From Incident to Inpatient: How Healthcare Coalitions Can Improve Urban Incident Response

Alex F. Mills
Indiana University
Kelley School of Business
1309 E Tenth St
Bloomington IN 47405
(812) 855-3484
millsaf@indiana.edu

Jonathan E. Helm
Arizona State University
WP Carey School of Business
300 E Lemon St
Tempe AZ 85287
(480) 965-0931
jehelm2@asu.edu

Andres F. Jola-Sanchez
Indiana University
Kelley School of Business
1309 E Tenth St
Bloomington IN 47405
(812) 855-8449
ajolasan@indiana.edu

Mohan V. Tatikonda
Indiana University
Kelley School of Business
801 W Michigan St
Indianapolis IN 46202
(317) 274-2751
tatikond@indiana.edu

Bobby A. Courtney
Indiana University
Fairbanks School of Public Health
1050 Wishard Blvd
Indianapolis IN 46202
(317) 274-2000
bobcourt@iu.edu

Abstract

In recent years, many urban areas have established healthcare coalitions composed of autonomous (and often competing) hospitals, with the goal of improving emergency preparedness and response. We study the role of such coalitions in the specific context of response to multiple-casualty incidents in an urban setting, where on-scene responders must determine how to send casualties to medical facilities. A key function in incident response is multi-agency coordination. When this coordination is provided by a healthcare coalition, responders can use richer information about hospital capacities to decide where to send casualties. Using bed availability data from an urban area and a suburban area in the United States, we analyze the response capability of healthcare infrastructures under different levels of coordination, and we develop a stress test to identify areas of weakness. We find that improved coordination efforts should focus on decision support using information about inpatient resources, especially in urban areas with high inter-hospital variability in resource availability. We also find that coordination has the largest benefit in small incidents. This benefit is a new value proposition for healthcare coalitions, which were originally formed to improve preparedness for large disasters.

History: Received: December 2015; accepted: July 2017 by Nitin Joglekar after three revisions.

Keywords: Healthcare, Operations Management, Public Policy

---

1 Introduction

Urban areas present specific challenges and opportunities in responding to incidents where multiple persons need medical attention. Dense population leads to larger and more frequent incidents, and urban hospitals tend to be busier, leaving less slack capacity for incident response. On the other hand, cities tend to have many providers with a high level of skill and specialized resources. The presence of multiple public, private, and nonprofit entities providing rescue, transportation, and healthcare services leads to a major challenge in emergency preparedness: how to coordinate autonomous entities during a response.

In this article, we define an *urban multiple casualty incident (UMCI)* to be an incident with two or more persons needing medical attention, and multiple care resources involved in the response. The salient feature of a UMCI is not the absolute number of casualties, but the fact that multiple autonomous entities must work together to place them with an appropriate care resource. In our discussions with emergency medicine professionals, we learned that UMCIs are a weekly or even daily occurrence in urban areas, and that hospitals, which are a key resource in responding to UMCIs, are not commonly part of the decision-making structure during a response.

In the US, the federal government is the largest payer for healthcare services (Schoenbaum et al., 2003), even though most services are provided privately. This system structure means autonomous healthcare entities, such as hospitals, do not necessarily have direct incentives to improve community response to incidents. With this problem in mind, in the last decade, many communities have established *healthcare coalitions* with the aim of improving emergency preparedness and response. A healthcare coalition (HCC) is a "group of individual healthcare organizations in a specified geographic area that agree to work together to maximize surge capacity and capability during medical and public health emergencies by facilitating information sharing, mutual aid, and response coordination" (Barbera and Macintyre, 2007). Each coalition's structure and governance are determined by the members. Much of the initial funding for HCCs was provided by government grants, but the business models of HCCs are constantly evolving to meet the unique needs of their respective communities. Every state in the US has at least one HCC, and many robust HCCs are located in urban areas. Broadly speaking, HCCs engage in two main activities: (i) coordination in the management of emergency response, and (ii) training, education, and group purchasing to improve emergency preparedness. We study the former due to the central role of operations in coordinating emergency response.

A main challenge in UMCI response is to coordinate multiple autonomous entities with the goal of removing patients from the scene of the incident and placing them in hospitals with sufficient capacity and capability. The practice of coordinating multiple autonomous entities is called *multi-agency coordination* (MAC), and usually focuses on first responders such as police, fire, and emergency medical services (EMS), as well as government agencies. Hospitals traditionally are not active participants in MAC. Individually, hospitals perceive little direct benefit by coordinating with one another (e.g., by sharing information about bed availability). Reasons for this may include anti-trust concerns, fear of losing customers to a competitor, concern with serving their own patient panel rather than system efficiency, and unwillingness to provide access to proprietary data.

A recent trend in some urban areas is for HCCs to facilitate the medical MAC. This role is the key functionality that we study. We formulate a model of healthcare infrastructure vulnerability and evaluate it with historical bed availability data from one urban area and one suburban area. This analysis identifies factors that lead to high vulnerability in the region studied. We demonstrate the value of sharing information about bed availability in UMCI response, and we show that the most valuable type of information is frequently about inpatient resources (such as Intensive Care Units), illustrating that commonly used casualty distribution policies based on availability of ED beds often perform poorly.

Based on our results, we argue that HCCs would be most effective at improving incident response if their roles include collecting and analyzing information about hospital inpatient capacity, particularly information about inter-hospital variability and temporal variability, and tracking patients from multiple-casualty incidents to determine where they were sent and which resources within that hospital they used.

2 Institutional Contextualization

In the spirit of OM research in Industry Studies and Public Policy, we integrated direct observation and institutional contextualization into our analysis of the organizations involved in making emergency response decisions (Joglekar et al., 2016). To accomplish this, we conducted a preliminary round of interviews and on-site observations with key entities involved in incident response. On-site observations included Emergency Departments (EDs) at major urban hospitals, management offices of healthcare coalitions, and "ride-alongs" on an ambulance. Following these initial observations, we developed a semi-structured interview protocol and interviewed diverse participants in emergency response management. A full list of informants and a brief description of the interview protocol is given in Appendix A. From our semi-structured interviews, we summarized the key players, decisions, and actions taken in response to a UMCI in temporal sequence (see

Figure 8). In the remainder of this section, we use a specific example to further contextualize this research.

2.1 Example of UMCI response with MAC involving HCC: Garfield Park chemical spill

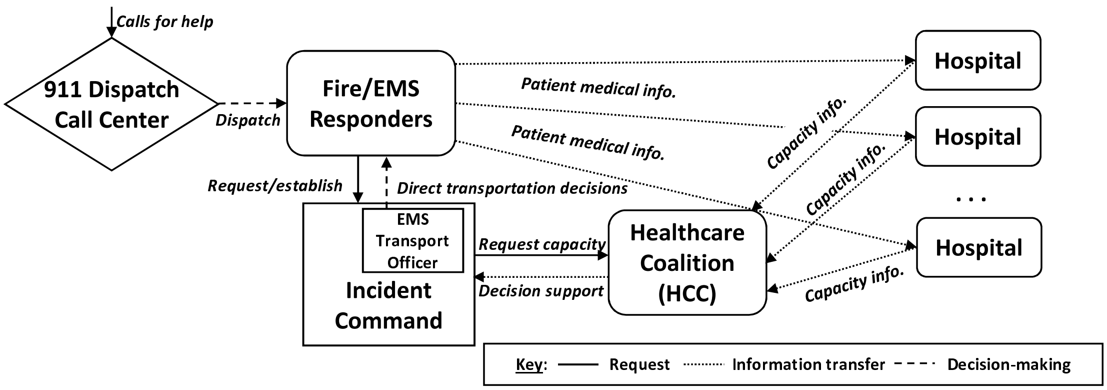
The following example details a chemical spill at the Garfield Park pool in Indianapolis, IN, with 71 total casualties. By interviewing the Chief of Operations for Indianapolis EMS, we constructed a basic timeline of the incident response and the coordinative role of MESH, which is an HCC that provides medical MAC functionality in Indianapolis, IN. The incident occurred in June 2012 when a chlorine dispenser malfunctioned and released a chlorine gas. Indianapolis 9-1-1 initially sent one ambulance and one fire truck to the scene. The first five minutes of the response unfolded as follows. After assessing the scene, responders initiated the hazardous materials protocol, which includes decontamination, triage, and treatment with oxygen. The EMS district lieutenant and 6 additional ambulances were dispatched to the scene. At the same time, the highest ranking person at the scene established an incident command and notified the MESH coalition. Staff at MESH received information about the number and severity of the casualties from on-scene responders and immediately began gathering information from local hospitals regarding available ED beds.

During the next five minutes, the Chief of Operations began making decisions about patient transportation with the aid of hospital capacity information provided by MESH. At the same time, hospitals initiated actions to expedite ED dispositioning (discharge, inpatient transfer, etc.). Within 10 minutes, the ten most critical patients had been transported from the scene to local hospitals that had available capacity to receive those patients. Because each ambulance can typically only take 2 casualties at a time, incident commanders requested and received buses from the local transit agency to transport less-severe patients. Approximately 24 patients were taken to hospitals by bus. The scene evacuation was complete between 45 and 50 minutes after the beginning of the response. We corroborated the events described above with three other individuals (some of whom were also informants in the interview protocol) and with local news media reports.

2.2 Benefit of MAC involving HCC

As the Garfield Park example shows, an HCC does not play a direct role in on-site response to a UMCI, but it may facilitate the MAC function, specifically involving hospitals in the coordination of multiple entities during the response. Our informants told us that an HCC's independence from the entities directly involved in the response gives its staff the ability to evaluate the situation in a way that is not subjective based on the potentially chaotic nature of the event, but objective based on the location, number of casualties, and available resources. When an HCC facilitates the MAC function in a community, HCC staff routinely monitor radio communication, hospital status (such as diversions), and news reports, with the goal of having an up-to-date system-wide viewpoint should a UMCI occur. When an incident occurs, the HCC provides general information or specific decision support to on-scene responders.

Figure 1: Information flow among autonomous entities in a UMCI response.



According to several of our informants, when hospitals do not participate in the MAC function, information flows only from the scene to the hospitals (see Figure 1 for a schematic diagram of information flow) once the patients are already en route, and on-scene responders must make decisions based on their experience. All of the individuals who we interviewed about the Garfield Park incident believed that the coordination by MESH provided substantial benefit to the response effort. Although different informants had different definitions of success, the informants cited the fact that MESH rapidly provided information on hospital capacities and capabilities to support decisions made by first responders. In particular, patients left the scene quickly and were able to be treated by appropriate providers *without overwhelming any of the hospitals.* At the same time, the hospitals were prepared to receive the patients because they had good information about how many patients they should expect and their conditions.

While it is conceivable that on-site incident commanders could attempt to collect capacity information from the hospitals themselves, in practice they have urgent patient care needs at the scene, and they typically lack a formal communication framework for obtaining this kind of information. HCCs provide this kind of framework: hospitals are *members* of an HCC and take advantage of its other services (like training and education). The HCC staff can leverage this working relationship to enable indirect information transfer from the hospitals to the responders, augmented by the HCC's knowledge of the system state and experience in disaster response. By using the HCC to provide decision support, incident command can make more informed decisions on patient distribution. Because MESH was founded in 2008 and our informants had many years of experience, the informants familiar with the Indianapolis area all stated that the level of coordination provided by MESH in the Garfield Park incident was substantially higher than the level of coordination that was achieved prior to the establishment of MESH's role in MAC. Still, questions remain regarding: how to quantify the benefits of MAC; what type of information enables a maximally effective response; and in what contexts and scenarios can HCCs be most effective in supporting MAC?

2.3 Research Objectives

From our in depth study of the institutional context of HCCs, we learned that there is an open question about the role of MAC in HCCs. We have evidence from our interviews that (1) the MAC functionality is important to emergency response, and (2) a specific HCC (namely, MESH) is able to facilitate the MAC functionality well. We also learned that HCCs often have good access to real-time hospital capacity information that could be used to facilitate MAC. However, we also learned that not all communities use an HCC to facilitate the MAC function (which in turn means that in those communities, hospitals are not active participants in MAC). Moreover, management lacks a way to quantify the benefits of effective coordination, which would allow us to assess the benefit of including hospitals in the coordinative effort.

In this research, we first develop a performance measure for emergency response that can quantify the operational benefit of multi-agency coordination in a UMCI in terms of patient access to care. Using this performance measure, we study the effect of increasing the intensity of MAC, specifically by using information about capacities at hospitals. We study how the community healthcare infrastructure, incident size, exogenous demand patterns, and hospital capacities affect the performance. We show how to use our model for several practical applications, including as decision support for hospital selection.

3 Literature Review

We now review three streams of the healthcare literature that especially inform the incident response problem: emergency response and healthcare coalitions, hospital capacity and patient flow, and EMS operations.

3.1 Emergency Response and Healthcare Coalitions

There are two especially foundational papers in the robust healthcare operations literature on pre-hosptial and in-hospital emergency preparedness and response. Barbera and Macintyre (2007) outline a comprehensive framework for managing medical resources during large-scale incidents. They define six tiers of resource management, ranging from individual healthcare assets, such as hospitals (Tier 1) to federal government resources (Tier 6). The authors place HCCs in Tier 2, stating an HCC "provides a central integration mechanism for information sharing and management coordination among healthcare assets". Courtney et al. (2009) review the Hospital Preparedness Program established by the U.S. Congress in 2002, which has led to the formation of many HCCs. The authors explain that HCCs provide the foundation for effective emergency response in mass casualty incidents.

As of 2013 the US had at least one HCC in each state and 496 HCCs total throughout all states and territories (Schmitt, 2016). These HCCs vary in structure and function. Not all HCCs provide MAC functionality in all kinds of emergencies, but many do undertake coordination activities. Table 1 illustrates types of coordination that different HCCs in the US provide, and shows the heterogeneity in HCC participants and coordination mechanisms. Given the above, there is a significant public policy opportunity for research that can demonstrate effective ways to use coalition capabilities.

While our empirical context is US HCCs, we note that many hospital and health IT coalitions also exist outside the US, including agencies that maintain relationships among autonomous healthcare providers. For example, in the EU, the European Hospital and Healthcare Federation (HOPE) is a hospital association representing local, regional and national health federations. Data collection and dissemination functionality similar to that studied in this paper is within the purview of such organizations. The fact that the US healthcare system is highly complex and decentralized, with individual hospitals having financial objectives potentially disincenting them from coordination activity, suggests it should be more feasible to establish coordination in more centralized, less complex healthcare systems.

Government regulations on information sharing are a key concern for healthcare coordination. The US and EU have especially intensive regulatory environments. US regulations define certain data as Protected Health Information (PHI) (Department of Health and Human Services, 2002). The EU has similar regulations on Personal Data (PD) (Council of European Union, 1995). The type of data considered in this study does not contain PHI or PD.

3.2 Hospital Capacity, Demand, and Patient Flows

One major factor in a healthcare infrastructure's ability to accommodate a sudden influx of demand is the *slack capacity* present in the system at the time of the incident. Hospital slack capacity exhibits predictable variability by time of day and day of week, with the middle of the week being very congested, while the beginning and end of the week are less congested (Harrison et al., 2005). Surgical schedule optimization helps

Table 1: Examples of US Healthcare Coalitions with Multi-Agency Coordination Functionality

Coalition name	Members	Coordination structure
MESH Coalition (Indianapolis)	35 hospitals, EMS, public safety.	Director of operations and emergency management coordinator on staff. Bed availability monitoring and reporting.
Michigan HCCs	8 healthcare coalitions: hospitals, EMS, and ancillary healthcare organizations.	Full-time coordinator and part-time medical director (per region). Fixed and mobile medical coordination centers.
Chicago Health System Coalition for Preparedness and Response	38 hospitals, EMS, 120 long-term care facilities, public health, and 5 regional health care agencies.	Interoperable communication systems and bed availability tracking.
Northern Virginian HCC	16 hospitals, fire, EMS, law enforcement and public health.	Incident notification, patient information sharing, patient transfer coordination, distribution of patients to hospitals.
Wisconsin HCCs	Hierarchical: local, regional, and state (7 regions). Public health, healthcare institutions, emergency management, and first responder agencies.	Bed capacity and medical capability database, incident monitoring and alert, information sharing, coordination of transport, situational awareness. Local medical coordination centers.
Missouri HCC	Public health agencies and EMS.	Medical incident coordination team. Information sharing: healthcare facility status, mobile medical asset status, statewide situational awareness. Text-message notifications.
Wyoming HCCs	5 HCCs: hospitals, public Health, EMS, long-term care providers, private sector healthcare providers.	Real-time information sharing: public health and medical information, situational awareness. Medical surge coordination: evaluation of medical infrastructure during incidents.
Northwest Healthcare Response Network (Seattle)	Healthcare organizations, state and local public health departments, and emergency response agencies.	Healthcare emergency coordination center. Information sharing and coordination before, during, and after emergency.

manage demand and balance workloads (Gallivan and Utley, 2005; Belien and Demeulemeester, 2007). Hospitals influence supply by making capital decisions on total beds (Green, 2002; Harper and Shahani, 2002) and operational decisions on staffing (Bard and Purnomo, 2005; Green et al., 2013).

We consider the Intensive Care Unit (ICU) and Medical/Surgical units (M/S) to be "downstream" resources because patients flow from the ED to one of these units. Related literature considers ICU admission control (Kim et al., 2014), load-based physician scheduling (Kazemian et al., 2014), and discharge decisions (Berk and Moinzadeh, 1998). Shi et al. (2015) develop in-hospital flow requirements from the ED to downstream resources. Chow et al. (2011) and Helm and Van Oyen (2014) propose surgical scheduling policies optimizing downstream patient flows through a network of hospital services.

We broaden this stream of literature by taking an empirical approach to study the distribution of patient demand, but with the main lever being coordination and information sharing rather than direct manipulation of hospital-controlled resources. We demonstrate the particular importance of coordinating downstream resources. We also consider a multiple-casualty situation, whereas most of the literature in capacity variability considers a single hospital during "normal" operations where patients arrive in unit increments.

3.3 Emergency Medical Services and Patient Transport

There is a rich literature on the operations of EMS. One stream of literature addresses unit response times to calls that occur randomly over a geographical service region, including problems such as allocating EMS units within a service region (Ingolfsson et al., 2008) and dynamically dispatching units based on availability (McLay and Mayorga 2013). UMCI casualties are concentrated at the disaster site(s) rather than spread geographically over a service region, so the objective of unit response time is of less concern. Since UMCIs occur in urban areas and urban areas have multiple hospitals, we assume that the hospitals are sufficiently close that travel distances will make a negligible impact on patient survival. Instead we focus our analysis on the role of information regarding hospital capacities, and on the ability for patients to receive necessary services in a resource constrained environment.

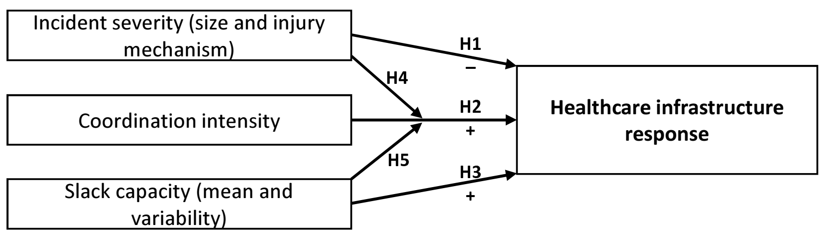
Recent work examines prioritization of patients for access to transportation resources in a disaster with a very large number of casualties. Mills et al. (2013) and Mills (2016) develop heuristic policies and Dean and Nair (2014) develop a mathematical program to distribute patients to hospitals. Neither approach considers uncertainty in available resources at the hospitals or downstream patient flow. These studies consider very large events and so use a mortality-related objective function. In contrast, UMCIs need not have a large number of patients. We consider the type of incident that may be operationally stressful for EMS and hospitals but for which patient volume does not substantially increase mortality.

A second stream of literature on EMS operations focuses on interactions between EMS and hospital EDs. "Ambulance offload delay" occurs when patients must physically wait inside the ambulance due to lack of ED capacity. Almehdawe et al. (2013, 2016) propose models where EMS patients are distributed to minimize offload delays. Deo and Gurvich (2011) and Allon et al. (2013) study ambulance diversion, a phenomenon where one hospital ED notifies EMS that it cannot accept additional patients, and identify system conditions where is likely to occur. These models, while applicable to day-to-day EMS operations, are less applicable in UMCIs. In the works cited above, patients arrive one at a time, which enables the decision-maker to re-route ambulances or change the hospital's diversion status, distributing the patients spatially across available hospitals. However, in a UMCI all patients arrive at once. Therefore, we focus on finding immediate bed availability for the patients, and on coordination policies that improve this outcome.

4 Model Development and Experimental Design

We study relationships among incident severity, coordination intensity, slack capacity, and healthcare infrastructure response (see Figure 2). Incident severity (size and injury mechanism) is exogenous and not controllable. In contrast, coordination intensity and slack capacity are potentially controllable, but with very different cost structure: adding slack capacity is slow, expensive, and permanent, while increasing coordination intensity is relatively quick, inexpensive, and flexible. While both of these variables' relationships to healthcare infrastructure response have management implications, we are primarily interested in finding cases where increasing coordination intensity can provide a significant benefit. Specifically, HCCs offer an opportunity to increase coordination intensity without any capital expenditures on the part of hospitals.

Figure 2: Conceptual framework of healthcare infrastructure response to UMCI.



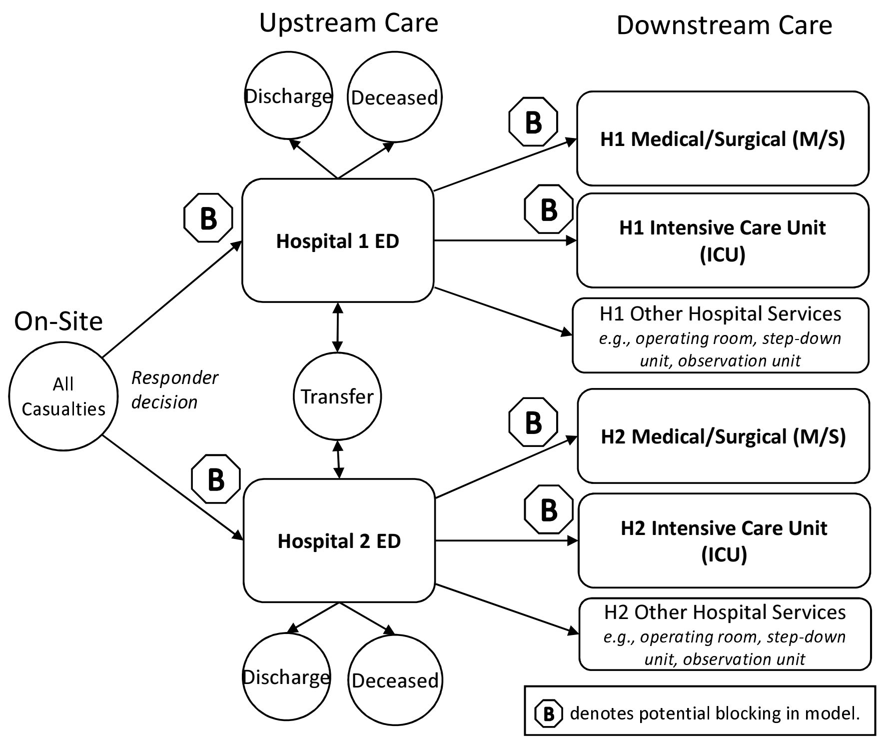
4.1 Measuring healthcare infrastructure response

The outcome variable, *healthcare infrastructure response*, is the probability a patient will be delivered to a hospital having sufficient available capacity to provide both emergency treatment and admit the patient (if needed). This variable differs from typical EMS performance evaluation, which considers only the timeliness with which patients are delivered to a hospital.

We operationalize healthcare infrastructure response by modeling a network of two hospitals with patient flow through an *upstream* resource (the ED) to multiple *downstream* resources (specifically, the Intensive Care Unit [ICU] and Medical/Surgical unit). When an incident occurs, responders decide how many casualties to send to each hospital. The two-hospital framework matches our case study with two trauma centers in an urban area and can be generalized to larger networks.

Once a casualty arrives at a hospital, we call her a *patient*, and we model her medical needs by a *patient* *flow pathway (PFP)*, which is a set of services that must be completed for her treatment to be successful. Based on our discussions with EMS managers and ED doctors, casualties involved in a UMCI will almost always require treatment in the ED. In our model, all PFPs include the ED as the upstream resource. If there are no staffed ED beds available for a patient, that patient is *blocked*. If the patient is not blocked upstream, she requires further service at a specific downstream resource with some probability; otherwise, she is deceased or discharged after ED treatment (see Figure 3).

Figure 3: Diagram of patient flows in the model.



PFPs occur according to a probability distribution that depends on the type of incident or the type of injuries sustained but not on the number of patients. We assume this distribution is known, but which downstream resource (if any) is required for a specific patient can only be determined after the patient enters the ED. Based on our interviews with EMS and ED personnel, this assumption is consistent with practice.

The performance variable, *patient flow pathway (PFP) completion rate*, is the fraction of all casualties whose PFP can be guaranteed to be completed. We calculate this rate by dividing the expected number of casualties who can complete their PFP based on the available resources by the number of casualties. To guarantee PFP completion, all services in the pathway must be available when the decision is made to transport the casualty to the hospital. Of course, patients spend time in one resource, during which existing patients might be discharged from the next needed resource, creating additional capacity. The resource availability data in our datasets is based on hospital personnel estimates of how many patients their hospital could accept if a UMCI occurred. In estimating capacity, hospital personnel consider projected discharges and arrivals. The only way to *guarantee* that a PFP can be completed is for all services to be available initially (e.g., when a patient spends as little as a few minutes in the ED before requiring ICU treatment). The PFP completion rate is therefore a conservative approximation of the proportion of patients who are treated without blocking.

Mathematically, we denote the number of casualties (hereafter, *incident size*) by m. At the time of the incident, hospital $i \in \{1, 2\}$ has available capacity c_i^j in resource $j \in R$, where $R = \{1, 2, \ldots, r\}$ is the set of hospital resources for some $r > 1$. Without loss of generality, we use index 1 to represent the upstream resource (ED). Therefore, $D = \{2, \ldots, r\}$ is the set of $r - 1$ downstream resources. Each capacity c_i^j is the realization of a random variable C_i^j. We will model the cases where all, some, or none of these realized capacities are known to the decision maker. The decision is (n_1, n_2), where n_i is the number of casualties to send to hospital i and $n_1 + n_2 \leq m$. A given patient requires only the upstream resource with probability $0 \leq q_1 \leq 1$, or she also requires the downstream resource type $j \in D$ with probability $0 \leq q_j \leq 1$, such that $\sum_{j \in R} q_j = 1$. That is, with probability q_1, a patient is discharged directly from the ED, while with probability $1 - q_1$, she first visits the ED and then visits one of the downstream resources. The distribution $\{q_j, j \in R\}$ varies by type of incident (see Section 5 for how we obtained these probabilities from the National Trauma Data Bank).

We denote by $S_i(n_i)$ the number of PFP completions at hospital i as a function of the number of casualties transferred there. Since the total number of casualties is known, we focus on computing the expected value of $S_i(n_i)$. We do so under three regimes: no information, information on ED capacity, and information on ED and inpatient capacity.

Lemma 1. *Let $B(n, p)$ denote a binomial random variable with n trials and success probability p. Then for a particular hospital i (whose subscript we drop to simplify the notation),*

1. *the expected number of successful PFP completions conditional on upstream and downstream capacity is*

$$\mathsf{E}\left[S(n) \middle| C^j = c^j,\ \forall j \in R\right] = q_1 n + \sum_{j \in D} \sum_{k=1}^{\min\{c^j, n\}} \mathsf{P}\left[B(n, q_j) \geq k\right], \quad \forall n \leq c^1, \tag{1}$$

2. *the expected number of successful PFP completions conditional on upstream capacity is*

$$\mathsf{E}\left[S(n)\middle|C^1 = c^1\right] = q_1 n + \sum_{j\in D}\sum_{k=1}^{n} \mathsf{P}\left[C^j \geq k\right] \mathsf{P}\left[B(n, q_j) \geq k\right], \quad \forall n \leq c^1, \tag{2}$$

3. *the unconditional expected number of successful PFP completions is*

$$\mathsf{E}\left[S(n)\right] = \sum_{k=1}^{n} \Bigg[q_1 \mathsf{P}\left[C^1 \geq k\right] + \sum_{j\in D} \mathsf{P}\left[C^j \geq k\right] \Bigg(\mathsf{P}\left[B(n, q_j) \geq k\right] \mathsf{P}\left[C^1 \geq n\right] + \sum_{c=k}^{n-1} \mathsf{P}\left[B(c, q_j) \geq k\right] \mathsf{P}\left[C^1 = c\right] \Bigg) \Bigg]. \tag{3}$$

See Appendix B for the proof of Lemma 1. We use the results of Lemma 1 to compute PFP completions in the experiment.

4.2 Factors and Levels

The experiment has three factors: incident type (size and mechanism), slack capacity, and coordination intensity.

Incident type. Incident type has two properties, size and mechanism. We systematically vary the incident size m from 2 to 100 casualties. We vary the incident mechanism, which indirectly varies the probabilities q_j of requiring various resources (detailed in Section 5).

Slack capacity. Slack capacity is the capacity available at each hospital in each service unit at the time of the incident. Slack capacity varies exogenously due to randomly occurring demand for emergency services, variability in patient length-of-stay, and scheduled procedures (i.e., elective surgeries). Some variability in slack capacity is predictable, allowing us to examine the effect of low, moderate, and high slack capacity on PFP completion without any data manipulation.

Coordination intensity. We vary coordination intensity along five ordinal levels.

Level 0 (no coordination). Responders must determine the number of casualties to send to each hospital right after the disaster occurs without information on available hospital service unit capacities. We assume responders use a naive policy of sending 50% of patients to each hospital. We calculate the expected PFP completion rate, $\left(\mathsf{E}\left[S_1(\lceil 0.5m \rceil)\right] + \mathsf{E}\left[S_2(\lfloor 0.5m \rfloor)\right]\right)/m$, using (3) to compute the expectation. Level 0 serves as the lower bound in our study.

Level 1 (historical upstream information). Responders have knowledge of the historical average number of available ED beds at hospital i. This is commonly observed in practice where EMS lacks access to recent or real-time information to support coordination, but does rely on previous experience. We assume that responders send patients to each hospital proportional to the historical average ED beds available (EMS experts corroborated to us that this is a common EMS practice). We calculate the expected PFP completion rate using (3). We consider Level 1 to be the "baseline" coordination intensity.

Level 2 (real-time upstream information). Upstream capacity information is available in real time through multi-agency coordination. Specifically, the coordinator observes realized capacity c_i^1 and forwards this information to responders, who can use it to help determine n_1 and n_2. Responders use a deterministic policy that maps (c_1^1, c_2^1) to (n_1, n_2) based only on the number of available ED beds. This policy captures the practical situation where a coordinator lacks access to all data. In most communities, EMS responders have little or no visibility into inpatient service unit capacities, even through an HCC. For the experiments experiments, we assume that $n_1 = \left\lceil \frac{c_1^1}{c_1^1 + c_2^1} m \right\rceil$ and set $n_2 = m - n_1$. We use (2) to compute the expected PFP completion rate $\left(\mathsf{E}\left[S_1(n_1) \middle| C_1^1 = c_1^1\right] + \mathsf{E}\left[S_2(n_2) \middle| C_2^1 = c_2^1\right]\right)/m$.

Level 3 (real-time upstream and downstream information). There is complete coordination between the EMS responders and the two hospitals. The coordinator observes realized capacities c_i^1 and c_i^j for all $j \in J$. Responders still do not know which casualties will require which downstream resources. Instead they use historical patient flow path probabilities to project downstream resource requirements. By leveraging this information, the coordinator can provide decision support to suggest the number of casualties to send to each hospital to maximize the service level. In our experiments, if $m \geq c_1^1 + c_2^1$, we send c_1^1 casualties to hospital 1 and c_2^1 casualties to hospital 2. Otherwise, we solve the following optimization problem:

$$
\begin{aligned}
\max_{n_1, n_2} \quad & \mathsf{E}\left[S_1(n_1) \middle| C_1^j = c_1^j,\ \forall j \in \{1\} \cup J\right] + \mathsf{E}\left[S_2(n_2) \middle| C_2^j = c_2^j,\ \forall j \in \{1\} \cup J\right] \\
\text{s.t.} \quad & n_1 + n_2 \leq m \\
& 0 \leq n_i \leq c_i^1 \quad i \in \{1, 2\} \\
& n_1, n_2 \text{ integer.}
\end{aligned}
$$

We divide the optimal objective, computed using (1), by m to obtain the PFP completion rate. This optimization is easily obtained using a line search on n_1.

Level 4 (pooled resources). All resources are pooled between the two hospitals. Theoretically this could be achieved if patient transfers are possible (e.g., when both hospitals belong to one integrated healthcare organization). We hasten to add that transfers between facilities with equivalent capabilities (i.e., transfers due only to capacity limitations) are medically and operationally undesirable, and are therefore unrealistic as a practical solution to the problem of geographically maldistributed slack capacity. We include this level as an upper bound on the achievable service level in the healthcare network, though full coordination at this level may be unachievable in many practical settings. We calculate the service level by pooling the two hospitals' capacities. For each $j \in R$, we define $c^j = c_1^j + c_2^j$. Thus, the pooled network behaves like a large single hospital and we compute the PFP completion rate using (1).

4.3 Hypotheses

We now present hypotheses regarding the effect of each experimental design dimension on healthcare infrastructure response performance. We test these hypotheses using empirical data (see Section 5).

Incident type. Incident type has two properties: size (i.e., number of casualties) and mechanism. Response effectiveness should be lower for larger UMCIs. Response effectiveness should also be lower with injury mechanisms that require greater levels of inpatient resources (reflecting more serious injuries). Thus:

Hypothesis 1a. *Larger incident size is negatively associated with healthcare infrastructure response.*

Hypothesis 1b. *Injury mechanism requiring greater use of inpatient resources is negatively associated with healthcare infrastructure response.*

Coordination intensity. The degree of information collection, sharing and use in responder decision-making varies from low coordination intensity (very little or no information collection and sharing) to high coordination intensity (extensive collection and sharing of high-quality, accurate, precise, granular, real-time information by an HCC). Greater coordination intensity offers responders greater awareness of the distribution of available bed capacity amongst community hospitals. Responders can use this information to better inform casualty distribution, especially by avoiding overload of one hospital when another has available capacity. We expect more effective response when there is more coordination:

Hypothesis 2. *Higher coordination intensity is positively associated with healthcare infrastructure response.*

Slack capacity. Healthcare infrastructures are not designed specifically for multiple casualty events. Each autonomous entity determines its own care design and capacity, typically with little awareness or consideration of the other entities' plans. This independent planning contributes to uneven distribution of bed availability across the community's hospitals. Non-stationary supply and demand (in both emergency and elective services) further contribute to temporal variability in slack capacity. While greater absolute slack capacity should improve the response, maldistribution of slack capacity should degrade performance (because each patient needs the correct resources to be available at her destination hospital, while resources at other hospitals are irrelevant to her).

Hypothesis 3a. *Higher slack capacity is positively associated with healthcare infrastructure response.*

Hypothesis 3b. *Higher inter-hospital variability in slack capacity is negatively associated with healthcare infrastructure response.*

Moderation Effects of Size and Slack Capacity Variation. We hypothesize that the (H2) positive effect of coordination intensity on response performance is diminished for larger incidents and magnified for smaller incidents. Coordination benefits arise from better use of existing available capacity, not from increasing capacity. When an incident is sufficiently large that there is not enough community capacity, coordinative activity cannot help:

Hypothesis 4. *The positive association between coordination intensity and healthcare infrastructure response is weaker for larger incidents than for smaller incidents.*

Greater coordination intensity enables better use of existing resources, so we expect higher inter-hospital variation in slack capacity to amplify the (H2) positive effect of coordination intensity on response:

Hypothesis 5. *The positive association between coordination intensity and healthcare infrastructure response is stronger in the presence of higher inter-hospital variation in slack capacity than in the presence of lower inter-hospital variation in slack capacity.*

5 Data and Methodology

This study employs several separate, proprietary, archival datasets: one large national trauma database (NTDB), daily census information from two autonomous trauma centers in the same urban area (case study 1), and daily census information from two autonomous hospitals in the same suburban area (case study 2).

5.1 National Trauma Data Bank

The American College of Surgeons (ACS) maintains a national database of ED visits at participating trauma centers in the United States. We obtained the 2012 ACS National Trauma Data Bank (NTDB) Research

Table 2: Emergency Department (ED) discharge disposition rates (%).

Disposition Classification	Motor Vehicle	Firearm	General
Inpatient bed (M/S)	36	24	50
Intensive care unit (ICU)	23	16	15
Operating Room	12	32	11
Discharged	14	11	7
Deceased	2	10	0
Other	13	7	16
Total patients	110,715	17,171	174,028

Italics indicate downstream resource considered in case study. Source: NTDB (2012).

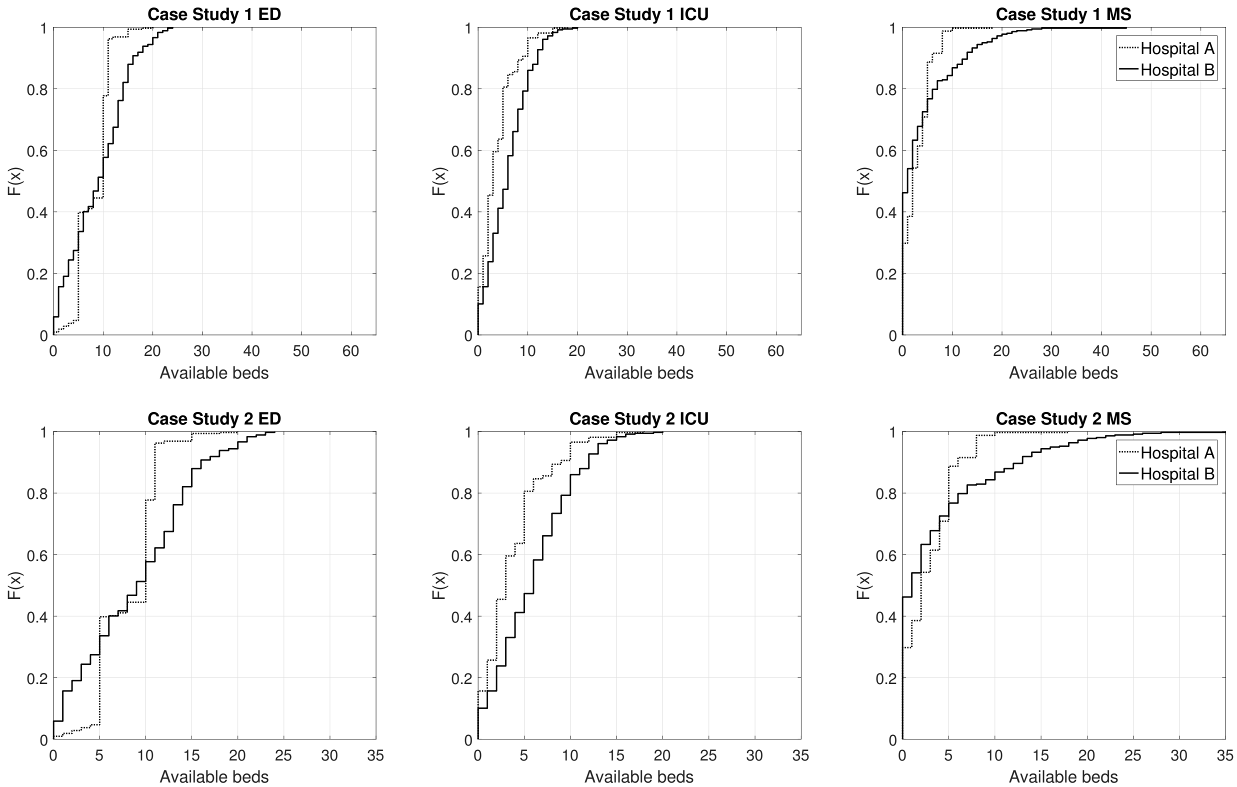
Data Set, which contains approximately 300,000 visit observations. We classified the observations in the database using ICD-9 E-Codes, each of which denotes a specific injury. We then mapped the E-Codes to injury mechanisms according to the method of the Centers for Disease Control and Prevention (CDC) (1997). We aggregated the 27 CDC mechanisms into three broad categories: (i) motor vehicle collisions, (ii) firearm injuries, and (iii) general accidents (such as falls, cuts, or chemical poisoning). We chose these three categories (which cover over 95% of the observations in the database) not necessarily to make specific recommendations for these types of incidents, but rather to get varying distributions on the types of services needed by the patients (and therefore different PFPs).

Table 2 shows the ED discharge disposition for the three trauma types described above. Although the NTDB provides an important resource for estimating the proportion of patients requiring downstream services, one limitation of the NTDB is that it is a convenience sample and only hospitals that choose to participate are represented. Despite that, frequencies in Table 2 demonstrate that different types of injuries result in widely varying requirements for downstream services, and because our two anonymous hospitals are also trauma centers (TC), we expect that this variability is also present at these two hospitals. In the case study, we take these probabilities as inputs to determine PFPs.

5.2 Hospital Census Data

We obtained daily bed availability data for two anonymous hospitals located in the same large urban area in the United States (case study 1, hospitals 1A and 1B) and two anonymous hospitals in the same suburban area in the United States (case study 2, hospitals 2A and 2B). Hospitals 1A and 1B have TC designations from the American College of Surgeons and both serve a similar catchment area. Hospitals 2A and 2B are somewhat smaller, but they still serve a highly populated catchment area and would be involved in the response to a multiple-casualty incident occurring within the large suburban area. Because all of the

Figure 4: Cumulative distribution of daily bed availability at hospitals used in case study 1 and in case study 2, for Emergency Department (ED), Intensive Care (ICU), and Medical/Surgical (MS).



hospitals studied consider their information proprietary, we could access the data only on the condition that their exact capabilities be disguised. The data span 15 months in 2014 and 2015. Each hospital's data contains information about the availability of ED and inpatient beds on a daily basis.

The hospitals used in the case studies do not formally share this information with each other. These data sets are representative of the kind of data that could reasonably be collected and used by an HCC. The data are self-reported and include resources that could be mobilized if needed (i.e., surge capacity). Table 3 shows summary statistics. We focus on the ED, ICU, and M/S beds because these are among the most common healthcare resources, and because both hospitals in each case study have sufficiently many beds in each of these units that deciding where to send casualties would be a reasonably complex decision problem.

A key concern for HCCs is the quality of the data provided by hospitals. For example, hospitals usually want to receive patients and therefore may overstate their capacity. Moreover, our data is captured only once per day, and thus has very coarse time granularity. Despite these identified limitations, our data exhibit clear day-to-day and inter-hospital variability in the number of available beds and provide sufficient detail to construct our high-level measure of healthcare infrastructure response (see Figure 4).

Table 3: Median number of available beds in each resource at hospitals in case study 1 and 2.

		Case Study 1		Case Study 2	
	Resource	H1A	H1B	H2A	H2B
Upstream	Emergency Department (ED)	30	26	10	9
Downstream	Intensive Care Unit (ICU)	18	0	3	6
	Inpatient Medical/Surgical (M/S)	20	0	2	1

Table 4: Median total resources available by day of week.

		Case Study 1					Case Study 2				
	Resource	Mon	Tue	Wed	Thu	Fri	Mon	Tue	Wed	Thu	Fri
Upstream	ED	55	52	58	57	58	16	18.5	18	19	19.5
Downstream	ICU	20	17	20	20	15	9	4	2.5	3	7
	M/S	20	15	13	22	30	11	11	9	8	10

In case study 1, we observe substantially different resource availability between the two hospitals. Hospital 1B is substantially more congested than Hospital 1A (see Table 3). Figure 4 shows the downstream capacity at Hospital 1A is frequently much higher than at Hospital 1B, although they have similar upstream (ED) capacity. In case study 2, while the hospitals overall have lower available capacity than in case study 1, available downstream capacity is evenly distributed between hospitals 2A and 2B.

In many healthcare services, the day of the week has a significant impact on availability of services. We focus on weekdays, when hospitals tend to be the busiest, and for which our data sets were more complete. While the number of available ED beds does not vary substantially by day of week, we see an impact on downstream resources, especially M/S (see Table 4). This difference is likely explained by scheduling of elective surgeries: the bias toward congestion within the inpatient resources later in the week can be caused by a large number of elective procedures performed early in the week.

5.3 Methodology

As we discussed in Section 4.2, we study three different factors in our experiment, two of which (incident type and coordination intensity) we vary systematically, and one of which (slack capacity) varies naturally in the data. For each combination of these factors, we use the formulas from Section 4.1 to compute the expected PFP completion rate using the empirical bed availability distribution extracted from the hospital census data described above. However, the computations in Section 4.1 do not provide a way to assess the variability of the PFP completion rate. Specifically, there are two sources of variability that affect the PFP completion rate: variability in bed availability (captured in the data) and variability in the actual patient flow paths. Therefore, in addition to computing the expected PFP completion rate, we also use Monte Carlo simulation to determine statistical significance. Specifically, for each combination of the levels, we perform 1,000 Monte Carlo simulations. In each simulation, the number of available beds in each unit of each hospital is drawn randomly according to the empirical distributions extracted from the census data, and each patient's PFP is sampled according to the probabilities determined from the NTDB. By using common random numbers, we construct paired T-tests to assess the statistical significance of the change in the PFP completion rate (and in the case of Hypothesis 4, the difference-in-differences in the PFP completion rate).

6 Results

Using the methodology of Section 5.3, we compute the expected PFP completion rate for four incident sizes by three injury mechanisms, five levels of coordination intensity, and three levels of slack capacity. UMCI size categories are defined in Table 5, slack capacity levels were defined based on observing the empirical bed availability, and mechanisms were associated with PFPs according to the data in Table 2.

Table 5: UMCI size classification.

UMCI size	Number of patients
Very Small	$2 \leq m \leq 10$
Small	$11 \leq m \leq 25$
Medium	$26 \leq m \leq 50$
Large	$51 \leq m \leq 100$

In Table 6, we present numerical results for case study 1, for the median number of patients in each size category. In Table 7, we present the results in the same way for case study 2, excluding large UMCIs because they always exceed the maximum combined ED capacity of the two hospitals (thus, additional hospitals would be included in the response in such an incident). Figure 5 summarizes the data differently, by plotting the PFP completion rate for more granular incident sizes, across different injury mechanisms and coordination intensities, but aggregating across all levels of slack capacity.

6.1 Impact of Incident Severity, Coordination Intensity, and Slack Capacity

Hypothesis 1 stated that increased incident size and severity would lead to lower PFP completion. Holding the other independent variables constant at any value, an increase in UMCI size results in a decrease in PFP completion rate. This trend can be seen in Table 5. The PFP completion rate is highest for firearm injuries, followed by motor vehicle and general accidents, due to the varying PFPs. Motor vehicle accidents result in more complex traumatic injuries than firearm incidents, which is reflected in Table 2: patients with motor

Table 6: Healthcare infrastructure response performance (PFP completion %) in case study 1.

	Mechanism	Vehicle			Firearm			General		
Coordination	Size Slack Cap.	H	M	L	H	M	L	H	M	L
Level 0	Very Small	80	80	77	87	87	84	77	76	73
	Small	72	73	69	82	82	80	68	68	64
	Medium	69	68	64	79	78	75	65	63	59
	Large	53	49	47	61	57	55	49	45	43
Level 1	Very Small	80	80	77	87	87	84	77	76	73
	Small	72	73	72	82	82	81	68	68	67
	Medium	69	68	67	79	78	77	65	63	61
	Large	53	49	48	61	57	56	49	45	43
Level 2	Very Small	72	73	72	82	82	81	69	70	67
	Small	71	71	69	81	81	80	67	66	64
	Medium	69	68	67	80	79	78	65	63	61
	Large	53	49	47	61	57	55	49	45	43
Level 3	Very Small	98	95	93	99	97	96	98	94	91
	Small	97	93	89	98	96	93	95	90	85
	Medium	84	80	78	91	88	86	81	75	72
	Large	54	50	48	63	59	57	51	46	44
Level 4	Very Small	98	96	94	99	97	96	98	94	92
	Small	97	94	91	98	97	95	95	91	87
	Medium	93	89	85	97	94	91	90	84	80
	Large	69	62	58	74	67	64	65	57	53

vehicle injuries are more likely to require treatment in the inpatient medical/surgical or intensive care units. Patients with general accidents are more likely to require medical/surgical than either other mechanism, but are less likely to require the ICU. Patients with firearm injuries are least likely to require an inpatient resource. Using the Monte Carlo simulation, we tested Hypotheses 1a and 1b and we found support for both hypotheses for all combinations of coordination and slack capacity (paired T-test, $\alpha = 0.01$).

Hypothesis 2 stated that an increase in coordination intensity would result in higher PFP completion rate. Here, the results are slightly more nuanced. If we exclude level 2 from the analysis, Hypothesis 2 is fully supported (see Table 8, which shows statistical significance results for case study 1). Using the Monte Carlo simulation, we also constructed 95% confidence intervals on the probability that the PFP completion rate would be (strictly or weakly) improved by increasing the coordination intensity (see Table 9 for the results for case study 1, motor vehicle type). These results show that coordinating based only on ED information is clearly less effective than coordinating on both ED and inpatient information. Specifically, level 3 and 4 provide significant improvement in expected PFP completion rate versus levels 0–2, with level 3 providing strict improvement over level 2 in the vast majority of very small, small, and medium size incidents. This result is of practical importance because EMS providers are primarily concerned with ED capacity, because their objective is to deliver each patient to the closest appropriate hospital with an available ED bed. However, assigning patient destinations based only on ED information performs relatively poorly.

Table 7: Healthcare infrastructure response performance (PFP completion %) in case study 2.

	Mechanism	Vehicle			Firearm			General		
Coordination	Size Slack Cap.	H	M	L	H	M	L	H	M	L
Level 0	Very Small	86	81	76	90	86	82	84	76	70
	Small	71	62	59	76	68	66	66	56	52
	Medium	43	37	36	47	42	41	39	33	31
Level 1	Very Small	85	79	74	89	84	80	82	74	68
	Small	71	63	59	76	69	66	66	56	52
	Medium	43	38	36	47	42	41	39	34	31
Level 2	Very Small	85	78	75	90	85	83	82	72	69
	Small	71	62	59	78	70	68	66	55	52
	Medium	42	36	34	47	41	40	38	31	29
Level 3	Very Small	95	91	87	97	94	92	93	87	83
	Small	78	69	65	83	76	73	73	63	58
	Medium	44	39	36	48	44	42	41	34	32
Level 4	Very Small	97	93	90	98	95	93	95	89	85
	Small	81	73	69	85	79	75	77	66	61
	Medium	47	41	39	50	45	43	43	37	34

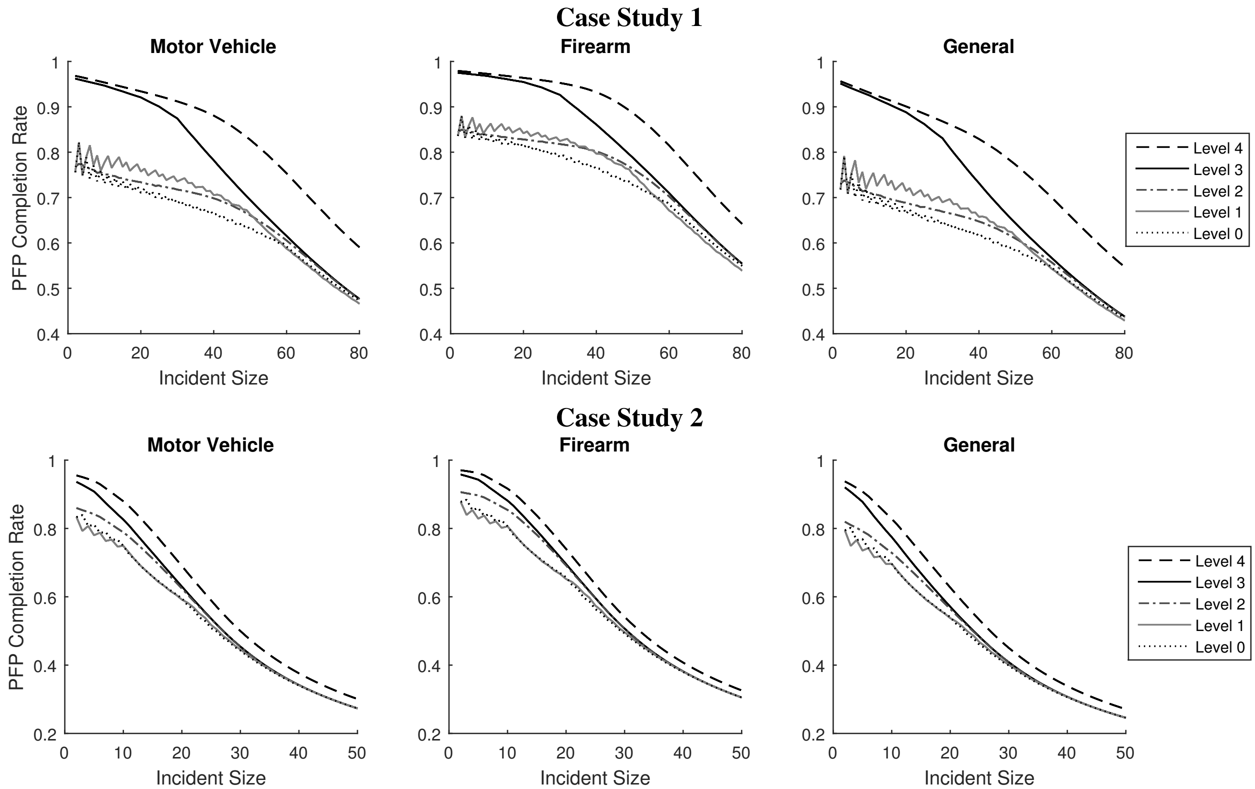
Table 8: Significance test for H2: increased coordination intensity increases PFP %. The symbol (-) denotes non-significance, while (*, **, *) denote significance at the (0.1, 0.01, 0.001) levels.**

	Mechanism	Vehicle			Firearm			General		
Coordination	Size Slack Cap.	H	M	L	H	M	L	H	M	L
Level 1 vs Level 0	Very Small*	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Small	***	***	***	***	***	***	***	***	***
	Medium	***	***	***	***	***	***	***	***	***
	Large	-	-	-	-	-	-	-	-	-
Level 2 vs Level 1	Very Small	-	-	-	-	-	-	-	-	-
	Small	-	-	-	-	-	-	-	-	-
	Medium	-	-	*	-	-	***	-	-	-
	Large	***	***	***	***	***	***	***	***	***
Level 3 vs Level 2	Very Small	***	***	***	***	***	***	***	***	***
	Small	***	***	***	***	***	***	***	***	***
	Medium	***	***	***	***	***	***	***	***	***
	Large	***	***	***	***	***	**	***	***	**
Level 4 vs Level 3	Very Small	-	**	***	-	*	***	-	**	***
	Small	***	***	***	***	***	***	***	***	***
	Medium	***	***	***	***	***	***	***	***	***
	Large	***	***	***	***	***	***	***	***	***

***Note:** For Very Small incident size, level 1 and level 0 are identical due to rounding.

In both case studies, we observe that using *real-time* ED information in level 2 offers surprisingly little improvement over the baseline level 1, which uses historical (non-real-time) information. In fact, we can find instances for which level 2 performs *worse* than level 1, for example, in case study 1, for the motor vehicle type with high slack capacity, level 2 performs worse than level 1 in more than half of small and medium sized incidents (at the 95% confidence level; see Table 9). See also the top-left panel of Figure 5, where the expected value of level 2 is below that of level 1 for some incident sizes. This phenomenon happens because much of the blocking occurs at the downstream resources, which are often more constrained. Level 2 will send casualties proportionally based on the number of available emergency department beds at the

Figure 5: PFP completion rate as a function of incident size, mechanism, and coordination policy.



Note: In coordination intensity levels 0 and 1, the policy is to send a fixed proportion of patients to hospital A. However, due to the discrete nature of the decision variables, rounding causes a deviation in the *actual* proportion of patients sent to hospital A, leading to oscillations in the graph. This phenomenon is particularly evident in small incidents, where the difference of one patient has a disproportionately large impact on the outcome.

Table 9: 95% confidence interval the probability that increased coordination improves (strictly or weakly) PFP completion.

			Strict Improvement Pr{Comparison > Baseline}			Weak Improvement Pr{Comparison ≥ Baseline}		
Comparison	Baseline	Size \ Slack	H	M	L	H	M	L
Level 1	Level 0	Very Small	(0.00, 0.00)	(0.00, 0.00)	(0.00, 0.00)	(1.00, 1.00)	(1.00, 1.00)	(1.00, 1.00)
		Small	(0.75, 0.80)	(0.64, 0.70)	(0.63, 0.68)	(1.00, 1.00)	(0.98, 1.00)	(0.97, 0.99)
		Medium	(0.83, 0.88)	(0.72, 0.77)	(0.72, 0.78)	(0.94, 0.97)	(0.89, 0.93)	(0.89, 0.93)
		Large	(0.03, 0.05)	(0.03, 0.06)	(0.07, 0.10)	(0.55, 0.61)	(0.68, 0.74)	(0.75, 0.80)
Level 2	Level 1	Very Small	(0.03, 0.06)	(0.03, 0.06)	(0.05, 0.08)	(0.77, 0.82)	(0.80, 0.85)	(0.86, 0.90)
		Small	(0.14, 0.18)	(0.11, 0.15)	(0.17, 0.22)	(0.39, 0.45)	(0.46, 0.52)	(0.57, 0.63)
		Medium	(0.20, 0.25)	(0.20, 0.25)	(0.28, 0.34)	(0.36, 0.42)	(0.46, 0.52)	(0.58, 0.64)
		Large	(0.41, 0.47)	(0.27, 0.33)	(0.21, 0.26)	(0.99, 1.00)	(0.99, 1.00)	(0.99, 1.00)
Level 3	Level 2	Very Small	(0.69, 0.75)	(0.60, 0.66)	(0.58, 0.64)	(1.00, 1.00)	(1.00, 1.00)	(1.00, 1.00)
		Small	(0.95, 0.97)	(0.90, 0.93)	(0.85, 0.90)	(1.00, 1.00)	(1.00, 1.00)	(1.00, 1.00)
		Medium	(0.87, 0.91)	(0.78, 0.83)	(0.74, 0.79)	(1.00, 1.00)	(1.00, 1.00)	(1.00, 1.00)
		Large	(0.05, 0.09)	(0.03, 0.05)	(0.01, 0.02)	(1.00, 1.00)	(1.00, 1.00)	(1.00, 1.00)
Level 4	Level 3	Very Small	(0.00, 0.00)	(0, 0.01.00)	(0.01, 0.03)	(1.00, 1.00)	(1.00, 1.00)	(1.00, 1.00)
		Small	(0.02, 0.04)	(0.05, 0.08)	(0.07, 0.10)	(1.00, 1.00)	(1.00, 1.00)	(1.00, 1.00)
		Medium	(0.83, 0.88)	(0.78, 0.83)	(0.72, 0.78)	(1.00, 1.00)	(1.00, 1.00)	(1.00, 1.00)
		Large	(0.94, 0.97)	(0.93, 0.95)	(0.90, 0.93)	(1.00, 1.00)	(1.00, 1.00)	(1.00, 1.00)

time of the incident, which will always perform better than level 1 in terms of ED completion. However, knowing nothing about inpatient availability, a policy that is based solely on real-time ED bed availability can backfire if there are few or no inpatient resources available at the hospital with more available ED beds. In our data for case study 1, it was often the case that hospital B had very little inpatient capacity (while hospital A had sufficient inpatient capacity) even though both hospitals had similar ED bed availability.

If we care about the ability to treat the patient successfully all the way from the incident to the resulting inpatient hospitalization, coordinating on ED information is insufficient, even in real time. This result highlights the importance of a system perspective that includes all aspects of a patient's treatment.

Table 10: Significance test for H3: increased slack capacity increases PFP %. The symbol (-) denotes non-significance, while (*, **, *) denote significance at the 0.1, 0.01, and 0.001 levels respectively.**

	Mechanism	Vehicle		Firearm		General	
Coordination	Size Slack Cap.	H vs M	M vs L	H vs M	M vs L	H vs M	M vs L
Level 0	Very Small	*	***	-	***	**	***
	Small	-	***	-	***	-	***
	Medium	***	***	***	***	***	***
	Large	***	***	***	***	***	***
Level 1	Very Small	*	***	-	***	**	***
	Small	***	***	*	***	***	***
	Medium	***	***	***	***	***	***
	Large	***	***	***	***	***	***
Level 2	Very Small	-	***	-	***	-	***
	Small	-	***	-	***	***	***
	Medium	***	***	***	***	***	***
	Large	***	***	***	***	***	***
Level 3	Very Small	***	***	***	***	***	***
	Small	***	***	***	***	***	***
	Medium	***	***	***	***	***	***
	Large	***	***	***	***	***	***
Level 4	Very Small	***	***	***	***	***	***
	Small	***	***	***	***	***	***
	Medium	***	***	***	***	***	***
	Large	***	***	***	***	***	***

Due to predictable variability in bed availability by day of week, we were able to study the system under different levels of slack capacity, confirming that when more slack capacity is available, system performance is generally better. Table 10 shows statistical significance for the hypothesis that increasing slack capacity leads to increased PFP completion. The difference is significant except when incident sizes are very small or small and slack capacity is medium (in which case there is already enough slack capacity that increasing it to high does not improve the outcome). In practical terms, this means that the healthcare infrastructure is more vulnerable in the middle of the week when slack capacity is low. We discuss the practical implications of predictable variability in more detail in Section 6.3.3.

To summarize, the results of the case studies broadly support Hypotheses 1(a-b), 2 (with the exception of level 2), and 3(a); namely, the PFP completion rate decreases in UMCI severity, and increases in coordination intensity and slack capacity.

6.2 Interaction Effects

Perhaps the result that is most relevant to HCCs is the interaction between UMCI size and the positive effect of coordination intensity on PFP completion rate. Hypothesis 4 states that increasing coordination will have a *smaller* positive impact on PFP completion rate for larger incidents. We test this hypothesis using the results of the Monte Carlo simulation by means of a paired T-test for the difference-in-differences. Specifically, for each coordination intensity level, we compute the difference in the improvement over the baseline between successive incident sizes. The results of this significance test for case study 1 are shown in Table 11. Again, if we exclude Level 2, Hypothesis 4 is fully supported: as incidents get larger, there is a significant decrease in the improvement due to coordination. The decrease in effect of coordination intensity is statistically significant for levels 3 and 4, for medium and large incidents, when level 1 is used as a baseline. In large incidents, each hospital reaches its own capacity under any reasonable transportation policy, and improvement in PFP completion can only be obtained by pooling inpatient resources (as in coordination intensity level 4). Visually, we can see this result by examining Figure 5 and Table 6: when the incident size is large, the differences between coordination levels is comparatively small.

Once again, level 2 proves to be an exception. This can be seen most clearly in Figure 5, where level 2 outperforms level 1 only when the incident size is medium. This result can be explained by thinking about the benefits provided by level 2 coordination, namely real-time ED capacity information. On the one hand, real-time ED capacity information does not provide much of a benefit when the incident size is sufficiently small, because the EDs in both case studies almost always had enough capacity to take on a few patients—adding information does not change this. On the other hand, when the incident size is larger, the system becomes more constrained, with the most constrained resources being inpatient services. Level 2 coordination does not provide information about those resources, and hence performs poorly.

The interaction effect highlights the large benefit offered by real-time information sharing for very small and small incidents–as much as 20 percentage points for the motor vehicle mechanism in case study 1. This result is important because smaller incidents occur more frequently, underscoring the potential impact of information collection and dissemination by HCCs on a daily basis, rather than just in large disaster situations. This result runs somewhat counter to the typical practice of MAC in communities lacking an

Table 11: Significance test for H4: larger incident size reduces the improvement due to coordination. (-) denotes non-significance, and (*, **, *) denote significance at the (0.1, 0.01, 0.001) levels.**

	Mechanism	Vehicle			Firearm			General		
Comparison level	Size Slack Cap.	H	M	L	H	M	L	H	M	L
Level 2 (vs Level 1)	Very Small vs Small	***	***	**	***	***	*	***	***	**
	Small vs Medium	-	-	-	-	-	-	-	-	-
	Medium vs Large	-	-	-	-	-	-	-	-	-
Level 3 (vs Level 1)	Very Small vs Small	-	-	*	-	*	*	-	**	***
	Small vs Medium	***	***	***	***	***	***	***	***	***
	Medium vs Large	***	***	***	***	***	***	***	***	***
Level 4 (vs Level 1)	Very Small vs Small	-	-	-	-	-	*	-	-	*
	Small vs Medium	-	-	-	-	-	-	**	*	***
	Medium vs Large	***	***	***	-	***	***	***	***	***

HCC, such as opening a command center only after a large disaster.

Turning to Hypothesis 5, which postulates that greater inter-hospital variability in slack capacity *increases* the positive effect of coordination intensity on PFP completion rate, we find supporting evidence from comparing the two case studies. Figure 4 shows that case study 1 has asymmetric inpatient slack capacity: bed availability is heavily concentrated at hospital 1. On the other hand, case study 2 has more evenly distributed inpatient slack capacity. We find that the positive effect of coordination intensity is more than twice as large in the asymmetric case as in the evenly distributed case (compare the upper and lower panels in Figure 5, being careful to note differences in scale on both the x and y axes).

Because inter-hospital variability in capacity does not naturally vary within each case study, we conducted a counterfactual analysis with the data of case study 1 to assess the statistical significance of H5 using Monte Carlo simulation. Specifically, we repeated the Monte Carlo simulation for case study 1 and artificially reduced the inter-hospital variability in ICU bed availability by taking the total ICU availability for each simulation and dividing it evenly between the two hospitals. Doing so results in higher PFP completion rate and an increased effect of coordination intensity on PFP completion rate (at the 0.01 level for all other combinations of the other factors), providing further support for Hypothesis 3(b) and Hypothesis 5.

6.3 Practical Applications

In this section, we demonstrate three practical applications for our model using the results of the experiment on case study 1.

6.3.1 Stress test for healthcare infrastructure vulnerability.

One use for our model is to assess the vulnerability of a healthcare infrastructure. We now demonstrate a way to make the results accessible for practitioners. We associate a *healthcare infrastructure stress level* with each of the four UMCI sizes in Table 5. We introduce a user-defined service level, denoted by SL, which is the is the minimum acceptable PFP completion rate, and we determine the largest incident size m for which the PFP completion rate is at least SL. If m is large [medium, small, very small], we say the healthcare infrastructure stress level is minimal [low, medium, high].

Table 12: Care infrastructure vulnerability: results of stress test on case study 1.

		Vehicle					Firearm					General				
SL	Coordination	M	T	W	R	F	M	T	W	R	F	M	T	W	R	F
0.75	Level 0	●	●	●	●	●	○	○	○	○	–	●	●	●	●	●
0.75	Level 1	●	●	●	●	●	○	○	○	–	–	●	●	●	●	●
0.75	Level 2	●	●	●	●	●	–	○	○	–	–	●	●	●	●	●
0.75	Level 3	○	○	○	○	○	–	–	–	–	–	○	○	○	○	○
0.75	Level 4	–	–	–	–	–	–	–	–	–	–	–	○	○	–	–
		M	T	W	R	F	M	T	W	R	F	M	T	W	R	F
0.85	Level 0	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
0.85	Level 1	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
0.85	Level 2	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
0.85	Level 3	○	○	○	○	○	○	○	○	○	○	○	◐	◐	○	○
0.85	Level 4	○	○	○	○	–	–	–	–	–	–	○	◐	○	○	○
		M	T	W	R	F	M	T	W	R	F	M	T	W	R	F
0.95	Level 0	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
0.95	Level 1	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
0.95	Level 2	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
0.95	Level 3	●	●	●	◐	◐	◐	●	◐	○	○	●	●	●	●	◐
0.95	Level 4	◐	●	●	◐	○	○	◐	◐	○	○	●	●	●	●	◐

Key: – minimal stress, ○ low stress, ◐ moderate stress, ● high stress.

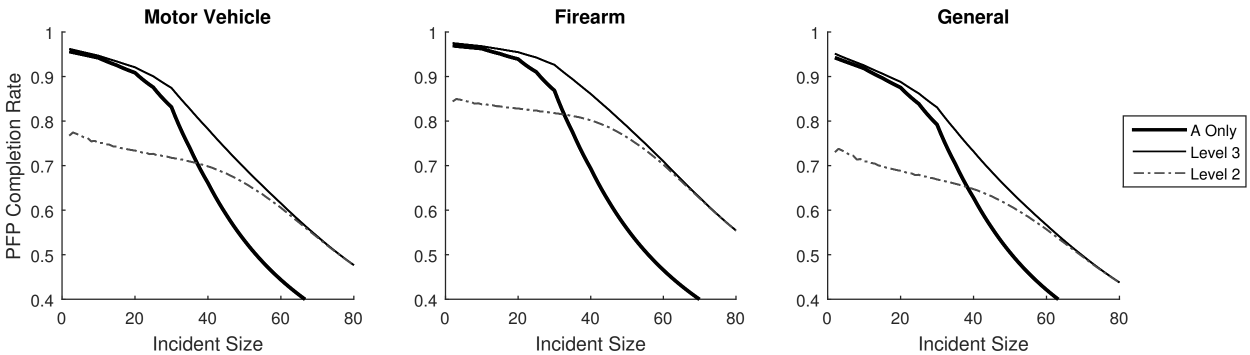
Results of the stress test for case study 1 are presented in Table 12 for three different values of SL. The stress test highlights many of the results we discussed in Sections 6.1 and 6.2. For example, we observe higher stress levels in the middle of the week, even with high coordination intensity, due to the predictable variability in slack capacity. Since midweek congestion is driven by hospital operational policies, this result highlights a further opportunity to involve hospitals in incident response.

The healthcare infrastructure stress test and the presentation of the data in the format of Table 12 provides several benefits for managers. First, it allows them to define a specific target service level that corresponds to an organizational goal. Second, it provides a quick comparison across different mechanisms and days of the week. Finally, it enables the identification of weaknesses through quick visual inspection.

6.3.2 Decision support for hospital inclusion in response.

One strong application for our model is to support the hospital inclusion decision. In a UMCI, it is not always immediately clear which hospital(s) (or how many hospitals) should be involved in the response. Calculating the PFP completion rate for different hospitals, or for different combinations of hospitals, can support this decision. For example, in case study 1, it is clear from Figure 4 that hospital B is often close to capacity in its inpatient services. Therefore, responders might wonder whether it is worthwhile to include hospital B at all. Moreover, this decision depends on the level of coordination intensity.

Figure 6: PFP completion rate as a function of incident size, disaster type, and coordination policy.



We calculated the expected PFP completion rate for hospital A alone and compared it with the PFP completion rate for the network of two hospitals given in Table 6 and Figure 5. The results are shown in Figure 6. We see that at level 3 (real-time ED plus inpatient information) , it is always beneficial to include hospital B, although the benefit is minimal when the incident size is Very Small. However, if coordination is limited to level 2 (real-time ED information), it is better to use *hospital A alone* when the incident size is less than about 40 patients. Once the incident size is large enough, hospital B should be included even if limited coordination is available. This application can be repeated for different hospitals and different size networks. For instance, in case study 1, once the incident size is large enough that all the coordination levels converge (in Figure 5), it would probably be worthwhile to include a third hospital, even though we did not have sufficient data to explore this possibility.

6.3.3 Information exposes the hidden problem of variable hospital workloads

The consideration of slack capacity in Hypothesis 4 is particularly relevant to the practical issues facing urban healthcare infrastructures because slack capacity in hospital inpatient units varies significantly by day of the week. This predictable variability in workloads is a nearly ubiquitous feature of hospital services, and is often a result of scheduling practices that concentrate elective procedures on certain days of the week (see, e.g., Helm and Van Oyen 2014).

Our results show that increased coordination intensity does not mitigate the impact of predictable variability. In fact, it intensifies it (see Figure 7). The relative effect of workload variability increases with more

Figure 7: PFP completion rate by UMCI size and day of week (Motor Vehicle incident type).

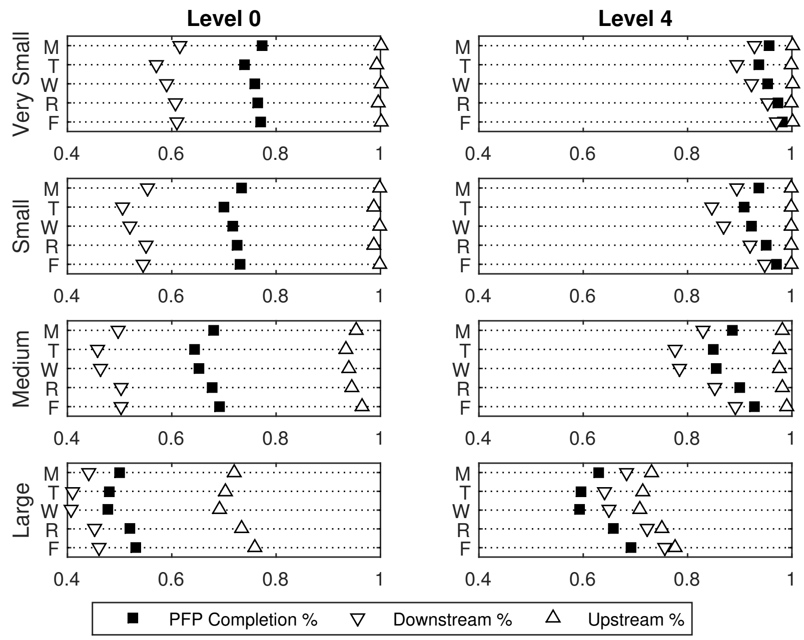


Table 13: Percent change in coefficient of variation (CV) of PFP completion across days of the week (case study 1, motor vehicle and firearm mechanisms).

	Motor Vehicle				Firearm			
	Very Small	Small	Medium	Large	Very Small	Small	Medium	Large
Level 3 vs Level 0	21%	55%	24%	1%	14%	46%	9%	0%
Level 4 vs Level 0	5%	30%	24%	37%	-4%	19%	11%	32%

information and better coordination, revealing that the system is more vulnerable to resource constraints caused by unbalanced hospital workloads. This result makes sense because resources (and in particular, downstream resources) are more highly utilized when care is coordinated.

Table 13 shows the percent increase in the coefficient of variation (cv) of PFP completion rate across days of the week caused by implementing a high level of coordination intensity. Specifically, level 3 coordination intensity can result in PFP completion rates that are as much as 55% more variable than the baseline.

The operational explanation for this phenomenon is that the system suffers from serious capacity problems on certain days of the week (most notably Tuesday and Wednesday), but such capacities are less likely to be reached in the uncoordinated case, due to patient maldistribution that occurs because of the lack of information. We argue that predictable variability in workloads creates unnecessarily high levels of stress on days when inpatient resources are highly utilized. As providers and HCCs improve operational response to incidents by increasing coordination and information sharing, other operational issues, namely predictable midweek capacity limitations, will be exposed. This result highlights the importance of continuous monitoring and evaluation of performance under increasing coordination.

There is a clear interaction between predictable and random variability, the latter of which is increased when there is more uncertainty about the type of resources a patient will need. Comparing the results for motor vehicle mechanism to firearm mechanism in Table 13, we see that the negative effect of predictable variability is driven by the availability of inpatient resources, because the increase is smaller for the firearm mechanism, which is less likely to require inpatient resources.

In addition to concerns about temporal variability, policy makers should understand how regulations affect inter-hospital variability in resource availability, which was a driver in healthcare infrastructure response in our study. For example, some jurisdictions require justification (e.g., a Certificate of Need) before expanding the physical space in a hospital, which limits the ability of hospitals to adjust capacity of inpatient services and could therefore lead to increased inter-hospital variability in resource utilization.

7 Implications for Practice and Public Policy

Public policy affects operational decisions made by organizations, and in turn context-specific observations about operational problems inform public policy (Joglekar et al., 2016). We study healthcare coalitions (HCCs), which are organizations created by public policy. In turn, the results of this study should inform policy makers interested in improving emergency response.

MAC should be included in urban HCC. HCCs were created with the goal of improving disaster response, but each HCC is established by members in a local or regional area, and so HCCs vary in terms of structure and function (see Section 3.1). Specifically, some HCCs include multi-agency coordination (MAC) as a core functionality and others do not. Our study supports the conclusion that MAC is valuable in an UMCI; furthermore, higher levels of coordination can only be provided by involving a HCC. Coordination intensity levels 0 and 1 can definitely be achieved by EMS acting alone, while levels 3 and 4 require hospital involvement, in turn making it unlikely that they can be implemented without participation of an HCC. The feasibility of level 2 coordination intensity without an HCC is debatable, but we found that coordinating based on ED information alone (i.e., level 2) provides a limited benefit compared to levels 3 and 4. Therefore, we recommend that in an urban setting, MAC should be a core functionality of an HCC.

MAC should be expanded to cover small incidents. In large incidents (and for case study 2, medium incidents), the improvement due to information sharing is limited. This occurs because when the entire system is overwhelmed, the sheer number of patients ensures that both hospital EDs will be fully utilized. Once the upstream resources reach their capacity, the resulting expected usage of downstream resources is the same for all coordination intensities. In contrast, when a metropolitan area experiences inter-hospital variability in resource availability, coordination through information sharing can provide a surprisingly large benefit even in very small incidents (those with only a few patients).

We emphasize that this conclusion is a potential new value proposition for HCCs: although these organizations were initially conceived to prepare for large disasters, they can have a substantial benefit in more routine incidents, suggesting that communities should find ways to more closely integrate HCCs into daily emergency response.

HCCs should collect information about inpatient resources. In Sec. 6.1 we observe several results that suggest that bed availability at *downstream* resources (inpatient medical/surgical and intensive care units) are the primary drivers of PFP completion in a metropolitan area where those resources are congested.

We find that real-time information about ED bed availability had disappointingly poor performance compared with sharing only historical mean bed availability at the ED in both case studies (see Figure 5). The relatively small benefit of real-time ED capacity information versus historical information is not because the information itself is not valuable, but rather because the patterns of congestion in hospitals tend to create bottlenecks in highly-utilized inpatient resources, such as the ICU. In both of our case studies, inpatient resources were more limited than ED resources. Unlike in studies that look at only one hospital, however, we find the presence of high inter-hospital variability in inpatient resource availability makes information about those resources much more valuable. For example, in the data for case study 1, we can find days where hospital A has several ICU beds available, while hospital B has none, and we can also find days where hospital A has none and hospital B has several. This inter-hospital variability makes it difficult to predict which hospital is going to have available ICU beds without real-time information. Moreover, because ICU beds are usually extremely limited, the value of knowing which hospital has available ICU beds is large.

Recall from Section 2.2 that placing MAC functionality within an HCC enables hospitals to participate actively in coordination. Our results clearly show that MAC should incorporate information about downstream resources *in addition to* information about ED bed capacity. It will be difficult, if not impossible, to collect this information in a useful way without the active participation of hospitals. Therefore, our results suggest there is a clear benefit if the MAC function is assigned to an HCC, or if HCCs are a major participant in MAC, because doing so allows downstream information to be leveraged.

Information sharing is insufficient in large incidents. In a completely integrated system, patients who receive upstream care at one hospital can move to a downstream resource at the other hospital. Inter-hospital flow of trauma patients would be rare in practice for both medical and operational reasons, but it might occur if both hospitals are part of the same healthcare delivery network, or if the HCC is given broad latitude to coordinate inter-hospital transfers due to the scale of the emergency. We argue that coordination intensity level 4 is more representative of a theoretical maximum system performance than a realistically achievable outcome in most instances, but this option shows that it is possible to make additional improvements using an extreme mechanism (hospital-to-hospital transfers precipitated only by capacity constraints). In a very large incident, mobilizing this more expansive level of coordination may be justified.

Our observations about the power and limitations of information sharing lead us to two conclusions about coordination in incident response. For small incidents, most of the benefit can be achieved through information sharing, which is relatively inexpensive (in terms of both monetary cost and palatability to competing organizations). On the other hand, for very large incidents, only full integration of care resources provides much improvement over any level of information sharing. During routine operations, most competing hospitals would be unwilling to participate in full care integration. However, large incidents are comparatively rare, suggesting that this extreme measure need not be employed frequently.

Increasing coordination provides substantial value for marginal cost. Our results show that HCCs can provide value by providing the MAC function, expanding this function to include smaller, non-disaster-type, incidents (e.g., UMCIs), and expanding information collection and dissemination to include data on inpatient capacities. Before implementing these policy suggestions, HCCs and policy makers must consider the effectiveness of these measures compared to the marginal cost of implementation. This value becomes apparent considering the fact that most of the costs of running an HCC are sunk and smaller incidents occur more frequently than large disasters.

Recall from Section 3.1 that HCCs were originally funded to improve disaster preparedness and response. As such, they already have much of the infrastructure required for MAC. For example, many HCCs have invested in information systems, data collection and monitoring capability, and communication equipment and processes. See Table 1, in which all of the example HCCs have some kind of technology, staff or facilities dedicated to coordination. Given that most HCCs already have invested in coordination capabilities, extending their reach to cover smaller incidents (that occur more frequently) would require minimal variable costs. In fact, some HCCs already have a full-time coordinator (for example, MESH Coalition; see Table 1). Those that do not may need to add such a staff position to monitor and disseminate relevant data more actively. We estimate this position would be at most one full-time equivalent, and therefore a relatively small portion of the HCC's total budget.

Expanding the MAC function to include smaller incidents would result in a much higher utilization of the HCC's coordination capability. Data from the National Highway Traffic Safety Administration (2008) show that between 2005 and 2007 (the most recent years for which this data is available), over half a million motor vehicle crashes in resulted in two or more casualties transported to hospitals (see Table 14). We searched the EM-DAT international disaster database (Guha-Sapir et al., 2017) over the same time period and geographic area, which returned just 18 disasters with patient injuries (see Table 15). Although neither of these databases covers all incidents that would meet the definition of UMCI, the difference in scale demonstrates that there is tremendous opportunity for increasing the utilization of coordination capabilities that already exist in many HCCs. Expanding coordination to cover smaller incidents would allow HCCs to provide more "bang for the buck" by amortizing their high fixed costs over more events. Moreover, the increased use of these capabilities provides more opportunities to practice coordinating, potentially providing locality-specific lessons that can be applied in the event that a large disaster strikes.

Table 14: Frequency of motor vehicle collisions with multiple patient transports in the United States (2005-2007) (National Highway Traffic Safety Administration, 2008).

UMCI Size	Number of incidents	Number of patients
Very Small ($2 \leq m \leq 10$)	552,387	1,399,981
Small ($11 \leq m \leq 25$)	1,272	17,846

Table 15: Frequency of disasters in the United States (2005-2007) (Guha-Sapir et al., 2017).

UMCI Size	Number of incidents	Number of patients
Very Small ($2 \leq m \leq 10$)	2	16
Small ($11 \leq m \leq 25$)	6	90
Medium ($26 \leq m \leq 50$)	2	71
Large ($51 \leq m \leq 100$)	4	324
Extra Large ($101 \leq m$)	6	1,402

8 Limitations and Future Work

There are some limitations to our model. Our data was not granular enough to obtain estimates for time intervals smaller than one day. Our model also was not designed to incorporate triage or mortality: we consider only incidents where the number of casualties is not so large that all hospitals in the area will be overwhelmed, but integrating triage and hospital selection is a clear avenue for future research. On the other hand, our model is tailored to the type of data that HCCs can obtain from their members. HCCs that have data about hospital census can use the model to study their healthcare infrastructure vulnerability and to conduct scenario-based stress tests, examine the decision of hospital inclusion, and understand how predictable variability affects incident response. In particular, understanding how predictable variability in hospital workloads impacts community capacity is deserving of further study.

As we discussed in Section 3.3, our model focuses on the role an HCC can play in coordinating multiple autonomous entities, as opposed to the operations of EMS. Whereas most models of EMS operations consider a response-time objective, we consider an objective that maximizes the probability a patient will be able to access the right kind of hospital resources, which is specific to the UMCI setting.

We did not consider the effect that our patient distribution policies would have on patients already in the hospital. However, prior research has shown that hospital overcrowding (i.e., waiting patients) has an effect on service of existing patients, both in the ED (Batt and Terwiesch, 2016) and in the ICU (KC and Terwiesch, 2012). Because coordination improves usage of existing beds, it could reduce hospital congestion and consequently reduce the negative externalities imposed by the UMCI on existing hospital patients.

In our discussions with the informants who participated in our study, one of the key challenges in incident response is patient tracking. Our work highlights the importance of improving patient tracking, since using information about downstream resources requires understanding of patient flow pathways. Because detailed information about patient flow pathways in the studied metropolitan areas was not available, we used a national survey to estimate the first downstream resource used. Ideally, our model would be expanded to consider subsequent downstream resources, and to tailor patient flow pathways to a specific metropolitan area using granular patient-level data. Although challenging in practice, collecting data about PFPs is a potential future role of HCCs that should be explored.

9 Conclusion

In contrast to the original impetus for HCCs, which was the desire on the part of communities to prepare for large disasters, we identify coordination of autonomous healthcare entities in smaller, more frequent incidents as a clear value proposition. We demonstrate these results using actual bed availability data from one urban and one suburban area. To our knowledge, our model is the first to incorporate downstream care resources (such as the intensive care unit) in decision-making about casualty distribution, which is also a sharp departure from the common EMS practice of considering only ED availability when making transportation decisions. We show that this distinction is very important: in our result, sharing ED information provided disappointingly little benefit compared to sharing downstream resource information.

The US healthcare system is characterized by decentralized control of healthcare infrastructure, with some hospitals having a profit-seeking motive. In the absence of an HCC, there is little incentive for hospitals to engage in coordination with their competitors. In this sense, the US system is a “worst-case” scenario for coordination. Coordination would likely be easier to achieve in a system where all hospitals are operated by the same entity (e.g., in the UK). In the US, HCCs make coordination possible because when the HCC provides MAC functionality, no hospital would be better off by leaving the coalition as they would be less involved in UMCI response.

HCCs are growing and diverse organizations, and because of their relative novelty there is not a prototypical funding and operational model. HCCs also rely on hospitals and other community healthcare providers to be willing participants, and they are subject to network effects, particularly if communities desire an HCC to provide MAC functionality. Therefore, a pressing concern is determining an appropriate mix of services for an HCC to provide in any given community in order to be sustainable. Because policy makers are intimately involved in funding HCCs in the US, they should consider providing incentives to autonomous healthcare entities to participate in data collection efforts and in coordinative activities.

Acknowledgments and Disclosures

The authors wish to thank all sources that have provided us with data for this work as well as the Senior Editor and anonymous referees, whose feedback on earlier versions of this manuscript was invaluable. The 2012 NTDB remains the full and exclusive copyrighted property of the American College of Surgeons. The American College of Surgeons is not responsible for any claims arising from works based on the original Data, Text, Tables, or Figures.

References

Allon, G., S. Deo, W. Lin. 2013. The impact of size and occupancy of hospital on the extent of ambulance diversion: Theory and evidence. *Operations Research* **61**(3) 544–562.

Almehdawe, E., B. Jewkes, Q.-M. He. 2013. A markovian queueing model for ambulance offload delays. *European Journal of Operational Research* **226**(3) 602–614.

Almehdawe, E., B. Jewkes, Q.-M. He. 2016. Analysis and optimization of an ambulance offload delay and allocation problem. *Omega* **65** 148–158.

Barbera, J. A., A. G. Macintyre. 2007. *Medical surge capacity and capability: A management system for integrating medical and health resources during large-scale emergencies*. US Department of Health and Human Services.

Bard, J. F., H. W. Purnomo. 2005. Short-term nurse scheduling in response to daily fluctuations in supply and demand. *Health Care Management Science* **8**(4) 315–324.

Batt, R. J., C. Terwiesch. 2016. Early task initiation and other load-adaptive mechanisms in the emergency department. *Management Science* Articles in Advance.

Belien, J., E. Demeulemeester. 2007. Building cyclic master surgery schedules with leveled resulting bed occupancy. *European Journal of Operational Research* **176**(2) 1185.

Berk, E., K. Moinzadeh. 1998. The impact of discharge decisions on health care quality. *Management Science* **44**(3) 400–415.

Centers for Disease Control and Prevention. 1997. Recommended framework for presenting injury mortality data. MMWR 46 (RR-14).

Chow, V. S., M. L. Puterman, N. Salehirad, W. Huang, D. Atkins. 2011. Reducing surgical ward congestion through improved surgical scheduling and uncapacitated simulation. *Production and Operations Management* **20**(3) 418–430.

Council of European Union. 1995. Directive 95/46/EC of the european parliament and of the council.

Courtney, B., E. Toner, R. Waldhorn, C. Franco, K. Rambhia, A. Norwood, T. V. Inglesby, T. O'Toole. 2009. Healthcare coalitions: the new foundation for national healthcare preparedness and response for catastrophic health emergencies. *Biosecurity and bioterrorism: biodefense strategy, practice, and science* **7**(2) 153–163.

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

Deo, S., I. Gurvich. 2011. Centralized vs. decentralized ambulance diversion: A network perspective. *Management Science* **57**(7) 1300–1319.

Department of Health and Human Services. 2002. Administrative data standards and related requirements. Tech. Rep. 45 CFR 160-164.

Gallivan, S., M. Utley. 2005. Modelling admissions booking of elective in-patients into a treatment centre. *IMA Journal of Management Math* **16**(3) 305–315.

Green, L. V. 2002. How many hospital beds? *INQUIRY: The Journal of Health Care Organization, Provision, and Financing* **39**(4) 400–412.

Green, L. V., S. Savin, N. Savva. 2013. Nursevendor problem: Personnel staffing in the presence of endogenous absenteeism. *Management Science* **59**(10) 2237–2256.

Guha-Sapir, D., R. Below, P. Hoyois. 2017. Em-dat: International disaster database. http://www.emdat.be/. Accessed 01 March 2017.

Harper, P., A. Shahani. 2002. Modelling for the planning and management of bed capacities in hospitals. *Journal of the Operational Research Society* **53**(1) 11–18.

Harrison, G., A. Shafer, M. Mackay. 2005. Modelling variability in hospital bed occupancy. *Health Care Management Science* **8**(4) 325–334.

Helm, J. E., M. P. Van Oyen. 2014. Design and optimization methods for elective hospital admissions. *Operations Research* **62**(6) 1265–1282.

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

Joglekar, N. R., J. Davies, E. G. Anderson. 2016. The role of industry studies and public policies in production and operations management. *Production and Operations Management* **25**(12) 1977–2001.

Kazemian, P., Y. Dong, T. R. Rohleder, J. E. Helm, M. P. Van Oyen. 2014. An IP-based healthcare provider shift design approach to minimize patient handoffs. *Health Care Management Science* **17**(1) 1–14.

KC, D. S., C. Terwiesch. 2012. An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing & Service Operations Management* **14**(1) 50–65.

Kim, S.-H., C. W. Chan, M. Olivares, G. Escobar. 2014. ICU admission control: An empirical study of capacity allocation and its implication for patient outcomes. *Management Science* **61**(1) 19–38.

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

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

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

National Highway Traffic Safety Administration. 2008. National motor vehicle crash causation survey: Report to congress. Tech. Rep. DOT HS 811 059.

Schmitt, K. 2016. How many healthcare coalitions is too many? http://bparati.com/Full-Article/ArticleID/945/How-Many-Healthcare-Coalitions-Is-Too-Many. Accessed 27 June 2016.

Schoenbaum, S. C., A.-M. J. Audet, K. Davis. 2003. Obtaining greater value from health care: the roles of the us government. *Health Affairs* **22**(6) 183–190.

Shi, P., M. C. Chou, J. Dai, D. Ding, J. Sim. 2015. Models and insights for hospital inpatient operations: Time-dependent ED boarding time. *Management Science* **62**(1) 1–28.

Figure 8: Temporal sequence of UMCI response activities, participants, and roles.