Mass-Casualty Distribution for Emergency Healthcare: A Simulation Analysis

Mohsin Nasir Jat¹ · Raza Ali Rafique²

Published online: 16 March 2020
© The Author(s) 2020

Abstract This study focuses on the casualty-load distribution problem that arises when a mass casualty incident (MCI) necessitates the engagement of multiple medical facilities. Employing discrete event simulations, the study analyzed different MCI response regimes in Lahore, Pakistan, that vary in terms of the level of casualty-load distribution and the required coordination between the incident site and the responding hospitals. Past terrorist attacks in this major metropolitan area were considered to set up experiments for comparing delays in treatment under the modeled regimes. The analysis highlights that the number of casualties that are allowed to queue up at the nearest hospital before diverting the casualty traffic to an alternate hospital can be an important factor in reducing the overall treatment delays. Prematurely diverting the casualty traffic from the incident site to an alternate hospital can increase the travel time, while a delay in diversion can overload the nearest hospital, which can lead to overall longer waiting times in the queue. The casualty distribution mechanisms based only on the responding hospitals’ available capacity and current load can perform inefficiently because they overlook the trade-off between the times casualties spend in traveling and in queues.

Keywords Casualty distribution · Emergency response · Mass casualty incidents · Pakistan · Urban terrorism

1 Introduction

Changes in geopolitical dynamics have brought conflicts and modern warfare into civilian arenas. As a result, a high number of urban terrorism incidents have occurred during the past decade. Many of these incidents can be classified as large-scale disasters based on the number of casualties. Terrorism risk is increasingly becoming a major concern for today’s societies (Sawalha 2017). A significant portion of the world population now lives with the threat of terrorism, while the brunt of terrorism has mainly been faced by low- and middle-income countries, such as Pakistan. The top 10 countries in terms of the number of terrorist attacks in 2018 were all low- and middle-income countries. These countries, including Pakistan, faced more than 80% of the total worldwide terrorist attacks in 2018 (START 2018).

Terrorist attacks, besides industrial and transportation accidents, can instantaneously generate a large number of casualties at a concentrated location. To minimize mortality, the response to such mass casualty incidents (MCI)—reaching the incident location, performing triage, and evacuating the casualties to medical facilities—needs to be performed in a rapid and smooth manner. However, MCIs pose several challenges to emergency and healthcare services. Mass casualty incidents are unpredictable in terms of their time and location of occurrence. After an MCI occurs, the aim is to identify critically injured casualties (triage) and rapidly evacuate the casualties, which are conflicting priorities. The greater the number of casualties, the more time it may take to find those who need immediate care, and the greater the likelihood of preventable deaths that are caused due to treatment delays. But the focus on rapid scene clearance to reduce treatment delays can result in triage errors, which can increase...
The purpose of the study was not to prescribe a situation-specific tool for determining mass-casualty traffic control but to generate general insights that can assist in devising flexible casualty distribution plans and guidance in the wake of an MCI. Section 2 provides a brief literature review, and Sect. 3 outlines the research methodology. Section 4 presents the analyzed approaches for casualty distribution and the experimental settings. The simulation experiment results are presented in Sect. 5, the results and research limitations are discussed in Sect. 6, and the conclusion is stated in Sect. 7.

2 Literature Review

The research related to the casualty distribution and coordination in response to an MCI can be classified into two streams. The first stream consists of commentary articles that discuss past incidents and shortcomings in their response while highlighting opportunities for better responses, mainly through better coordination and casualty distribution. A significant volume of this research is found in medical and emergency-related journals such as *Trauma and Injury*. The second stream focuses on mathematical modeling, presenting axiomatic prescriptive models that use operations research techniques. Comparatively fewer in number, the articles in the second stream are mostly published in operational research and operations management journals. We briefly cover both streams in this section.

A fundamental issue in responding to an MCI is the constraint on the capacities of medical facilities. Mass casualty incidents can lead to an imbalance between available medical resources, which can lead to chaos and compromise with respect to the quality of care. Several research articles and MCI reports have commented on this phenomenon being the result of the casualty surge that is created by the incident scene clearance (Hirshberg et al. 2001; Halpern et al. 2003; Gutierrez de Ceballos et al. 2005; Aylwin et al. 2006; Fuentes and Diaz 2007; Lennquist 2012; VanVactor 2012; Gates et al. 2014; Haverkort et al. 2017). The problem can amplify, if the distribution of casualties between the medical facilities is not properly organized (Mulyasari et al. 2013). There can be significant load disparities among the casualty receiving medical facilities involved in the MCI response (Hirshberg et al. 1999; van Vugt 2001; Frykberg 2002, 2004; Mulyasari et al. 2013). We contribute to the literature by quantitatively analyzing this apparently positive role rather than providing a retrospective discussion. The analysis provides some interesting insights. For example, a simple predefined casualty distribution mechanism, that requires no real-time coordination between the incident site and the hospitals, can bring significant improvements in the waiting times when compared to the original practice that directs all casualties to the nearest hospital first. The results also show that in certain cases it can be better to direct all casualties to the nearest hospital (the original practice) even if the casualties have to wait at the hospital. Furthermore, the results highlight the potential of a threshold policy in which the nearest hospital only receives casualties when the number of casualties it holds is less than a certain level; beyond this threshold level, the casualties are diverted to an alternate hospital. Overall, the results suggest that the efficiency of a distribution regime depends on a number of factors.

Rapid scene clearance further creates a surge in the rate that casualties arrive at the receiving medical facilities and puts the available resources under stress. The problem can amplify, if the distribution of casualties within the medical facilities is not properly organized—the issue that is the focus in this study. The strain at medical facilities and the mortality rate can presumably be reduced through better utilization of the available capacity of the existing medical facilities through an efficient distribution of the MCI casualties. Another approach to mitigating the overwhelming of medical facilities in response to an MCI is to set aside medical capacity in the form of dedicated MCI hospitals. However, this would be an investment in a capacity that is routinely unutilized.

Considering the cases of terrorist attacks between 2012 and 2017 in Lahore, Pakistan, this study focused on situations in which multiple hospitals are engaged in an MCI response, which raises the problem of efficient management of the casualty traffic from the incident site to the hospitals. Different approaches to transfer casualties from the incident site to the hospitals were analyzed. These approaches, which vary in terms of the level of casualty-load distribution and the required coordination, were simulated to determine and compare the expected waiting times for a casualty before treatment at a hospital’s emergency department (ED). Several studies have commented on the positive role of coordination and casualty-load distribution when responding to an MCI (Hirshberg et al. 1999; Simon and Teperman 2001; van Vugt 2001; Frykberg 2002, 2004; Mulyasari et al. 2013). We contribute to the literature by quantitatively analyzing this apparently positive role rather than providing a retrospective discussion. The analysis provides some interesting insights. For example, a simple predefined casualty distribution mechanism, that requires no real-time coordination between the incident site and the hospitals, can bring significant improvements in the waiting times when compared to the original practice that directs all casualties to the nearest hospital first. The results also show that in certain cases it can be better to direct all casualties to the nearest hospital (the original practice) even if the casualties have to wait at the hospital. Furthermore, the results highlight the potential of a threshold policy in which the nearest hospital only receives casualties when the number of casualties it holds is less than a certain level; beyond this threshold level, the casualties are diverted to an alternate hospital. Overall, the results suggest that the efficiency of a distribution regime depends on a number of factors.
rarely employed option (Haverkort et al. 2017). For most urban centers the issue remains to reduce the strain on the system, and the mortality rate, through better utilization of the routine emergency medical facilities through the efficient distribution of the MCI casualties. However, the efficient distribution of MCI casualties has been an issue even in developed countries with advanced emergency services (Klein and Weigelt 1991; Frykberg 2002; Postma et al. 2013). Equal distribution of patients across regional hospitals is often unsuccessful because of the dysfunctional use of distribution plans or lack thereof (Haverkort et al. 2017). VanVactor (2012) suggested that it is important to have foundational logistics planning for the effective distribution and reallocation of medical resources once mass-casualty events begin.

Generally, there has been significant research on the efficient management of disaster and emergency response employing mathematical modeling and operations research techniques (for literature reviews in this area, see Simpson and Hancock 2009, and Caunhye et al. 2012). However, there are few studies that specifically focus on managing mass-casualty traffic from the incident site to the medical facilities. Recently, Jacobson et al. (2012), Dean and Nair (2014), Mills (2016), and Repouissis et al. (2016) have presented optimization models related to this specific issue.

Jacobson et al. (2012) formulated a priority assignment model using stochastic dynamic programming to ration the emergency healthcare resources (for example, ambulances and operating rooms). The model determines each patient's priority level by considering the number of patients as well as their injury characteristics. Mills (2016), stressing the importance of simple decision making in an MCI scenario, proposed a heuristic-based approach for patient prioritization to determine the order in which to transport the casualties. Considering the available resources and casualty survival probabilities, Dean and Nair (2014) formulated an optimization model to determine the optimal distribution of MCI casualties between multiple hospitals. Repouissis et al. (2016) proposed a mixed integer program through which they aimed to optimize ambulance dispatching and treatment order, besides patient-to-hospital assignment. These studies assume patient assessment and classification at the incident site. It can, however, be argued that the assessment and classifications of casualties using a mathematical model at the incident site may not be realistic. Dean and Nair (2014) and Mills (2016) recognized that solving a mathematical program in real time can be a barrier to implementation in an urgent situation. Aylwin et al. (2006) even suggested that simplifying casualty classification at the incident site and systematic casualty reassessment at different stages can be a better approach. To our knowledge, minimal to no formal casualty classification takes place at incident sites in Pakistan.

In this study, we essentially consider an MCI response as a dynamic casualty routing problem, focusing specifically on the transfer of casualties between the incident site and hospitals. In its nature, this study is similar to Fawcett and Oliveira (2000), which simulated a broad casualty movement model, avoiding details like casualty triage classifications, with the aim to generate insights for planning purposes. By providing a tractable quantitative analysis, compared to the existing literature on the topic, this study goes beyond retrospective commentaries on the possible impact of coordination and casualty-load distribution in response to an MCI, while generating insights that can have practical implications.

3 Methodology

This study relies on discrete event simulations (DES) to model a terrorist attack and investigate the impact of various casualty distribution regimes on casualties’ waiting time in the system (Fig. 1). Discrete event simulation is one of the most popular quantitative modeling methodologies that have been used for research in various fields, including disaster and healthcare management (for the evolution and applications of DES, see Nance and Sargent 2002, and Hollocks 2006). In management science research, DES provides a much more realistic and flexible way to model operations compared to the conventional mathematical modeling methods. The lack of accuracy in the output has been flagged as a major drawback of using simulations; however, the proponents of the methodology argue that a valid model with less accurate results is better than an invalid model with accurate results. The advancements in simulation technology and computing power have now provided substantial confidence in simulation results.

![Fig. 1 Research framework of the simulation analysis of casualty distribution in a mass casualty incident. ED emergency department, DES discrete event simulation](image-url)
times were determined using Stat::Fit, a software program for statistically fitting analytical distributions to data; Stat::Fit can generate expressions for directly exporting the fitted distributions as modeling inputs into various simulation programs including Simio, the program used to implement and analyze the simulation models in this study. Simio is a versatile and domain neutral simulation environment based on a graphical object-oriented framework. It supports multiple modeling paradigms including event-, process-, object-, and agent-based paradigms. The object orientation allows a comparatively more natural way of modeling a system by describing objects (for example, patients and hospitals) that make up the system. The system behavior emerges from the interaction of these objects. Pegden (2008) provided detailed information on Simio’s simulation modeling framework and features.

To verify that the program components perform as anticipated, information stored in the variables and the data members of interest are observed for a process walkthrough. The simulation results are verified through Jackson’s network method that provides estimates and output bounds through analytical queueing models. For variance reduction, Simio uses the common random numbers technique to compare the alternative system configurations, that is, the different casualty distribution regimes. This means that the same random number sequence is generated for each configuration, ensuring that the state of the system, that is, the state of the hospitals, is exactly the same at the time of the MCI in each casualty distribution configuration. That means, there is more confidence that the observed differences in the performance are due to the differences in the system configuration rather than the fluctuations in the experimental conditions.

4 System Description

This section presents the system configurations modeled to analyze different mass-casualty distribution approaches, the performance measures used to compare these approaches, and the system parameter values set for the simulation experiments.

4.1 Configurations

As the base case, we modeled the MCI response regime practiced by the emergency services in Lahore, the second largest city of Pakistan, with a population of over 11 million (2017). In this regime, after an MCI, the major local public hospitals are alerted and the ambulances and personnel from the district emergency service and non-governmental organizations (NGOs) reach the incident site. The non-serious casualties are treated on the spot, while serious casualties are transferred to the nearest hospital, that is, the primary response hospital. Once the capacity of the primary hospital is exhausted, the incoming casualty traffic is diverted from there to the alternate hospital, which is the major hospital nearest to the primary hospital. In case the capacity of the alternate hospital is also exhausted, the traffic is diverted to the second alternate hospital. We label this practice as “original” (Fig. 2) and compare it with three alternate regimes (Figs. 3, 4, 5) considering the waiting time performances defined in Sect. 4.2.

Alternate 1 (Fig. 3): Set proportions of the MCI casualties are directed to each hospital in the system. These proportions are set in accordance with the normally available capacity of the hospitals’ EDs. The proportion of casualties directed towards a hospital equals the ratio between its average unutilized capacity and the total average unutilized capacity of all responding hospitals. For example, two hospitals, Hospital A and Hospital B, are engaged in responding to an MCI. On average, Hospital A has 15 beds unoccupied, while Hospital B has five beds unoccupied. In this case, 75% of the MCI casualties will be directed towards Hospital A, while 25% will be directed towards Hospital B.

Alternate 2 (Fig. 4): The site commander is in continuous coordination with the involved hospitals. An ambulance from the site is directed to the hospital with the lowest load, that is, the hospital with the largest difference between the capacity and the number of casualties present and on the way.

Alternate 3 (Fig. 5): Like Alternate 2, the incident commander coordinates the evacuation operation with the hospitals. Casualties are directed from the incident site towards the primary hospital until some capacity threshold level is reached, triggering the diversion of the traffic to the

---

1 http://www.geerms.com.
2 http://www.simio.com.
alternate hospital from the incident site. These threshold levels, in relation to the total capacity level, can preemptively stop casualty flow to the primary hospital to avoid its capacity exhaustion.

The original practice requires no prior evaluation of the hospitals’ capacity and no coordination between the incident site and the facilities during the response (Table 1). Alternate 1 requires a prior evaluation of the hospitals’ capacity but no coordination between the incident site and the facilities during the response. Alternates 2 and 3 require a high level of coordination between the incident site and the hospitals during the response.

We assume that each hospital responding to the MCI has a 24/7 emergency department, with a fixed capacity, routinely receiving patients at a certain rate. Though there are different wards in the EDs, the beds are moveable within the wards; hence, it is assumed that the capacity of the EDs is equal to the total number of beds. The MCI casualties are evacuated from the incident site and arrive at the hospitals in addition to the routine emergency patient arrivals. There are sufficient ambulances from the rescue service and NGOs, such that there is no constraint on the number of ambulances for evacuation to the hospitals. To manage the casualty influx in an ED, off-duty doctors and the doctors from the other hospital departments are called in, and, as such, the care is normally constrained by the available bed space.

Fig. 2 Original (practiced) response to a mass casualty incident (MCI) in Lahore, Pakistan

Fig. 3 Fixed distribution response (Alternate 1) to a mass casualty incident (MCI) in Lahore, Pakistan

Fig. 4 Load-based distribution response (Alternate 2) to a mass casualty incident (MCI) in Lahore, Pakistan
4.2 Performance Measures

We compare the configurations based on (1) the expected average waiting time; and (2) the expected maximum waiting time of a casualty before being treated at an ED. For convenience, we label expected average waiting time as “average waiting” and expected maximum waiting time as “maximum waiting.” Both measures are formally defined below.

Under a particular configuration, let

\[ n = \text{the total number of casualties in a simulated MCI}; \]
\[ m = \text{the number of hospitals responding to the MCI}; \]
\[ r = \text{the total number of times the MCI is simulated (simulation runs)}; \]
\[ t_{ij} = \text{the total time spent by casualty } i \text{ in the system, that is, time between the beginning of the evacuation operation at the incident site and casualty } i \text{'s exit from the system after a treatment in an ED, in run } j. \]

Where \( i = 1 \ldots n \) and \( j = 1 \ldots r; \)
\[ t_{\text{max}_j} = \text{the maximum of } t_{ij} \text{ in simulation run } j. \]

Where \( j = 1 \ldots r; \)
\[ P_k = \text{the proportion of casualties treated at hospital } k. \]

Where \( k = 1 \ldots m; \)
\[ T_k = \text{the average treatment time of MCI casualties at hospital } k. \]

Where \( k = 1 \ldots m. \)

Then,

\[
\text{Average waiting} = \frac{\sum_{j=1}^{r} \left( \sum_{i=1}^{n} \left( t_{ij} / n \right) \right)}{r} - \sum_{k=1}^{m} P_k T_k
\]

and,

Table 1 Response regime characteristics/requirements in a mass casualty incident

| Regime      | Prior capacity evaluation | Coordination between the incident site and hospitals | Casualty distribution from the incident site |
|-------------|---------------------------|------------------------------------------------------|---------------------------------------------|
| Original    | –                         | –                                                    | –                                           |
| Alternate 1 | ✓                         | –                                                    | High\(^a\)                                  |
| Alternate 2 | –                         | ✓                                                    | High\(^a\)                                  |
| Alternate 3 | –                         | ✓                                                    | High/low (controlled\(^b\))                |

\(^a\)Assuming that the responding hospitals are equivalent in their capacity and routing load

\(^b\)Depends on the chosen threshold level. If the threshold level is low, the diversion of the casualty traffic to the alternate hospital will be triggered more frequently, leading to a high level of casualty distribution. With a higher threshold level, the engagement of the alternate hospital will be delayed and the casualty distribution will be low.
Maximum waiting = \( \sum_{j=1}^{r} t_{\text{max}_j} - \sum_{k=1}^{m} P_k T_k \),
(2)
where \( \sum_{j=1}^{r} t_{\text{max}_j} \) is the average of the average casualty times in the system in \( r \) simulation runs, that is, the expected average time in the system, \( \sum_{k=1}^{m} P_k T_k \) is the average of the maximum casualty times in the system in \( r \) simulation runs, that is, the expected maximum time a casualty can spend in the system, and \( P_k T_k \) is the expected treatment time of a casualty in the EDs. The expected average time in the system and the expected maximum time a casualty can spend in the system are default output statistics provided by Simio. The difference between the expected average time in the system and the expected treatment time is taken as average waiting (1), and the difference between the expected maximum time in the system and the expected treatment time is taken as maximum waiting (2). The simulation output on the expected times in the system and the proportions of casualties treated at each hospital are imported from Simio into MS Excel to compute and analyze Eqs. 1 and 2.

4.3 Experimental Settings

The experimental settings were built based on the terrorist bomb attacks that occurred in Lahore between 2012 and 2017. As the capital of Punjab, Pakistan’s most populated province, with a population of more than 110 million (2017), Lahore is the location of several important government installations that have been the targets of terrorist attacks. Mayo Hospital (labelled H1) is the largest hospital in the city with an overall capacity of 2399 beds. The surrounding area includes major markets and governmental buildings and has experienced several terrorist bomb attacks in the past years. In the simulation we assume that a bomb attack occurs in H1’s vicinity and that the distance between the simulated attack and H1 is equal to the average distance between the hospital and the bomb attack locations between 2012 and 2017 in its vicinity (Fig. 6). Jinnah Hospital (labelled H2) and General Hospital (labelled H3) are considered as the second and third closest major hospitals, respectively, to the simulated attack. The average distance between the recent attack locations and H2 is taken as the distance between the simulated attack and H2. Similarly, the average distance between the recent attack locations and H3 is taken as the distance between the simulated attack and H3. Precisely, if \( D_{ij} \) is the actual distance between hospital \( i \) and a recent attack location \( j \), where \( j = 1 \ldots n \), then the distance between hospital \( i \) and the simulated attack is taken as \((\Sigma D_{ij})/n\). Note that for the threshold policy (Alternate 3), we only consider two hospitals in the system (H1 and H2). It is straightforward to model three or more hospitals with the threshold policy but it will add more variables to control in the experiments, significantly increasing the number of possible scenarios while not impacting the main qualitative insights.

Table 2 lists the parameters and distributions used as the model inputs. The input distributions are based on the Anderson–Darling and Kolmogorov–Smirnov tests on the observed treatment times and patient arrival rates using Stat::fit software. For the routine patients we consider general treatment time observations, whereas, for the MCI casualties, we only consider the observations from surgery and medicine wards, which are more relevant to blast injuries based on the information from the hospitals.

Before entering the wards, the regular emergency patients undergo a quick check-in process in which their basic information is fed into the computer database. At H1, we found that the hold time for the check-in has an exponential distribution with a mean of 0.326 min. Because the check-in system is identical at all three hospitals, the same check-in time distribution is considered for the other hospitals. The casualties from an MCI do not go through a check-in process.

The results in Sect. 5 correspond to two sets of experiments. In the first set, labelled as the “Low influx” case, it is assumed that the MCI generates 150 casualties. In the second set of experiments, the “High influx” case, it is assumed that the number of MCI casualties is 200. The capacity at H1 is suppressed by 20% to represent an overloaded ED at the nearest hospital. Testing cases with a lower casualty-load than the Low influx case and a higher casualty-load than the High influx case provided insights that are not significantly different from the ones discussed below. Alternate 3, the threshold policy, is only explored with the High influx case, in which the casualty distribution issue is more prominent.

The results reported and analyzed are gathered from 34 experiments, with 1000 simulation runs each. Simulations are performed using Simio version 10—Academic edition. In all experiments, the hospital systems are simulated for 2 weeks before triggering the MCI so that the bed usage at the hospitals comes to a stable level when the MCI casualties start arriving. Triggering the MCI too early can mean that a higher than normal bed capacity will be available to the MCI casualties due to the transient stage of the system.

5 Results and Analysis

The simulation results under the low influx and the high influx cases show a contrast. When the casualty influx is low, the distribution of casualties among multiple hospitals may have an undesired impact on the casualty waiting
**Fig. 6** Hospital and bomb attack locations between 2012 and 2017 in Lahore, Pakistan. H1: Mayo Hospital; H2: Jinnah Hospital; H3: General Hospital. Source: National Consortium for the Study of Terrorism and Responses to Terrorism (START 2018)

**Table 2** Emergency department treatment time distributions at Mayo (H1), Jinnah (H2), and General (H3) hospitals in Lahore, Pakistan (ED: Emergency department; MCI: Mass casualty incident)

| Input                                      | Distributions                                                                 |
|--------------------------------------------|-------------------------------------------------------------------------------|
| Routine treatment times for patients       | Exponential distribution with a mean of 23.44 min and a displacement of 4 min:|
| occupying beds in ED                      | 4 + Exponential (23.443)                                                      |
| Treatment times for blast casualties      | Exponential distribution with a mean of 19.8 min and a displacement of 17 min:|
|                                           | 17 + Exponential (19.8)                                                       |
| Exponential distribution with a mean of    | 18.172 min and a displacement of 5 min: 5 + Exponential (18.172)              |
| 19.8 min and a displacement of 17 min:     | Exponential distribution with a mean of 31.028 min and a displacement of 12 min:|
|                                           | 12 + Exponential (31.028)                                                    |
| Lognormal distribution with a normal mean  | 2.79 min, normal standard deviation of 0.691, and a displacement of 11 min:  |
| of 2.79 min, normal standard deviation of  | 11 + Lognormal (2.79, 0.691)                                                 |
| 0.691, and a displacement of 11 min:       | Exponential distribution with a mean of 31.028 min and a displacement of 12 min:|
|                                           | 12 + Exponential (31.028)                                                    |
| ED bed capacity                            | 139                                                                           |
| Routine arrival rate of patients in ED     | 141.18                                                                        |
| (per hour)                                 | 118.11                                                                        |
| Travel time from the simulated MCI site    | 16 min                                                                        |
|                                           | 30 min                                                                        |
|                                           | 32 min                                                                        |

*Arrival rates are based on the information from the statistics departments. The arrival rate observations confirmed that the rates follow Poisson distribution.*
times. The proposed casualty distribution alternates essentially demonstrate a potential when the casualty influx is high.

5.1 Low Influx Case

In the Low influx case, the original scenario outperforms both Alternates 1 and 2, with Alternate 2 being considerably better than Alternate 1 in terms of both average and maximum waiting times (Fig. 7). Note that H1, the nearest hospital, has a large spare capacity considering the regular emergency flow, due to which few MCI casualties require diversion to the other hospitals after they arrive at H1 in the original scenario.

In the experiments, only around 2.4% of the casualties are diverted to the other hospitals as a result of the exhaustion of H1’s capacity in the original scenario (Table 3). In Alternate 1, around 45.5%, 26.9%, and 27.6% of the casualties are routed to H1, H2, and H3, respectively. In Alternate 2, these percentages are 67.1%, 22.02%, and 10.9%, respectively. Compared to Alternate 2, Alternate 1 routes fewer casualties to H1 and more casualties to H3, the furthest hospital in the system. Compared to Alternate 1, the real-time load monitoring in the more efficient Alternate 2 finds H1 less loaded in most instances, sending more casualties on the shorter route towards H1.

Overall, the low casualty influx and a high capacity at the nearest hospital do not as such raise a casualty distribution issue. It is efficient to direct all casualties from the incident site to the nearest hospital as the hospital’s capacity rarely exhausts, making it unnecessary to route casualties to other hospitals that require longer travel time. If all casualties are treated at H1, that is the other two hospitals are not involved in the MCI response, the average waiting time is 1.523 h and the maximum waiting time is 3.63 h, which is a better result compared to the performance under the three distribution configurations.

5.2 High Influx Case

In the high influx case, the results show a potential improvement through the proposed alternates to the original practice (Fig. 8). Alternates 1 and 2 perform almost similarly in waiting times, with Alternate 1 providing marginally lower average waiting, while Alternate 2 provides marginally lower maximum waiting. Though Alternate 2 intelligently distributes the casualties considering the load at the hospitals, it disregards the distance between the incident site and the hospitals. Under Alternate 2, comparatively fewer casualties are directed towards H1 (Table 4), the closest hospital from the incident site. Overall, Alternate 3 performs the best and significantly better than the original configuration.

Figure 9 presents the percentage improvements in average and maximum waiting through the alternates in comparison with the original setting. A major improvement in the average waiting is achieved through the simple Alternate 1, whereas the more sophisticated Alternate 3 shows a marginal improvement beyond the Alternate 1 performance (Fig. 9a). Compared to the original setting, Alternates 1, 2, and 3 show a reduction of approximately 12.4, 12.1, and 13.7 min, respectively, in average waiting.

Alternate 3 performs significantly better than the other settings with respect to maximum waiting (Fig. 9b). Its results show a reduction of 48.7 min from the original setting’s performance, compared to the reductions of 19.7 and 20.8 min under Alternates 1 and 2, respectively. Note that under the original setting, a large proportion of casualties (around 48%) are diverted to H2 after reaching H1 (Table 4). There are also cases in which a casualty arrives at H3 via H1 and H2.

The threshold levels in Alternate 3 are determined by systematically evaluating various thresholds as the percentage of bed space capacity and selecting the ones providing the lowest average and maximum waiting (Fig. 10). The threshold levels above 100% mean that a queue is
allowed to develop before casualty traffic is diverted to H2 from the incident site. The average waiting forms a fairly clear U-shaped curve against the threshold levels (Fig. 10a)—a comparatively rapid drop, followed by an increase, providing a noticeable base corresponding to good threshold levels. However, the maximum waiting shows unsystematic fluctuations against the threshold levels and no clear trend (Fig. 10b). The analysis suggests that a queue should be allowed to accumulate at the nearest response hospital before the casualties are diverted to the alternate hospital.

The threshold levels that provide the lowest average and lowest maximum waiting times are different; hence, there could be a trade-off between minimizing the expected

---

### Table 3

System performance (Low influx) at Mayo (H1), Jinnah (H2), and General (H3) hospitals in Lahore, Pakistan

| Setting     | % of casualties | Time in system | Avg. waiting | Max. time in system | Avg. max. waiting |
|-------------|-----------------|----------------|--------------|---------------------|-------------------|
|             | H1   | H2    | H3      | Avg.\(^a\) | CI (½ width)\(^b\) | Avg.\(^c\) | CI (½ width)\(^b\) |
| Original    | 97.631 | 0.1887 | 2.1807  | 2.1508  | 0.0075     | 1.535376 | 0.0323     | 3.754576 |
| Alternate 1 | 45.451 | 26.931 | 27.618  | 2.2841  | 0.0079     | 1.666099 | 0.04       | 4.166789 |
| Alternate 2 | 67.092 | 22.023 | 10.885  | 2.21081 | 0.0072     | 1.605169 | 0.0395     | 3.880219 |

*Average of the average times casualties spend in the system over 1000 simulation runs

*95% CI

*Average of the maximum times casualties spent in the system over 1000 simulation runs

---

### Table 4

System performance (high influx) at Mayo (H1), Jinnah (H2), and General (H3) hospitals in Lahore, Pakistan

| Setting     | % of casualties | Time in system | Avg. waiting | Max. time in system | Avg. max. waiting |
|-------------|-----------------|----------------|--------------|---------------------|-------------------|
|             | H1   | H2    | H3      | Avg.\(^a\) | CI (½ width)\(^b\) | Avg.\(^c\) | CI (½ width)\(^b\) |
| Original    | 51.572 | 20.905 | 27.524  | 2.50976 | 0.0074     | 1.885964 | 0.0409     | 4.608404017 |
| Alternate 1 | 39.074 | 36.069 | 24.858  | 2.28725 | 0.0068     | 1.679249 | 0.04       | 4.280008977 |
| Alternate 2 | 37.441 | 37.851 | 24.709  | 2.29083 | 0.0066     | 1.684516 | 0.0411     | 4.260905463 |
| Alternate 3 (110%) | 59.306 | 40.705 | –       | 2.2361 | 0.0061     | 1.657764 | 0.0306     | 3.809063937 |
| Alternate 3 (102%) | 52.466 | 47.534 | –       | 2.23386| 0.0062     | 1.661406 | 0.0283     | 3.796055907 |

*Average of the average times casualties spend in the system over 1000 simulation runs

*95% confidence interval

*Average of the maximum times casualties spent in the system over 1000 simulation runs
average and maximum waiting times. The threshold level that provides the lowest average waiting time requires the development of a comparatively longer queue. It may be difficult to decide whether a reduction in the average waiting time is more important than the reduction in the maximum waiting time. However, the performance difference at the threshold level that minimizes the average waiting time and the threshold level that minimizes the maximum waiting time is small (Table 5).

6 Discussion

Rapid transfer of patients from an MCI site to definitive care is deemed a major factor in reducing critical mortality among casualties. Considering casualty evacuation, distribution, and ED treatment, this study provides original insights with practical implications for emergency logistics management for mass-casualty transfers to hospitals. The study focuses on the logistics link between an MCI site and medical facilities, which is a critical but under-researched area.

The analysis in this study shows that if the nearest/central hospital in the MCI vicinity has a sizeable excess capacity, then it can be efficient to direct all casualty traffic to this hospital, eliminating the casualty distribution problem. Hence it can be beneficial to maintain a surplus capacity at a central hospital to mitigate against overwhelming medical facilities with mass-casualty influx and to simplify MCI response. Having a surplus capacity at a running hospital can be a better alternative to establishing a dedicated MCI hospital in terms of providing economies of scale and lower setup costs. A high casualty influx can though understandably undermine the capacity of a single hospital, necessitating the distribution of casualties among multiple hospitals in the vicinity to achieve lower waiting times.

The simulation results suggest that directing a fixed, predefined proportion of casualties, depending on the capacities of the responding hospitals, may not be inferior to the load-based dynamic routing of the casualties. Determining the fixed proportions can be a one-off task as a disaster preparedness exercise, whereas the load-based dynamic routing requires a high level of coordination between hospitals and the incident site during the response. A more sophisticated threshold policy that defines a capacity threshold at the nearest hospital to trigger casualty diversions to an alternate hospital shows further potential, even with a reduced number of responding hospitals.

The literature suggests that casualties from an MCI should be distributed among multiple hospitals to avoid overloading a single facility. However, the studied threshold policy indicates that loading the nearest facility to some extent before diverting the casualty traffic can be more efficient. That is, a better response to a high-intensity incident can be to distribute casualties only when the local hospital is overloaded beyond a threshold. The policy also exposes a minor trade-off between the average and maximum waiting times for treatment, raising the question of whether reducing the average waiting time is more
important than the maximum waiting time. One may argue that following a threshold policy, in which threshold levels—depending on the casualty rate, number of hospitals, and distances—need to be determined in a critical situation, is not practical. However, studying local settings and predetermining threshold levels for different situations can provide the logistics strategy guidelines for the response to a specific incident.

### 7 Conclusion

The study highlights the need for flexible MCI response plans. The appropriate approach for shifting MCI casualties to hospitals depends on various factors. For example, the usefulness of casualty distribution can depend on the intensity of the incident; the appropriate distribution of casualties among the responding hospitals can depend on the hospitals’ capacities and routine loads; and the location of alternate hospitals may dictate the appropriateness of incurring a time penalty of routing a casualty to an alternate hospital to reduce the load at the nearest hospital. Our preliminary study has shown the potential for the threshold

| Capacity thresholds (%) | % of casualties | Time in system | Avg. a | CI (% width)b | Avg. c | CI (% width)b |
|-------------------------|-----------------|---------------|--------|---------------|--------|---------------|
| 92                      | 42.5705         | 2.25479       | 0.0062 | 1.69086       | 4.433  | 0.0312        |
| 93                      | 43.4205         | 2.2516        | 0.0062 | 1.686925      | 4.4196 | 0.0328        |
| 94                      | 44.3105         | 2.24724       | 0.0061 | 1.681800      | 4.39109| 0.0296        |
| 95                      | 45.377          | 2.24602       | 0.0061 | 1.679748      | 4.40188| 0.0313        |
| 96                      | 46.017          | 2.24666       | 0.0061 | 1.679702      | 4.41628| 0.033          |
| 97                      | 47.29           | 2.24666       | 0.0061 | 1.678657      | 4.4176 | 0.0318        |
| 98                      | 47.9245         | 2.24193       | 0.0063 | 1.673382      | 4.42125| 0.0332        |
| 99                      | 49.2915         | 2.24168       | 0.0063 | 1.671956      | 4.41167| 0.0325        |
| 100                     | 50.1165         | 2.2401        | 0.0061 | 1.669666      | 4.44642| 0.0347        |
| 101                     | 51.5415         | 2.2386        | 0.0061 | 1.666941      | 4.371  | 0.0295        |
| 102                     | 52.466          | 2.23386       | 0.0062 | 1.661406      | 4.36851| 0.0283        |
| 103                     | 53.436          | 2.23451       | 0.0061 | 1.661222      | 4.3919 | 0.0304        |
| 104                     | 54.4425         | 2.2346        | 0.0062 | 1.660446      | 4.40934| 0.0314        |
| 105                     | 55.0925         | 2.23265       | 0.0063 | 1.657937      | 4.39138| 0.0312        |
| 106                     | 56.317          | 2.23377       | 0.0062 | 1.658004      | 4.41969| 0.0326        |
| 107                     | 56.8175         | 2.23533       | 0.0063 | 1.659134      | 4.39767| 0.0294        |
| 108                     | 57.6665         | 2.23767       | 0.0061 | 1.660743      | 4.42778| 0.0309        |
| 109                     | 58.4665         | 2.23536       | 0.0064 | 1.658800      | 4.41594| 0.0317        |
| 110                     | 59.3055         | 2.2361        | 0.0061 | 1.657764      | 4.3874 | 0.0306        |
| 111                     | 60.515          | 2.23726       | 0.0062 | 1.657884      | 4.40511| 0.0313        |
| 112                     | 60.8345         | 2.23972       | 0.0062 | 1.660069      | 4.41663| 0.0317        |
| 113                     | 61.729          | 2.23907       | 0.0064 | 1.658650      | 4.41027| 0.0357        |
| 114                     | 62.32           | 2.23901       | 0.0063 | 1.658081      | 4.42879| 0.0327        |
| 115                     | 62.6325         | 2.23975       | 0.0061 | 1.658553      | 4.44177| 0.0313        |
| 120                     | 65.683          | 2.24725       | 0.0063 | 1.663429      | 4.40585| 0.0326        |
| 125                     | 68.3435         | 2.2555        | 0.0063 | 1.668891      | 4.42834| 0.0303        |
| 130                     | 70.8955         | 2.26342       | 0.0062 | 1.675117      | 4.45667| 0.0321        |

Time in hours

aAverage of the average times casualties spend in the system over 1000 simulation runs

b95% CI

cAverage of the maximum times casualties spent in the system over 1000 simulation runs
policy with some interesting results. A more rigorous study, perhaps employing optimization through simulation, with sensitivity analysis around the numbers and locations of hospitals, incident intensity, and evacuation rates will be more beneficial in providing the guidelines and exposing the relationship between the important factors.

As in any modeling work, the presented simulation study has a number of limitations due to the simplified assumptions. Considering a Poisson distribution for the evacuation rates with a constant mean, and deterministic travel times between the incident site and the hospitals, are two major simplifications in this study. Though it is known that the casualty evacuation rates vary during the response—a factor that can be handled by advanced simulation packages such as the one used—ignoring the time-varying rates have been common (Christie and Levary 1998; Aylwin et al. 2006). The relevant quantitative studies referred in Sect. 2 (Jacobson et al. 2012; Dean and Nair 2014; Mills 2016; Repoussis et al. 2016) also exclude this factor. Identifying the parameters for the change in evacuation rate remains a challenge. Both assumptions undermine the sources of variability and system strains; hence, the performance outputs are lower estimates. We have attempted to capture the system strain by increasing the number of casualties and suppressing the hospital capacity.

Acknowledgements The authors would like to acknowledge the support of Mayo, Jinnah, and General hospitals in Lahore, Pakistan with data gathering and model validation. The authors are also grateful to Mr. Sana Ullah, Research Associate at Lahore University of Management Sciences, for his assistance with this work.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Aylwin, C.J., T. König, N. Brennan, P.J. Shirley, G. Davies, M.S. Walsh, and K. Brohi. 2006. Reduction in critical mortality in urban mass-casualty incidents: Analysis of triage, surge, and resource use after the London bombings on July 7, 2005. The Lancet 368(9554): 2219–2225.

Caunhye, A.M., X. Nie, and S. Pokharel. 2012. Optimization models in emergency logistics: A literature review. Socio-Economic Planning Sciences 46(1): 4–13.

Christie, P.M.J., and R.R. Levary. 1998. The use of simulation in planning the transportation of patients to hospitals following a disaster. Journal of Medical Systems 22(5): 289–300.

Dean, M.D., and 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.

Fawcett, W., and C.S. Oliveira. 2000. Casualty treatment after earthquake disasters: Development of a regional simulation model. Disasters 24(3): 271–287.

Feliciano, D., G. Anderson, G.S. Rozycki, W.L. Ingram, J.P. Ansley, N. Namias, J.P. Salomone, and J.D. Cantwell. 1998. Management of casualties from the bombing at the centennial olympics. The American Journal of Surgery 176(6): 538–543.

Fransoo, J.C., and J. Bertrand. 2002. Operations management research methodologies using quantitative modeling. International Journal of Operations and Production Management 22(2): 241–264.

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

Frykberg, E.R. 2004. Principles of mass casualty management following terrorist disasters. Annals of Surgery 239(3): 319–321.

Fuentes, F.T., and P.D. Diaz. 2007. Medical response to the 2005 terrorist bombings in London. The Lancet 368(9554): 2188–2189.

Gates, J.D., S. Arabian, P. Biddinger, J. Blansfield, P. Burke, S. Chung, J. Fisher, F. Friedman, et al. 2014. The initial response to the Boston Marathon bombing: Lessons learned to prepare for the next disaster. Annals of Surgery 260(6): 960–966.

Gutierrez de Ceballos, J.P., T.F. Fuentes, P.D. Diaz, S.M. Sanchez, M.C. Llorente, and J.E. Guerrero Sanz. 2005. Casualties treated at the closest hospital in the Madrid, March 11, terrorist bombing. Critical Care Medicine 33(1): 107–112.

Halpern, P., M. Tsai, J.L. Arnold, E. Stok, and G. Ersoy. 2003. Mass-casualty, terrorist bombings: Implications for emergency department and hospital emergency response (Part II). Prehospital and Disaster Medicine 18(3): 235–241.

Haverkort, J.J.M., M.B. de Jong, M. Foco, D. Gui, M. Barhoum, G. Hyams, H. Bahouth, M. Halberthal, et al. 2017. Dedicated mass-casualty incident hospitals: An overview. Injury 48(2): 322–326.

Hirschberg, A., J.B. Holcomb, and K.L. Mattix. 2001. Hospital trauma care in multiple-casualty incidents: A critical view. Annals of Emergency Medicine 37(6): 647–652.

Hirschberg, A., M. Stein, and R. Walden. 1999. Surgical resource utilization in urban terrorist bombing: A computer simulation. Journal of Trauma and Acute Care Surgery 47(3): 545–550.

Hollocks, B.W. 2006. Forty years of discrete-event simulation—a personal reflection. Journal of the Operational Research Society 57(12): 1383–1399.

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

Klein, J.S., and J.A. Weigelt. 1991. Disaster management: Lessons learned. The Surgical Clinics of North America 71(2): 257–266.

Lemquist, S. 2012. Medical response to major incidents and disasters: A practical guide for all medical staff. Berlin: Springer.

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

Mulyasari, F., S. Inoue, S. Prashar, K. Isayama, M. Basu, N. Srivastava, and R. Shaw. 2013. Disaster preparedness: Looking through the lens of hospitals in Japan. International Journal of Disaster Risk Science 4(2): 89–100.

Nance, R.E., and R.G. Sargent. 2002. Perspectives on the evolution of simulation. Operations Research 50(1): 161–172.

Obermeyer, Z., S. Abujaber, M. Makar, S. Stoll, S.R. Kayden, L.A. Wallis, and T.A. Reynolds. 2015. Emergency care in 59 low-and
middle-income countries: A systematic review. *Bulletin of the World Health Organization* 93(8): 513–588.

Pegden, C.D. 2008. Introduction to Simio. In *Proceedings of the 40th conference on winter simulation*, 7–10 December 2008, Miami, Florida, USA. 229–235. [https://ieeexplore.ieee.org/document/4736072](https://ieeexplore.ieee.org/document/4736072). Accessed 28 Feb 2020.

Postma, I.L.E., H. Weel, M.J. Heetveld, I. van der Zande, T.S. Bijlsma, F.W. Bloemers, and J.C. Goslings. 2013. Patient distribution in a mass casualty event of an airplane crash. *Injury* 44(11): 1574–1578.

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

Sawalha, I.H. 2017. A context-centred, root cause analysis of contemporary terrorism. *Disaster Prevention and Management* 26(2): 183–202.

Simon, R., and S. Teperman. 2001. The world trade center attack: Lessons for disaster management. *Critical Care* 5(6): 317–319.

Simpson, N.C., and P.G. Hancock. 2009. Fifty years of operational research and emergency response. *Journal of the Operational Research Society* 60(1): 126–139.

START (National Consortium for the Study of Terrorism and Responses to Terrorism). 2018. *Global terrorism database*. Maryland: University of Maryland. [https://www.start.umd.edu/gtd](https://www.start.umd.edu/gtd). Accessed 11 Feb 2020.

VanVactor, J.D. 2012. Strategic health care logistics planning in emergency management. *Disaster Prevention and Management* 21(3): 299–309.

van Vught, A.B. 2001. Disaster medicine: Lessons from Enschede and Volendam. *Nederlands Tijdschrift Voor Geneeskunde* 145(48): 2309–2312.