severity is often classified as severe or non-severe (binary) in the literature, non-severe injuries (ISS ≤ 15) are further classified into minor (ISS 1-8) and moderate (ISS 9-15) injuries, with the latter often complex to separate them from severe injuries (ISS>15). We, therefore, consider such subgroups to further understand the effect of assessment mistriages in each of these groups on system performance.

For a given ISS group, an assessment-related mistriage (arUT and arOT) represents the % of patients that have been assessed to have injury that is different from their true underlying injury. That is, in case of ISS 1-8 and 9-15 groups (i.e., underlying injuries are non-severe), an assessment-related mistriage (arOT) means the patient is assessed to have severe injuries. However, for the ISS >15 group (i.e., underlying injuries are severe), an assessment-related mistriage (arUT) refers to the patient assessed to have non-severe injuries. We model these binary mistriages via group-specific Bernoulli distributions with probability of injury perception being severe as x for ISS 1-8, y for ISS 9-15, and (1-z) for ISS >15 group. Note that a mistriage in ISS > 15 group means a severely injured patient was mistriaged as non-severe; so if z refers to the probability of assessed non-severe (a mistriage - arUT), then (1-z) refers to assessed severe (which is what we need for the Bernoulli distribution for this group).

For given assessed injury severity, we use the Notional Tasking Algorithm proposed in Hirpara et al. (2022) to determine the destination hospital, mode of transportation, and resulting system-related triage type.

Given this background, we now present our model under the following assumptions:

- True injury severity is known only through the ISS score recorded during the patient's hospital stay.
- The number of patients and their locations are deterministic and known.
- In the case of multiple patients at the scene (say, during a multi-vehicle crash), each patient is evaluated individually and transported based on their medical condition.
- The candidate locations for the MTCs and NTCs are known and finite.
- If a severely injured patient is initially transported to an NTC, they are eventually transferred via ground to the nearest MTC from the incidence location for definitive care.
- While ground ambulance services are available without constraints, the availability of air ambulances is restricted, and the trips are only allowed to transport assessed severely injured patients.

- Ground and air transport times are known and deterministic.

Tables 16 and 17 summarize the parameters and decision variables, respectively, used in the model.

Table 16. Parameters in the model

Notation	*Definition*
I	Set of incidences for trauma patients; $i \in I$
J	Set of candidate locations (for MTC and NTC); $j \in J$
α	'Access' time threshold to determine srUT (in minutes)
β	'Bypass' time threshold to determine srOT (in minutes)
S_i^T	True underlying injury severity of patient *i*, 1 if severely injured (ISS>15); 0, otherwise (non-severe)
$\widetilde{S_i^P}$	EMS's on-scene assessment of severity of patient *i*; 1 if assessed severely injured; 0 otherwise (assessed non-severe)
T_{in}, T_{load}	Inbound time from base-to-incidence location and loading time of patient at the incidence location for an air ambulance
Z	Maximum allowable patients via air ambulance
V^{min},V^{max}	Minimum and maximum allowable volume of underlying severely injured patients at MTC
C	Maximum number of allowable MTCs in the TSA
TG_{ij}, TA_{ij}	Travel time from patient *i* to any candidate location *j* via ground and air
OT^{max}	Maximum allowable effective over-triage patients calculated as the proportion of underlying non-severe patients

Table 17. Decision variables in the model

Notation	*Definition*
x_j	1, if a candidate location *j* is designated to be an MTC; 0, otherwise (designated as an NTC)
$n_{ij}^{MTC\_G}$	1, if location *j* is marked as MTC and is the nearest MTC for patient *i* via ground; 0, otherwise
$n_{ij}^{MTC\_A}$	1, if location *j* is marked as MTC and is the nearest MTC for patient *i* via air; 0, otherwise
n_{ij}^{NTC}	1, if location *j* is marked as NTC and is the nearest NTC for patient *i* via ground; 0, otherwise
$srAT_{ij}^{P}$	1, if patient *i* is assessed as severely injured and transported to MTC *j* via ground or air; 0, otherwise
$srUT_{ij}$	1, if patient *i* is assessed as severely injured and transported to NTC *j* via ground; 0, otherwise
$srAT_{ij}^{N}$	1, if patient *i* is assessed as non-severely injured and transported to NTC *j* via ground; 0, otherwise
$srOT_{ij}$	1, if patient *i* is assessed as non-severely injured and transported to MTC *j* via ground; 0, otherwise
$UT_{ij}\|sr$	1, if patient *i* has underlying severe injuries and is also assessed to be severe, but still transported to NTC *j*; 0, otherwise
$UT_{ij}\|ar$	1, if patient *i* has underlying severe injuries but is assessed to be non-severe (incorrect assessment) and transported to NTC *j*; 0, otherwise
$OT_{ij}\|sr$	1, if patient *i* has underlying non-severe injuries and is also assessed to be non-severe, but transported to MTC *j*; 0, otherwise
$OT_{ij}\|ar$	1, if patient *i* has underlying non-severe injuries but is assessed to be severe (incorrect assessment) and transported to MTC *j*; 0, otherwise
v_j	Volume of underlying severe patients if location *j* marked as MTC; 0, otherwise

Minimize: $E\left[\sum_i \sum_j UT_{ij}|sr\right] + E\left[\sum_i \sum_j UT_{ij}|ar\right]$

Subject to:

Determining the nearest NTC and MTC

$$\left(n_{ij}^{MTC\_G}, n_{ij}^{MTC\_A}, n_{ij}^{NTC}\right) = f_1(\widetilde{S_i^P}, TG_{ij}, TA_{ij}, x_j); \forall i \in I, \forall j \in J \quad (1)$$

Determining the system-related triage of patients using on-scene assessed severity and notional tasking algorithm

$$\left(srAT_{ij}^{P}, srAT_{ij}^{N}, srUT_{ij}, srOT_{ij}\right) = f_2\left(\widetilde{S_i^P}, T_{in}, T_{load}, \alpha, \beta, Z, n_{ij}^{MTC\_G}, n_{ij}^{MTC\_A}, n_{ij}^{NTC}\right); \quad (2)$$
$$\forall i \in I, \forall j \in J$$

Determining the effective mistriages

$$(UT_{ij}|ar, OT_{ij}|ar, UT_{ij}|sr, OT_{ij}|sr) = f_3(S_i^T, \widetilde{S_i^P}, srAT_{ij}^P, srAT_{ij}^N, srUT_{ij}, srOT_{ij}); \forall i \in I, \forall j \in J \quad (3)$$

Determining the volume at MTC based on the true underlying severity of patients

$$v_j = f_4(S_i^T, srAT_{ij}^P, srUT_{ij}, srOT_{ij}, srAT_{ij}^N, \; n_{ij}^{MTC\_G}); \forall j \in J \quad (4)$$

Allowable number of MTCs, and their minimum and maximum volume

$$x_j\, V^{min} \leq v_j \leq x_j\, V^{max}; \forall j \in J \quad (5)$$

$$\sum_j x_j \leq C \quad (6)$$

Allowable maximum effective over-triage

$$\sum_i \sum_j (OT_{ij}|sr + \; OT_{ij}|ar) \leq \; OT^{max} \quad (7)$$

Bounds on decision variables

$$x_j, UT_{ij}\big|sr, UT_{ij}\big|ar, OT_{ij}\big|sr, OT_{ij}\big|ar, n_{ij}^{MTC_G}, n_{ij}^{MTC_A}, n_{ij}^{NTC}, srAT_{ij}^P, srAT_{ij}^N, srUT_{ij} \in \{0, 1\}; \; \forall i \in I, \forall j \in J \quad (8)$$

$$v_j \geq \; 0; \forall j \in J \quad (9)$$

The proposed model minimizes expected effective under-triage (*efUT*) patients in the TSA, which comprises of UT|sr and UT|ar. Constraints (1) determine the nearest MTC and NTC via ground for each patient, and the nearest MTC via air for assessed severe patients; this is depicted as function f_1 and can be expressed in closed analytical form (similar to Hirpara et al., 2022). The function f_2 represents the notional tasking algorithm that uses several system parameters, such as assessed severity of injury ($\widetilde{S_i^P}$), nearest MTC and NTC for each patient (depend on trauma network x_j), clinically recommended thresholds (α and β), and helicopter-related parameters (see Hirpara et al., 2022). Constraints (2) classify each patient into one of the 4 system-related triage types; under-triage (*srUT*), over-triage (*srOT*), appropriate-triage to MTC ($srAT^P$), and appropriate-triage to NTC ($srAT^N$).

Constraints (3) determine effective mistriages (*efUT* and *efOT*) based on true underlying severity and system-related triage. Recall effective mistriage occur either due to system-related mistriage (UT|sr, and OT|sr given correct assessment) or due to assessment-related mistriage (UT|ar, and OT|ar) in the first place (see Figure 24). Constraints (4) estimate the total volume of underlying severe patients if an MTC is located at a candidate location. Any underlying severe patient primarily transported to MTC ($efAT^P$ and $\widetilde{efAT}^P$ in Figure 24) are considered toward trauma volume at MTC. However, all underlying severely injured patients primarily transported to NTCs (UT|sr and UT|ar) are typically transferred from an NTCs to the nearest MTCs (on the same day or later after the patient stabilized) for definitive care; such patients are also considered towards the volume of MTC.

Constraints (4) specify the bounds on the allowable minimum and maximum volumes of severely injured trauma patients for an open MTC. The maximum allowed MTCs in the TSA is specified by Constraint (5). Constraints (6) limit maximum allowable *efOT* patients, which is a summation of correctly assessed non-severe patients taken to MTC (OT|sr) and incorrectly assessed non-severe taken to MTC (OT|ar). Finally, Constraints (7)-(8) specify bounds on the decision variables.

The proposed model is a stochastic, nested, multi-level, multi-transportation capacitated model. Such problems are combinatorial in nature and have been shown to be NP-hard (Daskin, 2011). Even a deterministic version of such a problem is difficult to solve optimally using commercial software such as CPLEX and Gurobi. Therefore, we propose a Simheuristic approach, which has been used to solve stochastic combinatorial optimization problems in different fields (Juan et al., 2015). We describe details of our Simheuristic-based solution approach in the next section.

4.4 Simheuristic

A Simheuristic approach is an integration of simulation (any of its variants) and metaheuristic to solve a real-life problem with uncertainty. Their popularity has increased recently due to their flexibility in solving various stochastic combinatorial optimization problems. Successful applications of Simheuristic include vehicle routing problem with stochastic demands (Juan et al., 2013), flow-shop problem with stochastic processing times (Juan et al., 2014), stochastic uncapacitated facility location problem (De Armas et al., 2017), and waste location-routing problem (Rabbani et al., 2019); see Juan et al., 2015 and Amaran et al., 2016 for a comprehensive review.

4.4.1 Proposed Simheuristic approach

In our proposed Simheuristic, we use a Genetic Algorithm (GA) as the underlying metaheuristic and Monte Carlo simulation (MCS) to capture uncertainty using repeated random sampling (Juan et al., 2013; Juan et al., 2014; Gruler et al., 2018; Rabbani et al., 2019; Quintero-Araujo et al., 2021; Sadrani et al., 2022). While a Simheuristic with GA has been proposed in the literature (Yoshitomi et al., 2003; Edison et al., 2011; Rabbani et al., 2019; Slama et al., 2021), the novelty in our implementation springs from two enhancements that help reduce computation time while achieving high quality solutions: (i) feasibility algorithm, which refers to converting infeasible offspring to feasible ones (as elaborated in Section 4.2) (ii) elite population, that uses evolution (to collect promising offspring) and refinement (to select the best among these).

Figure 25 illustrates a flowchart of our proposed Simheuristic approach, and the following sub-sections describe the various steps in detail.

4.4.1.1 Population initialization

We use binary string to form a chromosome indicating a network of MTCs with the following representation: H = {0, 1, 0, 1, 1, 0, ..., 0, 1}; where 1 represents an MTC and 0 represents an NTC, and |H| represents the total number of existing hospitals. The initial population of chromosomes (*P*) is generated using an 'initial solution generator' described in Section 4.3.

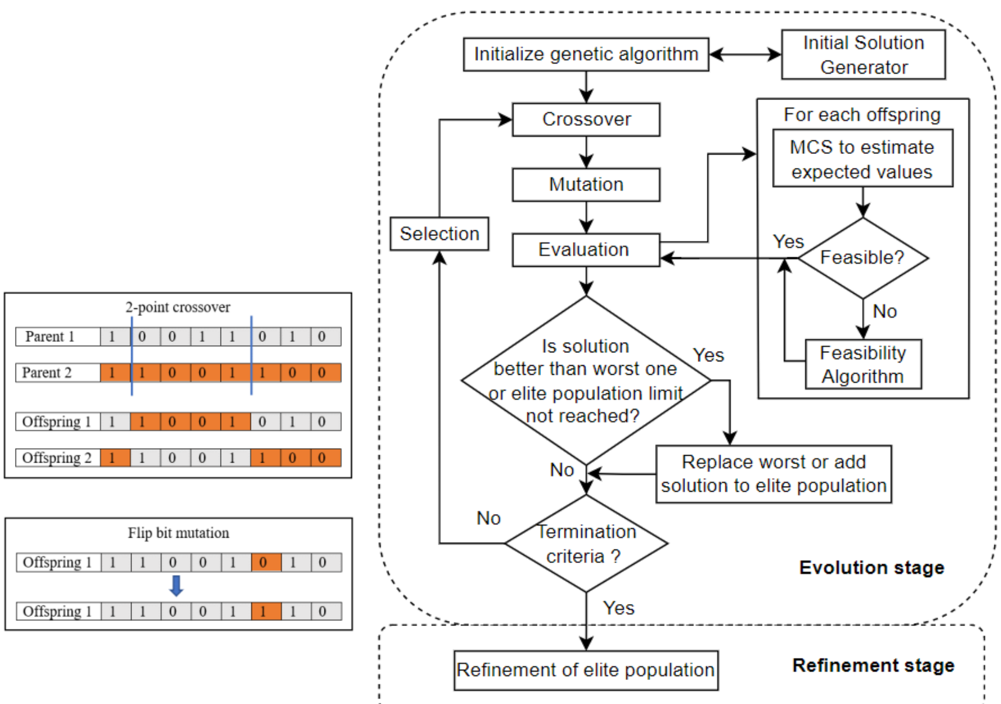


Figure 25. Flowchart of the proposed Simheuristic

4.4.1.2 Crossover

It is analogous to reproduction by exchanging genetic material between parents to generate new chromosomes known as offspring that are more likely to be better than the parents. The standard 2-point crossover is used to generate two offspring using the two parents with the crossover probability (*cp*) that controls the crossover operation.

4.4.1.3 Mutation

The mutation process diversifies the search process and prevents evolution from getting stuck in local optima. We use the flip bit mutation technique with mutation probability (*mp*) that controls the rate at which chromosomes' genes can mute.

4.4.1.4 Evaluation

Each offspring is evaluated using a MCS to estimate the expected efUT (objective), efOT, and volume at MTCs. For a given offspring, we first sample the assessed severity of each patient using MCS and use the notional tasking algorithm (per Hirpara et al., 2022) to determine the destination hospital. In each MCS replication, we use the ISS-group specific Bernoulli random variable to assign the assessed severity to each patient. Next, we use destination hospital type, along with true underlying and assessed injury severity, to classify each patient in one of the 4 system-related triage types and calculate efUT and efOT. Further, the volume at an MTC is calculated as a sum of number of underlying severely injured patients primarily transported to the MTC and the number of efUT patients that would be transferred from the NTC. We repeat these steps for *N* replications in the MCS to estimate the expected value of efUT, efOT, and volume at MTCs. If an offspring violates any constraints, we use our proposed ‘feasibility’ algorithm to convert an infeasible offspring to a feasible one (see Section 4.2).

4.4.1.5 Elite population

Elite population refers to a collection of promising offspring that are derived and evaluated in two stages, evolution, and refinement. During the evolution stage, each offspring in a generation is evaluated via a quick simulation run (limited number of replications, N) using the MCS. Feasible offspring that has the best solution in that generation is added to the elite population if either (i) the number of solutions in the elite population has not reached its maximum limit (P_{elite}) or (ii) this solution is better than the worst solution in the elite population. Once the maximum number of generations have been achieved, the refinement stage begins where all solutions in the elite population are then re-evaluated using a suitably large number of simulation replications (N_{elite}) (Juan et al., 2014; Pagès-Bernaus et al., 2019; Gruler et al., 2020; Quintero-Araujo et al., 2021). This evolution-refinement (similar to coarse-fine) approach enables identifying high quality solutions at reduced computational effort.

4.4.1.6 Selection

We use the elitist selection method that combines parents and offspring, and selects the best individuals among them for the next generation.

4.4.1.7 Stopping criteria

We use two termination criteria in the proposed approach: (i) a maximum number of total generations and (ii) a maximum number of generations without improvement in the best solution.

4.4.1.8 Refinement of elite population

This stage refines the best solutions preserved over the evolution (elite population) using an intensive simulation (large number of replications, N_{elite}) to precisely estimate the performance metrics.

4.4.2 Feasibility algorithm

An offspring is considered infeasible if it violates any of the four constraints in the model: (i) the lower bound on MTC volume (Constraints 5 in the optimization model), (ii) the upper bound on MTC volume (Constraints 5), (iii) limit on maximum MTCs in the

---

Feasibility Algorithm
New offspring = infeasible offspring
$k = 0$
Repeat
   If minimum volume violation:
      New offspring = Minimum volume repair
   If max volume violation:
      New offspring = Maximum volume repair
   If #MTCs (New Offspring) > C:
      New offspring = Maximum MTCs repair
   Else If efOT violation **and** no minimum volume violation:
      New offspring = Maximum efOT repair
   Evaluate New offspring
   If feasible:
      Return New offspring
   Else:
      $k = k + 1$
Until $k = k_{max}$

TSA (Constraint 6), and (iv) maximum allowable efOT (Constraint 7). The proposed feasibility algorithm iteratively repairs the offspring using one or more modifications corresponding to the constraint violations, as detailed below.

4.4.2.1 Minimum volume repair

Violation in the constraint representing the lower bound of MTC volume is the most common reason for the infeasibility of an offspring. If this happens, we use the following dynamics: if we downgrade an infeasible MTC to an NTC, then patients from downgraded MTC would be diverted to another infeasible nearby MTC. In that case, this nearby MTC may eventually receive enough patients to satisfy the minimum volume requirement at that MTC (making it feasible). Following this logic, if the minimum volume constraint is violated at only one MTC in an offspring, then that offspring can be repaired by simply downgrading the infeasible MTC to NTC (i.e., the decision variable for this hospital in that offspring is switched from 1 to 0). However, if >1 MTCs are violate this constraint in an offspring, then we create an offspring-specific list consisting of all these infeasible MTCs and downgrade only a handful of infeasible MTCs using the following steps. First, we pick one MTC in that list at random and downgrade it to an NTC. Second, then we randomly pick another MTC from this same list ensuring that this new MTC is relatively far away from the already downgraded MTC (e.g., >30 miles) and downgrade this MTC as well. We repeat the second step until for all MTCs in this list have been tested (ensuring that the chosen MTC is far away from the already downgraded MTCs). At this point, the entire offspring is again evaluated using the MCS. If this offspring is still infeasible, then the feasibility algorithm is repeated again until the constraint is satisfied.

4.4.2.2 Maximum volume repair

If any MTC violates the maximum allowable limit of severely injured patients, the algorithm randomly selects one of the five NTCs closest to this MTC and upgrades that NTC to an MTC. The idea is that upgrading a nearby NTC to an MTC would divert some severely injured patients to this new MTC, eventually reducing the burden on the violating MTC.

4.4.2.3 Maximum MTCs repair

A violation of maximum allowable MTCs in the TSA can only be repaired by downgrading all additional MTCs to NTCs in the same iteration this repair is invoked. Note that a downgrade of MTCs could increase access time for severely injured patients and subsequently increase efUT (objective). Therefore, the idea is to downgrade MTCs that have additional MTCs nearby (within 30 miles). To understand this, consider a situation when there are 13 MTCs in the offspring, while only 10 MTCs (C=10) are allowed in the TSA (Constraint 7). First, for each of the13 MTCs, the repair algorithm determines the number of other MTCs nearby and downgrades the MTC with a maximum number of other MTCs nearby. This process is repeated until the maximum limit (C) is satisfied; in this specific situation, the process is repeated 3 times.

4.4.2.4 Maximum allowable efOT

An infeasibility due to higher efOT patients is repaired by downgrading MTCs to NTCs. Recall that the higher the number of MTCs, the higher the likelihood of efOT. In this case, because of the downgrade of one or more MTCs in the minimum volume repair

(4.2.1) and maximum MTCs repair (4.2.3), efOT could already be within the maximum allowable value. Therefore, this repairing step is only performed if the offspring is not violating any of those two constraints (4.2.1 and 4.2.3). The repairing for this violation is similar to the previous repair (Section 4.2.3); however, it only downgrades one MTC with the highest number of other MTCs nearby and it usually repairs the issue. If it does not, we downgrade another one according to the same rule in the next iteration of the feasibility algorithm until the constraint is satisfied.

To achieve 100% of feasibility in the offspring, a higher number of iterations (k_{max}) are required in the feasibility algorithm. Initial analysis suggested that, in most cases, at least 95% feasibility can be achieved by using the feasibility algorithm with k_{max} = 4 compared to less than 40% without it. If there is still an infeasible offspring, then its objective value is set to a very high value.

4.4.3 Initial solution generator

To initiate the Simheuristic, we use this generator to generate several initial feasible solutions for a given number of underlying severe patients (*SI*), minimum volume requirement at MTC (V^{min}), and maximum allowable MTCs in the TSA (*C*). We embed the feasibility algorithm (Section 4.2) inside of this generator as well.

First, the generator calculates the number of MTCs to open (*m*) in the TSA as the minimum of two criteria: (a) maximum feasible MTCs for a given minimum volume requirement and (b) maximum allowable MTCs in the TSA (*C*). For each county in the TSA, the generator first calculates criteria (a), the maximum feasible MTCs (*n*) in the county, as the nearest smallest integer of the ratio of total severely injured patients in the

Initial Solution Generator
MTC = Ø
m = *min*(floor(# SI/V^{min}), C)
For each county in TSA:
  n = *floor*(# SI cases in county/V^{min})
  If $n \geq 1$:
    MTC = MTC ∪ {randomly choose n location(s) within the county}
If m - |MTC| > 0:
  MTC = MTC ∪ {randomly choose (m - |MTC|) non MTC location within TSA}
Return MTC
Chromosome = Generate binary string using MTC
Evaluate Chromosome
If infeasible:
  Run feasibility algorithm

county (#SI) and the minimum volume requirement at MTC (V^{min}). For a given county, n greater than or equal to 1 represents that there are sufficient underlying severe patients to satisfy the absolute minimum volume requirement of these n MTCs. Therefore, the generator randomly chooses n candidate locations from that county as MTCs. Next, if the total number of chosen MTCs across all counties (|MTC|) is less than the maximum feasible MTCs (m) in the TSA, the generator randomly chooses additional candidate locations as MTCs. The generator creates a binary chromosome using chosen MTC locations and evaluates the chromosome. If the chromosome is infeasible, the 'feasibility' algorithm is invoked to repair the chromosome. This process repeats until we generate P initial feasible solutions to initiate the evolution.

We used R to implement our proposed Simheuristic on a computer with 16 virtual CPUs, each with a 2.9 gigahertz processor and a total of 32GB RAM. We used parallel processing in R that allowed parallel feasible solution generation, evaluation of offspring during evolution, running feasibility algorithms for infeasible offspring, and refinement of the elite population.

Based on preliminary experiments, we use a population size (P) of 30 chromosomes, the mutation rate (mp) of $1/n$ (where n represents a number of candidate locations), crossover probability (cp) of 0.9, 50 replications of MCS during evolution (N), 500 replications of MCS during refinement (N_{elite}), and the elite population size of 50 (P_{elite}). We used two termination criteria (i) maximum generations (set to 1,000) and (ii) non-improving 75 generations.

4.5 Experimental Study

To analyze the impact of assessment-related mistriages, we selected (i) a TSA and (ii) identified corresponding trauma incidences. For (i), we considered a group of real counties in an existing midwestern US state (Figure 26). We ensured that this TSA is representative of an average TSA in the US for which 62% counties are rural; in the chosen TSA, there are 34 counties with 21 rural.

For (ii), we estimated median population of a US state, which is 4.55 million. Based on state-wide annual trauma registry data published by nine states across the US, we estimated that there are an average of 5.2 trauma patients per thousand citizens; this corresponded to 23,680 trauma patients in the TSA (given 4.55 million population). In line with National Trauma Data Bank (NTDB) report, we considered 15.63% of patients as severely injured and the rest as non-severely injured (NTDB, 2016). We used ArcGIS

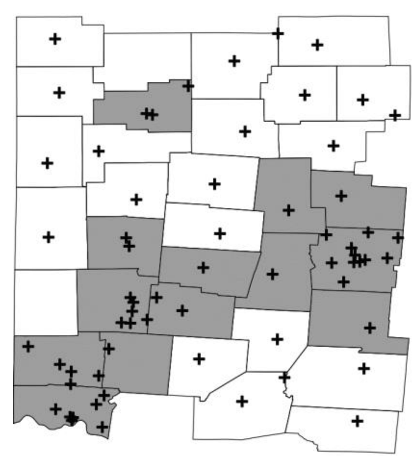


Figure 26. TSA with counties; grey filled areas are urban counties; '+'represents candidate locations

Pro 2.9.1 to generate locations (latitude and longitude) of trauma incidences in the TSA; we used a Gini index of 0.5 to represent distribution of trauma incidences in this TSA with moderate clustering around urban areas.

We next modeled the injury assessment mistriages and refer to them as assessment-related (ar) to distinguish them from system-related (sr) that occur during destination determination. For each group of ISS, we generated ar-related mistriages using the scheme shown in the Table 18; because there are relatively few mistriages in assessing minor injuries (ISS 1-8 group), we assumed 100% certainty in assessing minor injuries. We considered up to 75% mistriages in assessing moderate injuries (ISS 9-15) as it is hard to separate from severe injuries (ISS >15), resulting in often higher mistriages among all groups. The scenario with no assessment mistriage is considered the base case (marked in bold).

Table 18. Summary of the parameters, levels, and values in the sensitivity analysis

ISS group	**Levels**	**Assessment mistriage in percentage**	**Probability (in Bernoulli distribution)**
9-15	4	**0,** 25, 50, 75 (resulting in arOT)	**0**, 0.25, 0.5, 0.75
>15	3	**0,** 25, 50 (resulting in arUT)	**0**, 0.75, 0.5

There are a total of 64 hospitals currently in this TSA. We considered all hospitals in the TSA as candidate locations and used the underlying transportation network in these counties to generate the actual drive time matrix (TG_{ij}) from an incidence location to a candidate hospital using ArcGIS Pro 2.9.1. Airtime matrix (TA_{ij}) was generated using the Haversine formula assuming the helicopter speed of 120 mph.

As per ACS recommendation, we used a minimum of 240 severely injured patients as a lower bound for MTCs (V^{min}), 1,000 as the upper bound on volume (V^{max}) and maximum allowable efOT as 35% of total underlying non-severe patients (OT^{max}). In line with the literature, we used access time threshold as 30 minutes, bypass time threshold as 0 minutes, C = 64 (total candidate locations), and helicopter usage of 15% to transport severely injured patients.

Insight 1: Assessment-related mistriages in the ISS >15 group (underlying severe injuries) leads to poor patient safety

Recall that ISS >15 group is patients with severe underlying injuries. For this group, an assessment mistriage means assessing the injuries to be non-severe. When such mistriages are at 50% (resulting in arUT), we noticed that the optimal trauma network had an increase of 799% in efUT (1,357 vs. 151 in the base case). Further, the network required

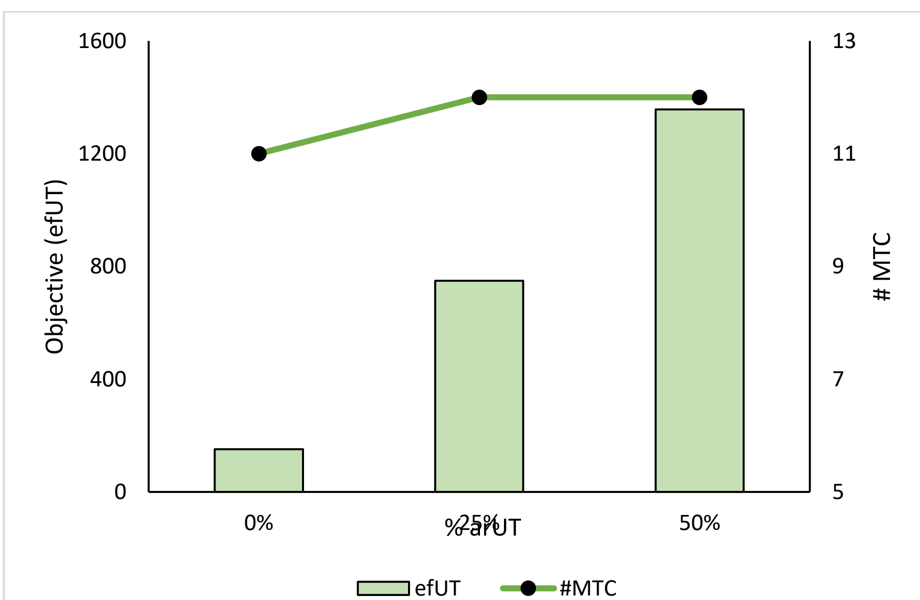


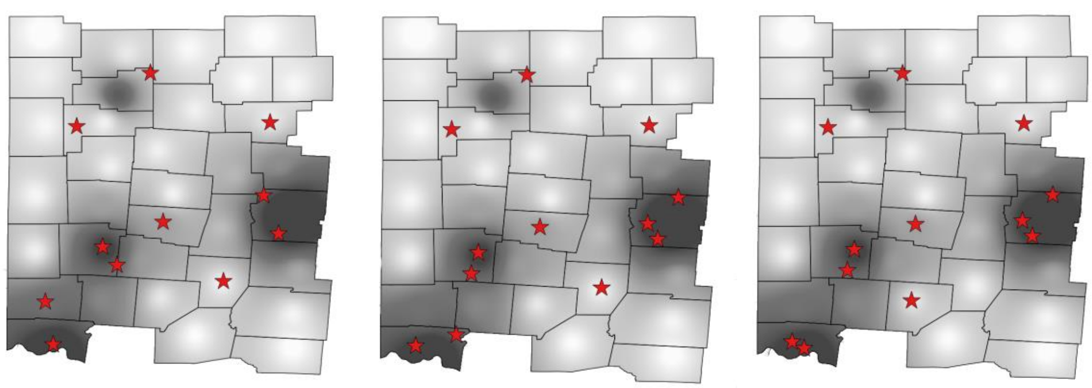
Figure 27. efUT and # MTCs for different % of arUT

an additional MTC (12 vs. 11) compared to the base case (see Figure 27); these 12 MTCs are clustered around high incidence rates (see Figure 28).

To understand this further, recall that when all patients are assessed correctly (base case) in the ISS >15 group, a severe patient may only experience UT|sr if the EMS has to transport this patient to an NTC (incorrect hospital) because the nearest MTC is outside of the access threshold (30 minutes via air or ground). This means that to minimize UT|sr, the optimal network corresponding to the base case would likely have a dispersed distribution of MTCs to increase access to MTC.

However, when the assessment mistriage in the ISS >15 group is high (say, 50%), the EMS would aim to transport mistriage patients (deemed non-severe) to an NTC, thus experiencing UT|ar. This increases the efUT as pointed out earlier.

Insight 2: Assessment-related mistriages in the ISS >15 group (underlying severe injuries) may lead to the clustering of MTCs near high trauma incidence rates



0% arUT　　25% arUT　　50% arUT

Figure 28. Locations of MTCs for different percentage of arUT; dense areas represent higher trauma incidence rate and stars represent MTCs

This insight is clearly evident from Figure 28 where the MTCs appear to cluster as the assessment mistriages increase. Recall clustering of MTCs around those areas may prone EMS paramedics to transport assessed non-severely injured patients (especially, the incorrectly-assessed in ISS >15 group) to MTCs (per the bypass threshold condition in the Notional Tasking Algorithm). That is, although an ISS >15 patient is mistriaged as having non-severe injuries, that patient may end up in an MTC (experiencing srOT), the correct hospital based on true underlying injury indicating an effective appropriate triage patient. This essentially means, clustering of MTCs would help reduce UT|ar patients in the trauma network with higher mistriages in assessing severe injuries. The gradual clustering of MTCs with an increase in assessment-related mistriages in the ISS >15 group is evident from Figure 27.

Insight 3: Assessment-related mistriages in the ISS 9-15 group (underlying non-severe injuries) may lead to poor patient safety and potentially a disperse distribution of MTCs

Recall that for the ISS 9-15 group, an assessment-related mistriage indicates that the EMS assessed the true underlying non-severe injuries of a patient as being severe. When such mistriages are increase, we observed an increase in the objective function (suggesting poor patient safety). At 75% assessment mistriage in this group, efUT increased from 151 (base case) to 586 (a 288% increase); see Figure 29. We also observed a more dispersed distribution of MTCs with an increase in the mistriages than the base case (Figure 30), along with a decrease in MTC at high mistriage values.

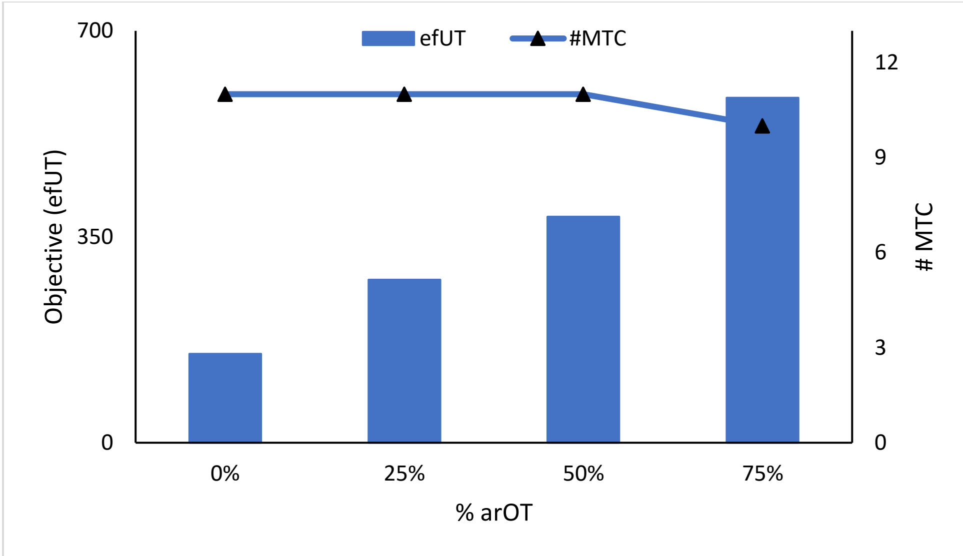
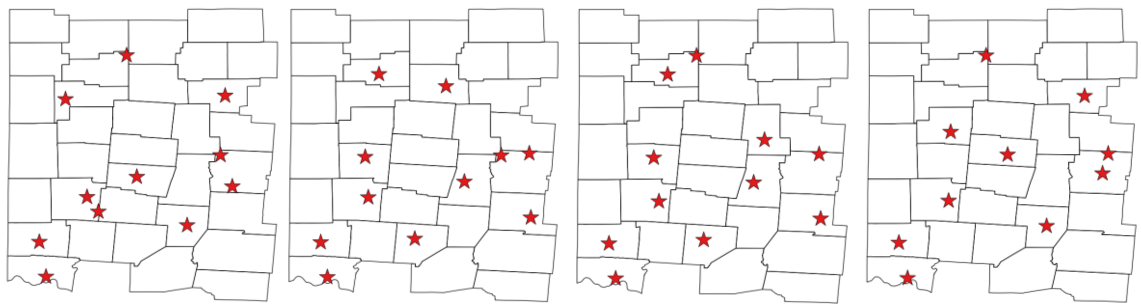


Figure 29. Objective and # MTCs for different % of arOT in ISS 9-15 group

To understand this phenomenon, note that a mistriaged patient in this group would be assessed by the EMS as severe, and, in turn, will prompt the EMS to transport them to the nearest MTC (the correct hospital based on assessed severity) if it is within the access threshold; a case of over-triage (OT|ar). Consequently, if such OT|ar patients need to be limited due to the efOT constraints (constraint 7), the ideal thing would be to lower the accessibility of MTCs from the scene (i.e., increase the access time), which would induce the EMS to transport such mistriaged (assessed severe) patients to nearby NTCs. This would result in the patient experiencing appropriate effective triage (an eventual correct outcome for the patient). To enable this, the resulting network tends to have a dispersed distribution of MTCs, and potentially fewer MTCs at high mistriage rates for this group. The obvious downside of this change in the network is the inability of EMS to transport severely injured patients to MTCs (both correctly assessed in the ISS >15 group and incorrectly assessed in the ISS 9-15 group), resulting in higher efUT (Figure 29).



0% arOT 25% arOT 50% arOT 75% arOT

Figure 30. Comparison of trauma network for different % of arOT in ISS 9-15 group

4.6 Summary

Realizing that an accurate on-scene injury assessment is critical for timely patient care, national and state organizations have proposed guidelines to assist EMS paramedics. Even though the field triage guidelines (FTGs) have been developed by medical experts and updated periodically, literature and anecdotes from practice suggested that these FTGs do not always provide 100% accurate determination of injury severity. Errors in clinical assessment of a patient's injury can have negative implication on overall patient safety, and even the design of the trauma network.

To understand the impact of such errors, not evaluated in prior literature, we proposed a stochastic nested multi-level, multi-transportation capacitated model. This model explicitly considers injury assessment mistriages, and determines the number and locations of major trauma centers to maximize patient safety. Because the EMS on-scene clinical injury assessment (severe vs. non-severe) is a binary decision problem, we modeled this assessment as a Bernoulli random variable with the probability of success referring to a patient classified as having severe injuries. Because mistriage in injury assessment for

ISS 1-8 group is low, we specifically focused on ISS 9-15 (moderately severe) and ISS >15 (severe) groups as mistriages in these groups have a higher likelihood, with major implications on patient safety.

We proposed an optimization model that captures the impact of assessment-related mistriages on the final outcome for each trauma patient. To solve this model efficiently, we proposed a Simheuristic approach that integrates Monte Carlo Simulation (MCS) with a genetic algorithm (GA). Two enhancements were proposed: elite population and feasibility algorithm. To quantify the impact of assessment mistriages on the resulting trauma network and patient safety, we considered a representative TSA generated using data from a state trauma system in the US. Key insights from our experimental study suggest that assessment-related mistriages:

- in assessing patients with underlying severe injuries (ISS>15) can substantially increase efUT.
- in assessing such patients with severe injuries (ISS>15) may also lead to the clustering of MTCs around areas with high trauma incidence rates.
- in assessing patients with underlying non-severe injuries (ISS 9-15) increase efOT, resulting in the reduction and dispersed distribution of MTCs.

Our research has several practical implications. It allows for quantifying the impact of clinical assessment-related mistriages on the performance of the existing network for trauma policy makers. Decision makers can use our approach for quantitative evaluation of various existing or new injury assessment protocols and even for validation of changes in existing protocols before implementing them in practice. In addition, decision-makers can quantify the impact of reducing assessment-related mistriages in specific ISS groups and modify existing EMS training programs for such groups. Better training can lead to reduced errors (if not fully eliminated), resulting in improve outcomes of trauma patients.

CHAPTER 5

CONCLUSION

Timely access to a trauma center is a key determinant of patient outcomes. Motivated by the limitations of existing approaches in locating trauma facilities, this dissertation studied the trauma network design problem to maximize patient safety, modeled using under-triage (UT) and over-triage (OT). In this dissertation, we developed multiple optimization-based approaches that account for various critical elements of a trauma system, such as multiple patient types, multiple levels of hospitals, multiple choices for transportation, multiple destination determination criteria, equity of care, and consideration of mistriages in the clinical assessment of injuries. Below we summarize our key findings followed by opportunities for future work in this area.

5.1 Summary of Contribution 1

We proposed the TCLP to determine the optimal number and location of MTCs that minimize the weighted sum of mistriages (srUT and srOT). We considered both severe and non-severe patients and their on-field operational decision-making processes and accounted for associated mistriages in the optimization model. We also conducted an experimental study to analyze how 4 key system parameters (i.e., weights for two mistriages in the objective, volume requirements at a major TC, access, and bypass thresholds) impact patient safety. The experimental results revealed the following:

- An increase in the number of major TCs can reduce system-related UT (srUT) at the cost of an increase in system-related OT (srOT); the trauma decision maker should choose their weights in the objective function wisely.
- There is an inverse relationship between the minimum volume requirement at MTC and the number of MTCs in the network. A lower volume requirement can result in better patient safety, but may not be financially viable for some MTCs.
- Quickly transporting severely injured patients to the nearest MTC is desirable (reflected by a lower 'access' threshold); this can only be achieved through more MTCs in the network (with the corresponding increase in srOT as highlighted above).
- An illustration of our approach using real data from Ohio indicated that up to 51.9% reduction in srUT can be achieved (at nearly the same srOT rate) with just one additional MTC. The state can also redistribute the same 21 MTCs and still achieve a high reduction in srUT (46.6%) along with a 4.95% reduction in srOT.

5.2 Summary of Contribution 2

We proposed NTNDP and presented a mixed integer problem that jointly locates major, intermediate, and non-trauma centers to minimize a weighted sum of equity among regions and effectiveness across a trauma service area (TSA). In the proposed model, we also incorporated three dominant destination determination criteria that EMS use at the incidence location (i.e., protocol, patient choice, and closest facility). We generated a TSA using data collected from state-wide trauma agencies across the US and analyzed the impact of system parameters on the performance of the network. We demonstrated the use of the proposed approach on 2019 data for the state of Ohio. The key insights from our study include the following:

- The 100% use of the ‘protocol’ criterion (as per ACS recommendation) can achieve better patient safety even with fewer MTCs and ITCs. Increased use of the ‘patient choice’ criterion appears to reduce patient safety and may lead to a higher number of ITCs in the trauma network, specifically in suburban and rural areas.
- A clustered distribution of severely injured patients in the TSA appears to improve trauma system performance with fewer MTCs and ITCs.
- An emphasis on equity in the network may lead to a decline in overall patient safety; better performance can be achieved by balancing it with effectiveness across the TSA.
- The illustration of our approach using real data from Ohio suggested an over 30% improvement in patient safety with an existing mixer of destination determination criterion. The importance of 100% protocol use for destination determination was also verified.

5.3 Summary of Contribution 3

To understand the impact of mistriages in injury assessment, we proposed a stochastic nested multi-level, multi-transportation capacitated model that explicitly considers clinical mistriages in injury assessment in determining the number and location of trauma centers to maximize patient safety. We modeled on-scene clinical injury assessment as a Bernoulli random variable with the probability of success referring to a patient classified as having severe injuries. We generated a representative TSA using data from a state trauma system in the US, and quantified the impact of assessment mistriages on the resulting trauma network and patient safety. Our experimental study led us to the following insights:

- The trauma network is susceptible to clinical mistriages in assessing patients with severe injuries (ISS>15); higher mistriages in this group tends to increase effective UT (efUT).
- Higher mistriages in assessing patients with underlying severe injuries (ISS>15) may also lead to the clustering of MTCs around areas with high trauma incidence rates.
- The network with higher mistriages in assessing patients with non-severe injuries (ISS 9-15) resulted in higher effective OT (efOT) and dispersed distribution of MTCs.

5.4 Future Research

This research attempted to address several fundamental questions in trauma care provision in the US. This is a relatively less explored domain within IE/OR; there are many other opportunities for future research that we summarize below.

Inclusion of migration patterns within a state, between the state, and other patient distribution uncertainties would help evaluate the robustness of optimal (near-optimal) trauma network. Learning such migration patterns to design trauma networks that evolve over time would create a roadmap for the trauma decision makers to deploy in practice. Because the resulting model would be highly complex, it would more a sophisticated solution approach.

The inclusion of the cost incurred in upgrading an ACS-verification of a trauma center (e.g., from NTC to ITC/MTC and ITC to MTC) through a multi-criteria optimization model would allow trauma policymakers to appropriately tradeoff cost vs. care in designing their network.

Investigating the impact of uncertainty in the ground and air EMS arrival and on-scene assessment times on destination determination, network design, and patient safety would also be an interesting consideration. While our models assumed deterministic times for EMS arrival and on-scene assessment, in practice, those times could vary substantially between rural and urban areas, and can lead to subsequent network changes.

Patient safety in rural areas is a major public health concern. As MTCs are not financially viable in rural areas due to low patient volume, designing an optimization-based framework to tradeoff between financial loss of MTC due to fewer patients vs. compromise in patient safety due to the inability to access trauma centers can be considered. This model can further be extended to account for various regional or state-level government subsidies to recover the financial loss at the trauma center due to low volume.

REFERENCES

1. Ahmadi-Javid, A., Seyedi, P., & Syam, S. S. (2017). A survey of healthcare facility location. *Computers & Operations Research*, *79*, 223-263.
2. Amaran, S., Sahinidis, N. V., Sharda, B., & Bury, S. J. (2016). Simulation optimization: a review of algorithms and applications. *Annals of Operations Research*, *240*(1), 351-380.
3. American College of Surgeons Committee on Trauma (2015). *Needs-Based Assessment of Trauma Systems (NBATS) Tool*. Retrieved from https://lern.la.gov/wp-content/uploads/Appendix-I-acs-nbats-tool.pdf.
4. American College of Surgeons. (2016). *Strengthening our national trauma system*.
5. American Trauma Society (2022). *Trauma Center Levels Explained*. Retrieved from https://www.amtrauma.org/page/traumalevels.
6. Ares, J. N., De Vries, H., & Huisman, D. (2016). A column generation approach for locating roadside clinics in Africa based on effectiveness and equity. *European Journal of Operational Research*, *254*(3), 1002-1016.
7. Aringhieri, R., Bruni, M. E., Khodaparasti, S., & van Essen, J. T. (2017). Emergency medical services and beyond: Addressing new challenges through a wide literature review. *Computers & Operations Research*, *78*, 349-368.

8. Armstrong, J. H., Hammond, J., Hirshberg, A., & Frykberg, E. R. (2008). Is overtriage associated with increased mortality? The evidence says "yes". *Disaster Medicine and Public Health Preparedness, 2*(1), 4-5.

9. Balcik, B., & Beamon, B. M. (2008). Facility location in humanitarian relief. *International Journal of Logistics Research and Applications, 11*(2), 101-121.

10. Barringer, M. L., Thomason, M. H., Kilgo, P., & Spallone, L. (2006). Improving outcomes in a regional trauma system: impact of a level III trauma center. *The American Journal of Surgery, 192*(5), 685-689.

11. Bayram, V., Tansel, B. Ç, & Yaman, H. (2015). Compromising system and user interests in shelter location and evacuation planning. *Transportation Research Part B, 72*, 146-163.

12. Bélanger, V., Ruiz, A., & Soriano, P. (2019). Recent optimization models and trends in location, relocation, and dispatching of emergency medical vehicles. *European Journal of Operational Research, 272*(1), 1-23.

13. Beliën, J., De Boeck, L., Colpaert, J., Devesse, S., & Van den Bossche, F. (2013). Optimizing the facility location design of organ transplant centers. *Decision Support Systems, 54*(4), 1568-1579.

14. Branas, C. C., MacKenzie, E. J., & ReVelle, C. S. (2000). A trauma resource allocation model for ambulances and hospitals. *Health Services Research, 35*(2), 489.

15. Branas, C. C., MacKenzie, E. J., Williams, J. C., Schwab, C. W., Teter, H. M., et al., (2005). Access to trauma centers in the United States. *JAMA, 293*(21), 2626-2633.

16. Branas, C. C., Wolff, C. S., Williams, J., Margolis, G., & Carr, B. G. (2013). Simulating changes to emergency care resources to compare system effectiveness. *Journal of clinical epidemiology, 66*(8), S57-S64.

17. Brotcorne, L., Laporte, G., & Semet, F. (2003). Ambulance location and relocation models. *European Journal of Operational Research, 147*(3), 451-463.

18. Brown, J. B., Gestring, M. L., Guyette, F. X., Rosengart, M. R., Stassen, N. A., et al., (2016). Development and validation of the air medical prehospital triage score for helicopter transport of trauma patients. *Annals of Surgery, 264*(2), 378-385.

19. Brown, J. B., Rosengart, M. R., Billiar, T. R., Peitzman, A. B., & Sperry, J. L. (2016). Geographic distribution of trauma centers and injury-related mortality in the United States. *Journal of Trauma and Acute Care Surgery, 80*(1), 42-50.

20. Burkey, M. L., Bhadury, J., & Eiselt, H. A. (2012). A location-based comparison of health care services in four US states with efficiency and equity. *Socio-Economic Planning Sciences, 46*(2), 157-163.

21. Cardoso, T., Oliveira, M. D., Barbosa-Povoa, A., & Nickel, S. (2015). An integrated approach for planning a long-term care network with uncertainty, strategic policy and equity considerations. *European Journal of Operational Research, 247*(1), 321-334.

22. Carr, B. G., & Branas, C. (2010). Trauma center maps. *University of Pennsylvania Cartographic Modeling Laboratory*.

23. Caruso, V., & Daniele, P. (2018). A network model for minimizing the total organ transplant costs. *European Journal of Operational Research, 266*(2), 652-662.

24. Centers for Disease Control and Prevention (2021). *Economic Cost of Injury – United States, 2019*.

25. Centers for Disease Control and Prevention (2022). *Injuries and Violence Are Leading Causes of Death*. Retrieved from https://www.cdc.gov/injury/wisqars/animated-leading-causes.html

26. Çetin, E., & Sarul, L. S. (2009). A blood bank location model: A multiobjective approach. *European Journal of Pure and Applied Mathematics, 2*(1), 112.

27. Chanta, S., Mayorga, M. E., & McLay, L. A. (2014). Improving emergency service in rural areas: a bi-objective covering location model for EMS systems. *Annals of Operations Research, 221*(1), 133-159.

28. Chen, X., Guyette, F. X., Peitzman, A. B., Billiar, T. R., Sperry, J. L., & Brown, J. B. (2019). Identifying patients with time-sensitive injuries: Association of mortality with increasing prehospital time. *The journal of trauma and acute care surgery, 86*(6), 1015–1022.

29. Chen, Z., Chen, X., Li, Q., & Chen, J. (2013). The temporal hierarchy of shelters: a hierarchical location model for earthquake-shelter planning. *International Journal of Geographical Information Science, 27*(8), 1612-1630.

30. Cheng Siong Lim, Mamat, R., & Braunl, T. (2011). Impact of ambulance dispatch policies on performance of emergency medical services. *IEEE Transactions on Intelligent Transportation Systems, 12*(2), 624-632.

31. Chia-Feng Juang. (2004). A hybrid of genetic algorithm and particle swarm optimization for recurrent network design. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 34*(2), 997-1006.

32. Cho, S.-H., Jang, H., Lee, T., & Turner, J. (2014). Simultaneous location of trauma centers and helicopters for emergency medical service planning. *Operations Research, 62*(4), 751-771.

33. Clerc, M., & Kennedy, J. (2002). The particle swarm - explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation, 6*(1), 58-73.

34. Cocking, C., Flessa, S., & Reinelt, G. (2012). Improving access to health facilities in Nouna district, Burkina Faso. *Socio-Economic Planning Sciences, 46*(2), 164-172.

35. Committee on Trauma of the American College of Surgeons. (1986). Hospital and prehospital resources for optimal care of the injured patient and appendices A through J: United States-1986.

36. Committee on Trauma of the American College of Surgeons. (2006). Resources for optimal care of the injured patient 2006. *Chicago, IL: American College of Surgeons*.

37. Côté, M., Syam, S., Vogel, W., & Cowper, D. (2007). A mixed integer programming model to locate traumatic brain injury treatment units in the department of veteran's affairs: A case study. *Health Care Management Science, 10*(3), 253-267.

38. Daskin, M. S. (2011). *Network and discrete location: models, algorithms, and applications*. John Wiley & Sons.

39. Daskin, M. S. (2013). *Network and discrete location: Models, algorithms, and applications* (2nd ed.). New Jersey: John Wiley & Sons.

40. De Armas, J., Juan, A.A., Marques, J., Pedroso, J. (2017). Solving the deterministic and stochastic uncapacitated facility location problem: from a heuristic to a simheuristic. *Journal of the Operational Research Society*.

41. Doerner, K., Focke, A., & Gutjahr, W. J. (2007). Multicriteria tour planning for mobile healthcare facilities in a developing country. *European Journal of Operational Research, 179,* 1078-1096.

42. Edison, E., & Shima, T. (2011). Integrated task assignment and path optimization for cooperating uninhabited aerial vehicles using genetic algorithms. *Computers & Operations Research*, *38*(1), 340-356.

43. Enayati, S., Mayorga, M. E., Toro-Díaz, H., & Albert, L. A. (2019). Identifying trade-offs in equity and efficiency for simultaneously optimizing location and multipriority dispatch of ambulances. *International Transactions in Operational Research, 26*(2), 415-438.

44. Frykberg, E. R. (2002). Medical management of disasters and mass casualties from terrorist bombings: how can we cope? *Journal of Trauma and Acute Care Surgery, 53*(2), 201-212.

45. Gauss, T., Ageron, F. X., Devaud, M. L., Debaty, G., Travers, S., Garrigue, D., Raux, M., Harrois, A., Bouzat, P., & French Trauma Research Initiative (2019). Association of Prehospital Time to In-Hospital Trauma Mortality in a Physician-Staffed Emergency Medicine System. *JAMA surgery, 154*(12), 1117–1124.

46. Gianola, S., Castellini, G., Biffi, A., Porcu, G., Fabbri, A., Ruggieri, M. P., Stocchetti, N., Napoletano, A., Coclite, D., & D'Angelo, D. (2021). Accuracy of pre-hospital triage tools for major trauma: a systematic review with meta-analysis and net clinical benefit. *World journal of emergency surgery, 16*(1), 1-11.

47. Gruler, A., Panadero, J., de Armas, J., Pérez, J. A. M., & Juan, A. A. (2018). Combining variable neighborhood search with simulation for the inventory routing problem with

stochastic demands and stock-outs. *Computers & Industrial Engineering, 123,* 278-288.

48. Gruler, A., Panadero, J., de Armas, J., Pérez, J. A. M., & Juan, A. A. (2020). A variable neighborhood search Simheuristic for the multiperiod inventory routing problem with stochastic demands. *International Transactions in Operational Research*, *27*(1), 314-335.

49. Güneş, E. D., Melo, T., & Nickel, S. (2019). Location problems in healthcare. In *Location Science* (pp. 657-686). Springer.

50. Güneş, E., Yaman, H., Cekyay, B., & Verter, V. (2014). Matching patient and physician preferences in designing a primary care facility network. *Journal of the Operational Research Society, 65*(4), 483-496.

51. Haase, K., & Müller, S. (2015). Insights into clients' choice in preventive health care facility location planning. *OR Spectrum, 37*(1), 273-291.

52. Handing Wang, Yaochu Jin, & Jansen, J. O. (2016). Data-driven surrogate-assisted multiobjective evolutionary optimization of a trauma system. *IEEE Transactions on Evolutionary Computation, 20*(6), 939-952.

53. Hirpara, S., Vaishnav, M., Parikh, P. J., Kong, N., & Parikh, P. (2022). Locating trauma centers considering patient safety. *Health Care Management Science*, 1-20.

54. Holland John, H. (1975). Adaptation in natural and artificial systems. *Ann Arbor*.

55. Ingolfsson, A., Budge, S., & Erkut, E. (2008). Optimal ambulance location with random delays and travel times. *Health Care Management Science, 11*(3), 262-274.

56. Intrevado, P., Verter, V., & Tremblay, L. (2019). Patient-centric design of long-term care networks. *Health Care Management Science, 22*(2), 376-390.

57. Izquierdo, J., Montalvo, I., Pérez, R., & Fuertes, V. S. (2008). Design optimization of wastewater collection networks by PSO. *Computers and Mathematics with Applications*, *56*(3), 777-784.

58. Jansen, J. O., Moore, E. E., Wang, H., Morrison, J. J., Hutchison, J. D., Campbell, M. K., & Sauaia, A. (2018). Maximizing geographical efficiency: an analysis of the configuration of Colorado's trauma system. *Journal of Trauma and Acute Care Surgery*, *84*(5), 762-770.

59. Jansen, J. O., Morrison, J. J., Wang, H., He, S., Lawrenson, R., Hutchison, J. D., & Campbell, M. K. (2015). Access to specialist care: optimizing the geographic configuration of trauma systems. *The journal of trauma and acute care surgery, 79*(5), 756.

60. Juan, A. A., Faulin, J., Grasman, S. E., Rabe, M., & Figueira, G. (2015). A review of simheuristics: Extending metaheuristics to deal with stochastic combinatorial optimization problems. *Operations Research Perspectives*, *2*, 62-72.

61. Juan, A.A., Barrios, B., Vallada, E., Riera, D., Jorba, J. (2014). A simheuristic algorithm for solving the permutation flow shop problem with stochastic processing times. *Simulation Modelling Practice and Theory, 46*, 101–117.

62. Juan,A.A., Faulin, J., Jorba, J., Caceres, J., Marques, J. (2013). Using parallel & distributed computing for solving real-time vehicle routing problems with stochastic demands. *Annals of Operations Research*, *207*, 43–65.

63. Kennedy, J., & Eberhard R. C. (1995). Particle swarm optimization. *Proceedings of the International Conference on Neural Network*, Perth, Australia, 1942-1984.

64. Kennedy, J., & Eberhart, R. C. (1997). A discrete binary version of the particle swarm algorithm. *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation,* Orlando, FL.

65. Khanesar, M. A., Teshnehlab, M., & Shoorehdeli, M. A. (Jun 2007). A novel binary particle swarm optimization. *Proceedings of the Mediterranean Conference on Control & Automation, 1-6.*

66. Kim, D., & Kim, Y. (2013). A Lagrangian heuristic algorithm for a public healthcare facility location problem. *Annals of Operations Research*, *206*(1), 221-240.

67. Larsson, A., Berg, J., Gellerfors, M., & Gerdin Wärnberg, M. (2021). The advanced machine learner XGBoost did not reduce prehospital trauma mistriage compared with logistic regression: a simulation study. *BMC medical informatics and decision making, 21*(1), 1-9.

68. Latha Shankar, B., Basavarajappa, S., Chen, J. C. H., & Kadadevaramath, R. S. (2013). Location and allocation decisions for multi-echelon supply chain network – A multi-objective evolutionary approach. *Expert Systems with Applications, 40*(2), 551-562.

69. Lee, T., & Jang, H. (2018). An iterative method for simultaneously locating trauma centers and helicopters through the planning horizon. *Operations Research for Health Care*, 19, 185-196.

70. Lejeune, M. A., & Prasad, S. Y. (2013). Effectiveness–equity models for facility location problems on tree networks. *Networks*, *62*(4), 243-254.

71. Lerner, E. B. (2006). Studies evaluating current field triage: 1996-2005. *Prehospital Emergency Care*, *10*(3), 303-306.

72. Liao, C., Chao-Tang Tseng, & Luarn, P. (2007). A discrete version of particle swarm optimization for flowshop scheduling problems. *Computers and Operations Research, 34*(10), 3099-3111.

73. Lim, C. S., Mamat, R., & Braunl, T. (2011). Impact of ambulance dispatch policies on performance of emergency medical services. *IEEE Transactions on Intelligent Transportation Systems, 12*(2), 624-632.

74. Liu, B., Wang, L., & Jin, Y.-H. (2007). An effective PSO-based memetic algorithm for flow shop scheduling. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 37*(1), 18-27.

75. Lupton, J. R., Davis-O'Reilly, C., Jungbauer, R. M., Newgard, C. D., Fallat, M. E., Brown, J. B., Mann, N. C., Jurkovich, G. J., Bulger, E., & Gestring, M. L. (2022). Under-triage and Over-triage Using the Field Triage Guidelines for Injured Patients: A Systematic Review. *Prehospital Emergency Care*, 1-8.

76. MacKenzie, E. J., Rivara, F. P., Jurkovich, G. J., Nathens, A. B., Frey, K. P., Egleston, B. L., Salkever, D. S., & Scharfstein, D. O. (2006). A national evaluation of the effect of trauma-center care on mortality. *The New England Journal of Medicine, 354*(4), 366-378.

77. Marianov, V., & Taborga, P. (2001). Optimal location of public health centres which provide free and paid services. *Journal of the Operational Research Society, 52*(4), 391–400.

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

79. Mestre, A. M., Oliveira, M. D., & Barbosa-Póvoa, A. P. (2015). Location–allocation approaches for hospital network planning under uncertainty. *European Journal of Operational Research, 240*(3), 791-806.

80. Michael F. Rotondo, C. C., R. Stephen Smith (2014). Resources for optimal care of the injured patient. Committee on Trauma, American College of Surgeons. Retrieved from https://www.facs.org/media/yu0laoqz/resources-for-optimal-care.pdf

81. Michigan Department of Health and Human Services (2022). Michigan Statewide Trauma System. Retrieved from https://www.michigan.gov/mdhhs/doing-business/trauma.

82. Motallebi Nasrabadi, A., Najafi, M., & Zolfagharinia, H. (2020). Considering short-term and long-term uncertainties in location and capacity planning of public healthcare facilities. *European Journal of Operational Research, 281*(1), 152–173.

83. Nasrabadi, A. M., Najafi, M., & Zolfagharinia, H. (2020). Considering short-term and long-term uncertainties in location and capacity planning of public healthcare facilities. *European Journal of Operational Research, 281*(1), 152-173.

84. Newgard, C. D., Fu, R., Zive, D., Rea, T., Malveau, S., Daya, M., Jui, J., Griffiths, D. E., Wittwer, L., & Sahni, R. (2016). Prospective validation of the national field triage guidelines for identifying seriously injured persons. *Journal of the American College of Surgeons, 222*(2), 146-158. e142.

85. Newgard, C. D., Hsia, R. Y., Mann, N. C., Schmidt, T., Sahni, R., Bulger, E. M., Wang, N. E., Holmes, J. F., Fleischman, R., & Zive, D. (2013). The trade-offs in field trauma triage: a multi-region assessment of accuracy metrics and volume shifts associated with different triage strategies. *The journal of trauma and acute care surgery, 74*(5), 1298.

86. Newgard, C. D., Mann, N. C., Hsia, R. Y., Bulger, E. M., Ma, O. J., Staudenmayer, K., Haukoos, J. S., Sahni, R., Kuppermann, N., & Western Emergency Services Translational Research Network (WESTRN) Investigators (2013). Patient choice in the selection of hospitals by 9-1-1 emergency medical services providers in trauma systems. *Academic Emergency Medicine, 20*(9), 911-919.

87. Newgard, C. D., Nelson, M. J., Kampp, M., Saha, S., Zive, D., Schmidt, T., Daya, M., Jui, J., Wittwer, L., & Warden, C. (2011a). Out-of-hospital decision-making and factors influencing the regional distribution of injured patients in a trauma system. *The Journal of trauma, 70*(6), 1345.

88. Newgard, C. D., Zive, D., Holmes, J. F., Bulger, E. M., Staudenmayer, K., Liao, M., Rea, T., Hsia, R. Y., Wang, N. E., & Fleischman, R. (2011b). A multisite assessment of the American College of Surgeons Committee on Trauma field triage decision scheme for identifying seriously injured children and adults. *Journal of the American College of Surgeons, 213*(6), 709-721.

89. Ohio Department of Public Safety. (2019) *Ohio trauma registry annual report.* https://www.ems.ohio.gov/links/2018_Annual_Trauma_Report_FNL.pdf

90. Pagès-Bernaus, A., Ramalhinho, H., Juan, A. A., & Calvet, L. (2019). Designing e-commerce supply chains: a stochastic facility–location approach. *International Transactions in Operational Research*, *26*(2), 507-528.

91. Parikh, P. P., Parikh, P., Guthrie, B., Mamer, L., Whitmill, M., Erskine, T., Woods, R., & Saxe, J. (2017). Impact of triage guidelines on prehospital triage: comparison of guidelines with a statistical model. *Journal of Surgical Research*, 220, 255-260.

92. Parikh, P. P., Parikh, P., Hirpara, S., Vaishnav, M., Sebastian, S., McCarthy, M. C., Jansen, J., & Winchell, R. J. (2022). Performance-Based Assessment of Trauma Systems: Estimates for the State of Ohio. *The American Surgeon*, 00031348211065095.

93. Physicians, A. C. o. E. (1987). Guidelines for trauma care systems. *Annals of Emergency Medicine*, *16*(4), 459-463.

94. Quintero-Araujo, C. L., Guimarans, D., & Juan, A. A. (2021). A simheuristic algorithm for the capacitated location routing problem with stochastic demands. *Journal of Simulation*, *15*(3), 217-234.

95. Rabbani, M., Heidari, R., & Yazdanparast, R. (2019). A stochastic multi-period industrial hazardous waste location-routing problem: Integrating NSGA-II and Monte Carlo simulation. *European Journal of Operational Research*, *272*(3), 945-961.

96. Reuter-Oppermann, M., van den Berg, P. L., & Vile, J. L. (2017). Logistics for emergency medical service systems. *Health Systems*, *6*(3), 187-208.

97. Røislien, J., van den Berg, P. L., Lindner, T., Zakariassen, E., Uleberg, O., Aardal, K., & van Essen, J. T. (2018). Comparing population and incident data for optimal air ambulance base locations in Norway. *Scandinavian journal of trauma, resuscitation and emergency medicine*, *26*(1), 42. https://doi.org/10.1186/s13049-018-0511-4

98. Rotondo, M. F., Cribari, C., & Smith, S. R. (2014). Resources for optimal care of the injured patient. *Committee on Trauma, American College of Surgeons.* Retrieved from https://www.facs.org/quality-programs/trauma/tqp/center-programs/vrc/resources.

99. Sadrani, M., Tirachini, A., & Antoniou, C. (2022). Vehicle dispatching plan for minimizing passenger waiting time in a corridor with buses of different sizes: Model

formulation and solution approaches. *European Journal of Operational Research*, *299*(1), 263-282.

100. Salman, F. S., & Yücel, E. (2015). Emergency facility location under random network damage. *Computers & Operations Research*, *62*, 266-281.

101. Sasser, S. M., Hunt, R. C., Faul, M., Sugerman, D., Pearson, W. S., Dulski, T., ... & Galli, R. L. (2012). Guidelines for field triage of injured patients: recommendations of the National Expert Panel on Field Triage, 2011. *Morbidity and Mortality Weekly Report: Recommendations and Reports*, *61*(1), 1-20.

102. Saveh-Shemshaki, F., Shechter, S., Tang, P., & Isaac-Renton, J. (2012). Setting sites for faster results: Optimizing locations and capacities of new tuberculosis testing laboratories. *IIE Transactions on Healthcare Systems Engineering*, *2*(4), 248.

103. Schmid, V. (2012). Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. *European Journal of Operational Research*, *219*(3), 611-621.

104. Shanahan, T. A., Fuller, G. W., Sheldon, T., Turton, E., Quilty, F. M. A., & Marincowitz, C. (2021). External validation of the Dutch prediction model for prehospital triage of trauma patients in South West region of England, United Kingdom. *Injury*, *52*(5), 1108-1116.

105. Shariff, S. S. R., Moin, N. H., & Omar, M. (2012). Location allocation modeling for healthcare facility planning in Malaysia. *Computers & Industrial Engineering*, *62*(4), 1000-1010.

106. Shishebori, D., & Babadi, A. Y. (2015). Robust and reliable medical services network design under uncertain environment and system disruptions. *Transportation Research*, *77*, 268-288.

107. Slama, I., Ben-Ammar, O., Dolgui, A., & Masmoudi, F. (2021). Genetic algorithm and Monte Carlo simulation for a stochastic capacitated disassembly lot-sizing problem under random lead times. *Computers & Industrial Engineering*, *159*, 107468.

108. Smith, H. K., Harper, P. R., & Potts, C. N. (2013). Bicriteria efficiency/equity hierarchical location models for public service application. *Journal of the Operational Research Society*, *64*(4), 500-512.

109. Stone, D. A. (1997). Policy paradox: The art of political decision making (Vol. 13). Ww Norton New York.

110. Syam, S. S., & Côté, M. J. (2010). A location–allocation model for service providers with application to not-for-profit health care organizations. *Omega*, *38*(3), 157-166.

111. Teixeira, J. C., & Antunes, A. P. (2008). A hierarchical location model for public facility planning. *European Journal of Operational Research*, *185*(1), 92–104.

112. Texas Department of State Health Services (2022). Trauma System Development. Retrieved from https://dshs.texas.gov/emstraumasystems/etrauma.shtm.

113. Tinkoff, G. H., O’connor, R. E., Alexander III, E. L., & Jones, M. S. (2007). The Delaware trauma system: impact of Level III trauma centers. *Journal of Trauma and Acute Care Surgery*, *63*(1), 121-127.

114. Toro-Díaz, H., Mayorga, M. E., Chanta, S., & McLay, L. A. (2013). Joint location and dispatching decisions for emergency medical services. *Computers & Industrial Engineering*, *64*(4), 917-928.

115. Van Buuren, M., Jagtenberg, C., Van Barneveld, T., Van Der Mei, R., & Bhulai, S. (2018). Ambulance dispatch center pilots proactive relocation policies to enhance effectiveness. *Interfaces, 48*(3), 235-246.

116. Van Laarhoven, J. J., Lansink, K. W., van Heijl, M., Lichtveld, R. A., & Leenen, L. P. (2014). Accuracy of the field triage protocol in selecting severely injured patients after high energy trauma. *Injury, 45*(5), 869-873.

117. Van Rein, E. A., Sadiqi, S., Lansink, K. W., Lichtveld, R. A., Van Vliet, R., Oner, F. C., Leenen, L. P., & Van Heijl, M. (2020). The role of emergency medical service providers in the decision-making process of prehospital trauma triage. *European journal of trauma and emergency surgery, 46*(1), 131-146.

118. Van Rein, E. A., van der Sluijs, R., Houwert, R. M., Gunning, A. C., Lichtveld, R. A., Leenen, L. P., & van Heijl, M. (2018a). Effectiveness of prehospital trauma triage systems in selecting severely injured patients: Is comparative analysis possible? *The American journal of emergency medicine, 36*(6), 1060-1069.

119. Van Rein, E. A., van der Sluijs, R., Raaijmaakers, A. M., Leenen, L. P., & van Heijl, M. (2018b). Compliance to prehospital trauma triage protocols worldwide: A systematic review. *Injury, 49*(8), 1373-1380.

120. Van Rein, E. A., van der Sluijs, R., Voskens, F. J., Lansink, K. W., Houwert, R. M., Lichtveld, R. A., de Jongh, M. A., Dijkgraaf, M. G., Champion, H. R., & Beeres, F. J. (2019). Development and validation of a prediction model for prehospital triage of trauma patients. *JAMA surgery, 154*(5), 421-429.

121. Voskens, F. J., van Rein, E. A., van der Sluijs, R., Houwert, R. M., Lichtveld, R. A., Verleisdonk, E. J., Segers, M., van Olden, G., Dijkgraaf, M., & Leenen, L. P. (2018).

Accuracy of prehospital triage in selecting severely injured trauma patients. *JAMA surgery*, *153*(4), 322-327.

122. Wang, H., Jin, Y., & Jansen, J. O. (2016). Data-driven surrogate-assisted multiobjective evolutionary optimization of a trauma system. *IEEE Transactions on Evolutionary Computation*, *20*(6), 939-952.

123. Wang, X. J., Yang, M., & Fry, M. J. (2015). Efficiency and equity tradeoffs in voting machine allocation problems. *Journal of the Operational Research Society*, *66*(8), 1363-1369.

124. World Health Organization(2021). Injuries and violence. Retrieved from https://www.who.int/news-room/fact-sheets/detail/injuries-and-violence

125. Yapicioglu, H., Smith, A. E., & Dozier, G. (2007). Solving the semi-desirable facility location problem using bi-objective particle swarm. *European Journal of Operational Research*, *177*(2), 733-749.

126. Yasenovskiy, V., & Hodgson, J. (2007). Hierarchical Location-Allocation with Spatial Choice Interaction Modeling. *Annals of the Association of American Geographers*, *97*(3), 496–511.

127. Yoshitomi, Y., & Yamaguchi, R. (2003). A genetic algorithm and the Monte Carlo method for stochastic job-shop scheduling. *International Transactions in Operational Research*, *10*(6), 577-596.

128. Zhang, Y., & Atkins, D. (2019). Medical facility network design: User-choice and system-optimal models. *European Journal of Operational Research*, *273*(1), 305-319.

129. Zhang, Y., Berman, O., & Verter, V. (2009). Incorporating congestion in preventive healthcare facility network design. *European Journal of Operational Research, 198*(3), 922-935.

130. Zhang, Y., Berman, O., & Verter, V. (2012). The impact of client choice on preventive healthcare facility network design. *OR Spectrum*, *34*(2), 349-370.

131. Zhang, Y., Berman, O., Marcotte, P., & Verter, V. (2010). A bilevel model for preventive healthcare facility network design with congestion. *IIE Transactions, 42*(12), 865-880.

132. Zahiri, B., Tavakkoli-Moghaddam, R., & Pishvaee, M. S. (2014). A robust possibilistic programming approach to multi-period location–allocation of organ transplant centers under uncertainty. *Computers & Industrial Engineering*, *74*, 139-148.

APPENDIX

Appendix A: Notional Tasking Algorithm to Estimate srUT and srOT

Figure 31 presents a schematic of the notional tasking algorithm. Accordingly, let $t_{TC\text{-}gnd}$ and $t_{TC\text{-}air}$ refer to the total time from field to the TC via ground and air, respectively, and t_{NTC} is the time from field to NTC via ground. While t_{in} and t_{load} refer inbound and loading time for the air ambulance, respectively. If t_{access} and t_{bypass} refer to the 'access' and 'bypass' thresholds, then

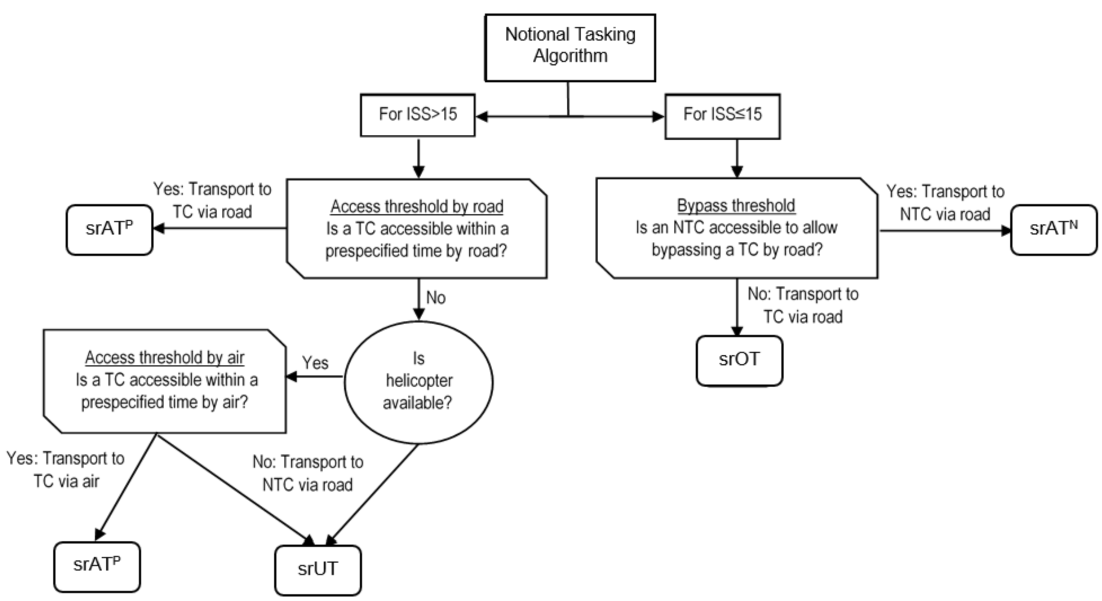


Figure 31. Notional Tasking Algorithm

- If ISS>15 (i.e., severe injuries), then
    - If $t_{TC\text{-}gnd} \leq t_{access}$, then transport to TC
    - Elseif (helicopter available), then
        - If $t_{TC\text{-}air} + t_{in} + t_{load} \leq t_{access}$, then transport to TC
    - Else transport to NTC (and mark the case as srUT)
- Elseif ISS $\leq$ 15 (i.e., less severe injuries), then
    - If $t_{NTC} - t_{TC\text{-}gnd} \leq t_{bypass}$, then transport to NTC
    - Else transport to TC (and mark the case as srOT)

Table 19. Illustration of Triage Classification by the Notional Tasking Algorithm

(t_{access} = 30 minutes and t_{bypass} = 15 minutes)

Instance	ISS	Ideal hospital	Time to nearest TC by road, $t_{TC\text{-}gnd}$ (mins)	Time to nearest TC by air, $t_{TC\text{-}air}$ (mins)	Time to nearest NTC by road, t_{NTC} (mins)	Likely EMS transport	Triage classification	Reason
1	18	TC	25	10	45	TC	AT^P	$t_{TC\text{-}gnd} \leq t_{access}$ TC is within access threshold by road
2	27	TC	40	15	55	TC	AT^P	$t_{TC\text{-}air} + t_{in} + t_{load} \leq t_{access}$ TC is within access threshold by air
3	24	TC	80	35	24	NTC	srUT	$t_{TC\text{-}gnd}$; $t_{TC\text{-}air} + t_{in} + t_{load} > t_{access}$ TC is not within threshold by road/air
4	10	NTC	30	-	16	NTC	AT^N	$t_{NTC} - t_{TC\text{-}gnd} \leq t_{bypass}$ NTC is within bypass threshold
5	14	NTC	25	-	8	TC	srOT	$t_{NTC} - t_{TC\text{-}gnd} > t_{bypass}$ NTC is not within bypass threshold

Table 19 presents a few representative cases to illustrate how the tasking algorithm helps classify a specific trauma incidence as appropriately triaged (AT^P for triaged to TC and AT^N for triaged to NTC) or mistriaged (srUT or srOT). In these cases, we assume t_{access} = 30 minutes and t_{bypass} = 15 minutes.

In Table 19 consider trauma incidence #1 with ISS>15, suggesting the need to transport this patient to the nearest TC. The algorithm first finds the nearest TC from the incident field in a given network and compares the EMS ground transportation to this TC ($t_{TC\text{-}gnd}$) to the 'access' threshold. Because $t_{TC\text{-}Gnd} < t_{access}$ = 25<30, driving to this TC is feasible, and so the case is categorized as AT^P. However, for incidence #2 also with ISS>15, $t_{TC\text{-}Gnd} > t_{access}$ (40>30), and so the possibility of air transportation is explored. The algorithm then compares the total flight time to this TC ($t_{TC\text{-}air}$), which accounts for inbound from the nearest helicopter base, patient loading, and outbound to the TC, with t_{access}. Assuming an inbound time of 5 minutes and a loading time of 5 minutes, the total air transportation time will result in 25 minutes. In this case, $t_{TC\text{-}air}<t_{access}$ (25<30), and thus this incidence is classified as transportation via air, also resulting in AT^P. But the total air transportation time incorporating inbound and loading time may not be feasible, as in the case of incidence #3 where $t_{TC\text{-}air} > t_{access}$ ({35+5+5} 45 >30), in which case the patient will be assigned to the nearest NTC by road, and the incidence will be classified as a srUT. Similarly, all the patients meeting the inclusion criteria are run through the tasking algorithm. A similar process is followed for patients with ISS≤15; air transportation is not considered as the injuries are less severe, in line with the actual EMS practice.

Appendix B: The Optimization Model to Estimate Coefficients of Patient Choice Utility Model

Our proposed model determines the coefficients of the utility model in order to minimize the misclassification of patients. A patient is misclassified if the destination hospital type estimated through the utility model differs from the actual destination type; e.g., a misclassification would be when a patient was taken to MTC (according to actual data), while the utility model's estimated choice is NTC. The optimization model is presented below with parameter and decision variables in Table 20 and 21, respectively.

Table 20. Parameters in the model

Notation	*Definition*
I	Set of trauma patients assigned via protocol criteria; $i \in I$
J	Set of candidate hospital locations (for MTC, ITC, and NTC); $j \in J$
L	Set of hospital type; $l \in L$; l = 1, 2, 3 represent MTC, ITC, and NTC, respectively
A^l	Attractiveness of hospital level *l*
C_i^l	1, if patient *i* chose hospital type *l*; 0, otherwise
TG_{ij}	Travel time from patient *i* to any candidate location *j* via ground
X_j^l	1, if a candidate location *j* is designated to be level *l*; 0, otherwise
M	Big number

Table 21. Decision variables in the model

Notation	*Definition*
β_1	Coefficient for the attractiveness of hospital
β_2	Coefficient for travel time between incidence location and hospital
u_{ij}, u_i^{max}	Utility of patient *i* receiving care at hospital *j*; $u_i^{max} = \max_j\{u_{ij}\}$
n_{ij}	1, if the highest utility for patient *i* occurs for a hospital *j*; 0, otherwise
m_i	1, if estimated choice through utility model is different than chosen hospital type; 0, otherwise

minimize: $\sum_i m_i$

subject to:

$$u_{ij} = \beta_1 \sum_l A^l X_j^l - \beta_2 \, TG_{ij}; \ \forall i \in I, \forall j \in J \tag{1}$$

$$u_i^{max} \geq u_{ij}; \ \forall i \in I, \forall j \in J \tag{2}$$

$$(u_i^{max} - u_{ij}) - M\,(1 - n_{ij}) \leq 0; \ \forall i \in I, \forall j \in J \tag{3}$$

$$\sum_j n_{ij} = 1; \ \forall i \in I \tag{4}$$

$$m_i \geq n_{ij}\, X_j^l - C_i^l\,; \ \forall i \in I, \forall j \in J, \forall l \in L \tag{5}$$

$$0 \leq \beta_1, \beta_2 \leq 1 \tag{6}$$

$$u_{ij}, u_i^{max} \in \mathbb{R}; \ \forall i \in I, \forall j \in J \tag{7}$$

$$n_{ij} \in \{0, 1\}; \ \forall i \in I, \forall j \in J \tag{8}$$

$$m_i \in \{0, 1\}; \ \forall i \in I \tag{9}$$

The objective of the model is to minimize the total misclassification of patients. Constraints (1)–(4) are similar to the NTNDP model to capture patients' choices using the utility model. For each patient *i*, Constraints (1) calculate the utility of receiving care at each hospital, Constraints (2) find maximum utility among all hospitals, while Constraints (3) and (4) record the hospital with the maximum utility. For each patient *i*, Constraints (5) record misclassification by comparing the estimated and actual hospital type. Constraints (6)-(9) define bound on decision variables.

The attractiveness for MTC, ITC, and NTC is set as 5, 3, and 1, respectively. We used 5627 cases assigned through patient choice criteria in the cleaned 2019 data from the state of Ohio, along with corresponding 2019 network of hospitals and their types. Further, we used ArcGIS to generate the ground travel time matrix and the Gurobi solver to find an optimal solution.

Appendix C: List of Abbreviations

ACS	American College of Surgeons
ACS COT	American College of Surgeons Committee on Trauma
$arAT^N$	Assessment-related Appropriate Triage Negative
$arAT^P$	Assessment-related Appropriate Triage Positive
arOT	Assessment-related Over-triage
arUT	Assessment-related Under-triage
BPSO	Binary Particle Swarm Optimization
CDC	Centre for Disease Control and Prevention
CP	Crossover Probability
efAT	Effective Appropriate Triage
efOT	Effective Over-triage
efUT	Effective Under-triage
EMS	Emergency Medical Services
FTDS	Field Triage Decision Scheme
FTG	Field Triage Guidelines
GA	Genetic Algorithm
ISS	Injury Severity Score

ITC	Intermediate Trauma Center
NTC	Non-trauma Center
MCS	Monte Carlo Simulation
MNL	Multinomial Logit
MP	Mutation Probability
MTC	Major Trauma Center
NBATS	Needs-Based Assessment of Trauma System
NTDB	National Trauma Data Bank
NTNDP	Nested Trauma Network Design Problem
ODPS	Ohio Department of Public Safety
OPTTDT	Ohio Prehospital Trauma Triage Decision Tree
OT	Over-triage
PBATS	Performance-Based Assessment of Trauma System
PC	Patient Choice
PSO	Particle Swarm Optimization
$srAT^N$	System-related Appropriate Triage Negative
$srAT^P$	System-related Appropriate Triage Positive
srOT	System-related Over-triage
srUT	System-related Under-triage

$srUT^S$	System-related Under-triage Stabilized
TC	Trauma Center
TCLP	Trauma Center Location Problem
TNDP-AM	Trauma Network Design Problem Considering Assessment-Related Mistriages
TRAMAH	Trauma Resource Allocation Model for Ambulance and Hospitals
TSA	Trauma Service Area
UT	Under-triage
WSM	Weighted Sum of Mistriages

CURRICULUM VITAE

NAME:	Sagarkumar Dhirubhai Hirpara
EDUCATION:	Ph.D., Industrial Engineering University of Louisville, Louisville, KY, 2022
	M.S, Industrial and Human Factors Engineering Wright State University, Dayton, OH, 2019
	B.E., Mechanical Engineering Gujarat Technological University, Ahmedabad, Gujarat, India, 2015
AWARDS:	Outstanding Research Award (IE department, University of Louisville, 2022)
	Institute of Industrial and Systems Engineers (IISE) • Lisa Zaken Award for Excellence Scholarship (2019) • Best Student Paper Award (Logistic and Supply Chain Track, 2019) • Best Track Paper Award (Facilities Design and Planning Track, 2020)
PROFESSIONAL POSITIONS:	Data Science Intern WestRock, Atlanta, GA, 2022