Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

EMERGENCY IMAGING AFTER A MASS CASUALTY INCIDENT: AN OPERATIONAL PERSPECTIVE VIA A SIMULATION STUDY

Joseph Sismondo
Vahid Sarhangian

Department of Mechanical and Industrial Engineering
University of Toronto
Toronto, ON M5S 3G8, CANADA

Emmett Borg

Business Insights and Analytics
Hoffmann-La Roche
Mississauga, ON L5N 5M8, CANADA

Eric Roberge

Department of Radiology
Madigan Army Medical Center
Tacoma, WA 98431, USA

Ferco H. Berger

Sunnybrook Health Sciences Centre
Department of Medical Imaging
University of Toronto
Toronto, ON M4N 3M5, CANADA

ABSTRACT

In the aftermath of a Mass Casualty Incident (MCI) many patients require lifesaving treatments and surgeries. Due to the sudden surge in demand, the resources of the hospital are overwhelmed, making proper planning and use of the available resources crucial in minimizing mortality and morbidity. To help with planning, patients are triaged into four levels based on the clinical assessment of the criticality of their conditions. The triage decisions are however subject to error and a patient may be under or overtriaged. Mistriages can be identified by performing imaging, e.g., a Computed Tomography (CT) scan, but imaging also takes non-negligible time and has limited capacity. We propose a queueing network model of patient flow during an MCI and use simulation experiments to quantify the value of identifying mistriaged patients. Our results demonstrate the value of performing imaging, but also point out to the importance of accounting for its limited capacity.

1 INTRODUCTION

A Mass Casualty Incident (MCI) is an event that results in a surge of demand for emergency and health services, overwhelms the available resources, and disrupts the normal operations. In the aftermath of an MCI, many patients require immediate lifesaving treatments and surgeries. As such, since treating all patients with a standard level of care in a timely manner is no longer feasible, the care paradigm shifts to minimizing the overall mortality.

After the MCI occurs, patients are often triaged at the scene of the incident and the local hospitals are informed to prepare for the influx of a large number of patients. Some victims may be able to walk to the hospital whereas others need to be transported. Upon arrival to the Emergency Department (ED) patients undergo clinical triage at the ED. Emergency triage protocol assigns incoming patients to four categories: T1 (urgent), T2 (urgent, able to wait), T3 (walking wounded, i.e., patients who can make their own way to the hospital) and T4 (unlikely to survive) (Berger et al. 2016). These triage codes greatly impact prioritization decisions within the hospital. The triage process is subject to error and can lead to

undertriage (falsely assigning a patient to a lower-acuity level) or overtriage (falsely assigning a patient to a higher-acuity level).

Medical imaging tools such as Computerized Tomography (CT) scans and ultrasound scans can provide crucial information and identify mistriaged patients. As indicated in Berger et al. (2016), the role of imaging can be summarized as increasing the accuracy of triage and thereby allowing for better allocation of resources and reducing mortality. Identifying overtriaged patients can free up resources for more critical patients, and identifying undertriaged patients would allow the medical team to provide the appropriate level of care, and hence increase their probability of survival. Despite these benefits, the imaging department resources (equipment, technologists, and radiologists) can also be overwhelmed during an MCI and hence the limited capacity needs to be carefully managed and allocated so as to minimize the mortality.

This work was initiated as part of a collaboration with the American Society of Emergency Radiology (ASER). In the 2018 annual meeting of the society, we organized a simulation workshop to introduce practicing radiologists to simulation modeling and illustrate how it can be used in planning for an MCI by examining various scenarios related to staffing and scheduling decisions, in particular those pertinent to the role of the imaging department. One set of experiments that particularly led to stimulating discussion, was quantifying the benefits of imaging (in terms of reduction in mortality) under various policies indicating how the available capacity should be used. In this paper, we examine this question more closely.

To quantify and study the value of imaging, we propose a stylized queueing network model of the patient flow in part of the ED during a MCI. In particular, we focus on the congestion of the surgery rooms which provide life-saving operations for patients in need. Upon arrival, patients are triaged into one of the 3 possible categories, but the triage decisions are subject to miss-specification errors. (Note that we exclude T4 (expectant) patients from our model.) The triage level is used to determine whether patients need surgery and if so what is their priority level. Patients who are triaged as T2 can be sent to imaging to conduct a CT scan prior to surgery. Upon performing a CT scan the true triage level of patients is identified which can in turn be used to either adjust the priority level of the patient for receiving surgery, or deem surgery unnecessary altogether. Similar to the surgery rooms, the capacity of imaging is however limited. In addition to the regular time required to perform and interpret a CT scan, patients may also need to wait before imaging can be performed. Our queueing model captures this *information-delay* trade-off faced by the clinical decision makers in deciding whether to utilize imaging or not. Indeed, identifying the true triage level is beneficial in reducing mortality, but it can result in an increased length-of-stay for patients who may die without receiving surgery in time.

We calibrate our model using data from the medical literature and in consultation with radiologists. Using simulation experiments, we investigate the following questions: (1) What is the value of information gained through imaging? We quantify the reduction in expected mortality by identifying the mistriaged patients as a function of over and undertriage probabilities and show that it could be significant in parameter regimes relevant to real MCIs. (2) How should the congestion and available capacity of the system be taken into account, when deciding whether a patient should be sent to imaging? We show that in cases where imaging capacity is scarce, one should be cautious with respect to when patients are sent to imaging in order to avoid creating idleness at the surgery rooms. We propose a simple threshold policy - that only uses information on the current number of patients in surgery - which determines whether a patient should be routed to imaging or not. We show that even when capacity of the imaging is scarce, by carefully allocating the available capacity one can still achieve a considerable reduction in the expected mortality compared to the case of performing no imaging.

2 RELATED LITERATURE

Queueing and simulation models have been previously proposed and studied in the Operations Research/Management literature to provide insights into various decision making problems related to MCIs. Here we discuss a few of the most relevant papers to our work.

Cohen et al. (2014) propose a queueing network model to study the dynamic allocation of surgeons to surgery and shock rooms during an MCI. They use a fluid approximation of the system to obtain insights and approximately optimal policies. Fluid models are useful as they capture the transient dynamics of the system well. For our problem derivation of fluid approximations is more complicated due to the existence of mistriaged patients (see also the discussion in Section 5). Yom-Tov and Mandelbaum (2014) study a queueing network with returns and illustrate its applications to various healthcare staffing problems including planning for an MCI. Motivated by the problem of prioritizing patients during an MCI, Argon et al. (2008) studies how jobs with different waiting costs and random remaining lifetimes should be scheduled in a clearing queueing system (i.e., starting with a positive number in queue but receiving no future arrivals). In a related study, Sun et al. (2018) study patient triage and prioritization during a MCI in battlefields. They study a single-server queueing model where the customer types are unknown but can be identified through triage and at the expense of additional delay. They show, in a simplified setting assuming infinite lifetimes, that the server should start performing triage when there are sufficiently many patients and stop when there are few patients. Our queueing model is different in that we assume that all patients are triaged but the triage decisions may be inaccurate. Similar to Sun et al. (2018), we assume that obtaining the additional information (i.e., identifying mistriages) requires additional time. In addition, we consider a finite dedicated capacity for identifying the mistriages. Inaccuracy or uncertainty in patient types has also been considered in Argon and Ziya (2009) who investigate prioritization of patients with imperfect information on their priority levels. For a more detailed discussion of the related literature in this area see Sun et al. (2018).

Hupert et al. (2007) uses simulation to study the relationship between overtriage and mortality. They find that the relationship is not necessarily monotone increasing and depends on other system characteristics such the the relative number of critical patients to available capacity. In contrast, in this work we examine the value of identifying overtriaged patients (in terms of relative reduction in mortality) as the overtriage probability increases.

Finally, we note that other papers in the literature have studied related tradeoffs to the information-delay tradeoff considered in this paper, although in different settings; see for example Levi et al. (2019) and Alizamir et al. (2013).

3 MODEL DESCRIPTION

3.1 The Model

We consider a queueing network with two server pools, one corresponding to the operating rooms and the other to imaging. Patients arrive to the system according to a non-stationary Poisson process with rate $\{\lambda(t); t \geq 0\}$ where we assume that there exists a t_0 such that $\lambda(t) = 0$ for $t \geq t_0$ and define,

$$\Lambda = \frac{1}{t_0} \int_0^{t_0} \lambda(s) ds,$$

to be the average arrival rate during $[0, t_0]$. Upon arrival, patients are triaged and assigned to class $i \in \{1,2,3\}$ (representing triage level Ti) with probability p_i. Patients may be over- or under-triaged. Denote by $\bar{q}_i$ and $\underline{q}_i$ the over- and under-triage probability for class i patients, respectively, that is,

$$\bar{q}_i = \mathbb{P}[\text{Triaged as class } i | \text{ class } (i+1)], \quad i \in \{1,2\},$$
$$\underline{q}_i = \mathbb{P}[\text{Triaged as class } i | \text{ class } (i-1)], \quad i \in \{2,3\}.$$

Class 1 and 2 patients have a random remaining lifetime following a continuous distribution with cdf denoted by $G_i(\cdot)$ and mean $1/\gamma_i$ for class $i \in 1,2$ patients. Class 3 patients have an infinite remaining lifetime. Patients who are triaged as class 3 do not enter the system as they do not require emergency surgery. Patients who are triaged as class 1 directly proceed to surgery upon arrival, whereas patients

triaged as class 2 *can* be sent to imaging node of the networks before the surgery. Imaging fully recovers the true triage class of patients. If a patient is found to be overtriaged he/she exits the system, whereas a patient found to be undertriaged proceeds to the surgery with the true triage type known. Class 1 patients are prioritized for surgery according to a non-preemptive priority policy. Among patients of each class, the queueing discipline is assumed to be First-Come, First-Served (with respect to the arrival time to the surgery node). A schematic representation of the proposed queueing network under the assumption that all class 2 patients are routed to imaging is presented in Figure 1.

There are n_S servers (surgery rooms) at the surgery node and n_I servers in the imaging node of the network. Processing times are assumed to be identically distributed for all patients and follow a continuous distribution with cdf $F_S(\cdot)$ in the operating room and $F_I(\cdot)$ in imaging. The average surgery times and scan times are $1/\mu_S$ and $1/\mu_I$, respectively.

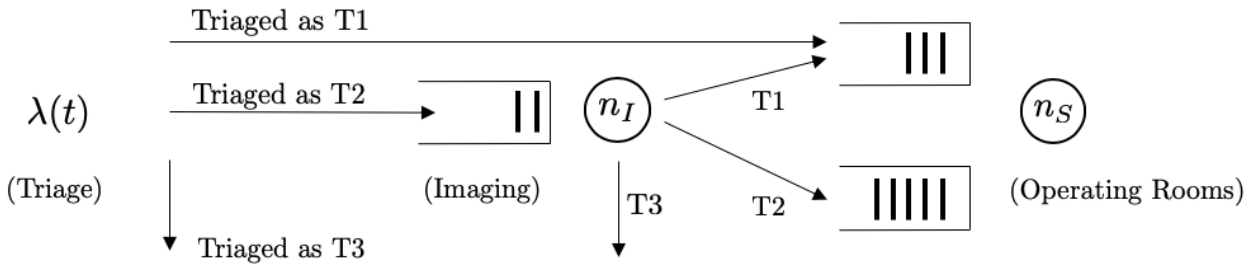


Figure 1: A schematic representation of the proposed queueing network under the assumption that all patients triaged as class 2 (T2) are routed to imaging prior to surgery.

For $i \in \{1,2,3\}$ denote by $\{X_i(t); t \geq 0\}$ and $\{Y_i(t); t \geq 0\}$ stochastic processes that keep track of the number of (actual) class i patients in the surgery node and imaging node, respectively. Note that class 3 patients may enter the system, only if mistriaged as class 2. Denote by $\{\tilde{X}_i(t); t \geq 0\}$ and $\{\tilde{Y}_i(t); t \geq 0\}$ the processes that keep track of the number of patients triaged as class $i \in \{1,2\}$ in the surgery and imaging nodes, respectively. (Patients triaged as T3 do not enter the system.) Note that $\tilde{X}_i$ and $\tilde{Y}_i$ are observable for all i. Denote by $\{D_i(t); t \geq 0\}$ for $i \in \{1,2\}$ the number of class i deaths occurred by time t. Our main performance measure of interest is the expected mortality in the system during the interval $[0, \tau]$, that is,

$$\mathbb{E}\left[D_1(\tau) + D_2(\tau)\right],$$

where $\tau = \inf\{t \geq t_0; \sum_{i=1}^{3}(X_i(t) + Y_i(t)) = 0\}$ is the time at which the system is empty after the arrivals stop. In all experiments, we assume that the system starts empty.

3.2 The Information-Delay Tradeoff and the Proposed Control Policy

Routing a class 2 patient to imaging identifies the true triage level of the patient and possibly uncovers one of the two possible mistriages: if a patient is overtriaged (i.e., is truly a class 3 patient) he/she no longer requires surgery. Hence, uncovering overtriaged patients reduces the demand for surgery, which in turn decreases the mortality for other patients by reducing the congestion. If on the other hand an undertriaged patient is identified (i.e., the patient is truly a class 1) the patient is prioritized for surgery, improving his/her chance of survival.

Despite these benefits, performing imaging requires time and resources. Since the imaging capacity is also limited, congestion can also occur at the imaging node of the network. If too many patients are routed to imaging, this might lead to idleness at the surgery rooms, and further delay patients who are correctly triaged and require surgery, possibly increasing mortality.

To better manage the utilization of the imaging and avoid idleness at the surgery rooms, here we propose a simple and intuitive control policy. According to this threshold-based policy, a class 2 patient is only

sent to imaging if the number of patients waiting for surgery exceeds a fixed threshold value. Formally, under a routing policy with threshold T, a patient arriving at time $t \geq 0$ is routed to the imaging node if and only if,

$$\tilde{X}_1(t) + \tilde{X}_2(t) \geq T,$$

and otherwise directly proceeds to surgery. Note that the policy only uses information on the number of patients currently waiting for or receiving surgery. Intuitively, the policy aims to send patients to imaging if the congestion at the surgery node is large enough so that with high probability the waiting time of the patient before receiving surgery exceeds the time required for the patient to receive imaging. We note that this policy is not the "optimal" policy with respect to minimizing mortality. Characterizing the structure of the optimal policy is beyond the scope of this paper and is left for future work. Our goal here is to use a simple and practical policy to illustrate the benefits and importance of carefully managing the imaging capacity and utilization during an MCI.

4 SIMULATION EXPERIMENTS

4.1 Choice of Model Parameters

In this section we discuss the calibration of the model input for the simulation experiments. We calibrate the model using information reported in the literature and in consultation with expert radiologists.

Patient types and mistriage probabilities. We set the triage probabilities to $p_1 = 0.15, p_2 = 0.33$, and $p_3 = 0.52$. These values are relevant to a realistic MCI, although we note that the probabilities could vary for different MCIs. For simplicity, we assume that only class 2 patients can be mistriaged and vary the mistriage probabilities in our experiments. We note that this is a reasonable assumption since the majority of mistriages occur for T2 triage level (Berger et al. 2016). According to Frykberg (2002) although undertriage rarely happens, overtriage probability is typically high, e.g., ranged between $8\% - 80\%$ (average 53%) in 10 bombing incidents. Hence, in our experiments we vary the overtriage probability in $\bar{q}_2 \in [0,1]$ and the undertriage probability in $\underline{q}_2 \in [0, 0.5]$.

Arrival process. We consider a bi-modal arrival rate that occurs over $t_0 = 210$ minutes. Specifically, we use the arrival rate,

$$\lambda(t) = \begin{cases} -a \times 10^{-4} t^2 + 9a \times 10^{-3} t, & 0 \leq t \leq 90, \\ 0, & 90 < t \leq 120, \\ -b \times 10^{-4} (t-120)^2 + 9b \times 10^{-3} (t-120), & 120 < t \leq 210, \end{cases}$$

where a and b are parameters which we will vary in our experiments to get different average arrival rates given by,

$$\Lambda = \frac{1}{210} \left(\frac{243a}{20} + \frac{243b}{20} \right).$$

The shape of arrival functions captures two waves of arrivals caused by the emergency transportation services going back to the scene of the incident as exemplified in Cohen et al. (2014) for a real MCI. Note that for 30 minutes, the arrival rate is equal to 0 before the second wave of arrivals begin to arrive. This structure of the arrival rate will allow us to demonstrate the drawbacks of over-utilizing the imaging when capacity is scarce. We consider two values for the average arrival rates in our experiments, both corresponding to an event that last for hours and lead to overloaded servers at both the imaging and surgery nodes.

Surgery and imaging times. As reported in Huber-Wagner et al. (2009) the duration of emergency surgical operations could vary significantly depending on the type of the surgery. We assumed surgery times are Log-Normal. In consultation with our collaborators, we considered an average surgery time of 60 minutes and to capture the high-variability set the standard deviation to 60 minutes as well. Further,

we assumed that service times in the imaging node are Normally distributed with mean 20 minutes and standard deviation 3 minutes using the data reported in Mueck et al. (2016) from a mid-scale exercise in preparation for MCIs.

Remaining lifetimes. We assumed that remaining lifetimes of class 1 and 2 patients follow a Weibull distribution with cdf $G_i(t) = 1 - e^{-(t/\beta_i)^{\theta_i}}$ and class-dependent parameters. Similar to Sun et al. (2018), we set the shape parameter $\theta_i = 1.5$ for both classes and set $\beta_i = (1/\gamma_i)/\Gamma(1+1/\theta_i)$ where $1/\gamma_i$ is the mean remaining lifetime for class $i \in \{1,2\}$. In consultation with our collaborators, we assume $1/\gamma_1 = 400$ minutes and consider two cases for $1/\gamma_2 \in \{800, 1200\}$ in our experiments.

4.2 Design of Simulation Experiments

In all experiments, we obtain the estimates by generating sample paths of the processes introduced in Section 3.1 using discrete-event simulation. In our experiments, we gradually increase the number of replications until the approximate absolute error of the estimated mortality is smaller than 0.2% (see, e.g., Nelson (2013) Chapter 8) or a maximum of 25,000 replications is reached. As such, the error bars in our figures (corresponding to 95% confidence intervals) are hardly visible. Finally, to reduce the simulation error, we use Common Random Numbers (CRNs) in our experiments when comparing difference scenarios.

4.3 Value of Information

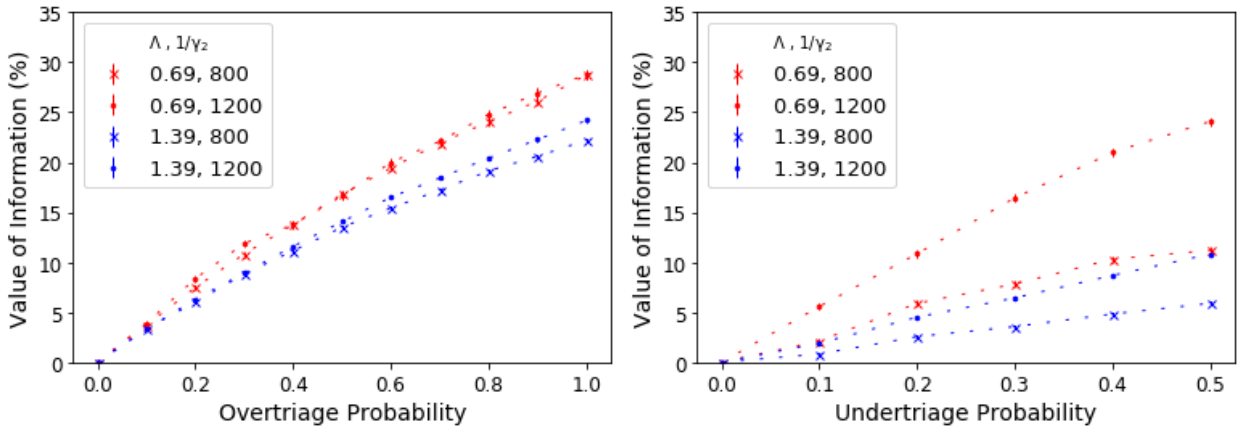
In this section, we conduct a set of simulation experiments to quantify the maximum value of information obtained by performing imaging. More specifically, we consider the queueing network without an imaging node (or equivalently under the policy of sending all patients triaged as class 2 directly to surgery) and compare the mortality under the following two cases: one with full information on patient types (i.e., with mistriage probabilities equal to zero) and another with fixed (non-zero) mistriage probabilities. We vary the undertriage and overtriage probabilities and estimate the expected *relative* reduction in mortality achieved under the case with full information. In doing so, we aim to understand the maximum value of using imaging in the absence of congestion and imaging times and identify parameter regimes where the benefits are high.

In Figure 2 we plot the estimated values of information for varying mistriage probabilities. In the experiments, we set $n_S = 9$, $1/\mu_S = 60$, $1/\mu_I = 20$, $1/\gamma_1 = 400$, and consider two cases for the average arrival rate $\Lambda \in \{0.69, 1.39\}$ and mean class 2 remaining lifetimes $1/\gamma_2 \in \{800, 1200\}$. In the left plot (a) we fix the undertriage probability at $\underline{q}_2 = 0$ and vary the overtriage probability in $\overline{q}_2 \in [0,1]$. In the right plot, we fix the overtriage probability at $\overline{q}_2 = 0$ and vary the undertriage probability $\underline{q}_2 \in [0, 0.5]$.

As expected, the value of information is increasing in both mistriage probabilities. Note that the value could be very significant even for moderate mistriage probabilities. Further, observe that identifying overtriages leads to a higher relative reduction in mortality when compared to undertriages. This implies that the value of reducing the arrival rate to the surgery node (by identifying patients who do not need surgery) dominates that of better prioritization for the set of parameters considered in our experiments.

Further, observe that the relative value of information is higher in the case of the smaller average arrival rate. This is because the baseline mortality rate (no mistriage) is smaller in this case, leading to a higher relative reduction in mortality. More interestingly, we observe that the value of identifying undertriage probabilities could be considerably higher for the case with the higher average remaining lifetime of class 2 patients. This can be explained noting that prioritizing class 2 patients has a larger impact on class 1 mortality when the remaining lifetime of class 2 patients is longer.

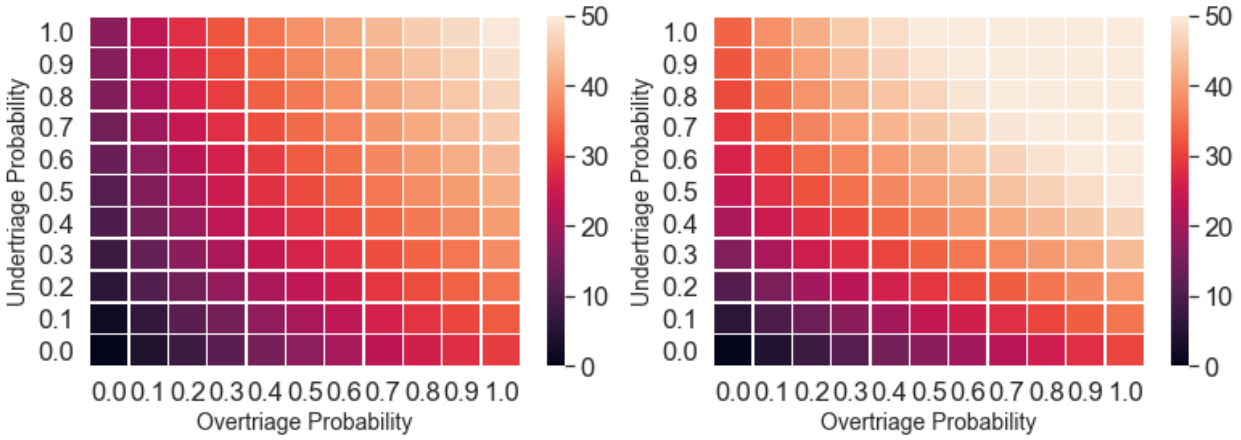
In Figure 2, we only vary one of the mistriage probabilities and keep the other at zero. We observe however that the observations are consistent when we jointly vary both mistriage probabilities, suggesting that there is little interaction between the two effects. To illustrate this, we plot the estimates in Figure 3 for varying under and overtriage probabilities and arrival rate $\Lambda = 0.69$. We keep the other parameters the same as in Figure 2. A lighter color corresponds to higher value of information.



(a) $\overline{q}_2$ varies in $[0,1]$, $\underline{q}_2 = 0$.

(b) $\overline{q}_2 = 0$, $\underline{q}_2$ varies in $[0,0.5]$.

Figure 2: Estimated expected relative reduction in mortality achieved under the full information case for varying overtriage (a) and undertriage (b) probabilities while keeping the other at zero; average arrival rates $\Lambda \in \{0.69, 1.39\}$; and mean class 2 remaining lifetimes $1/\gamma_2 \in \{800, 1200\}$. Other parameters are set to $n_S = 9$, $1/\mu_S = 60$, $1/\mu_I = 20$, $1/\gamma_1 = 400$.



(a) $1/\gamma_2 = 800$

(b) $1/\gamma_2 = 1200$

Figure 3: Heatmaps of estimated expected relative reduction in mortality achieved under the full information case for mean class 2 remaining lifetimes $1/\gamma_2 = 800$ in (a) and $1/\gamma_2 = 1200$ in (b). Other parameters are set to $\Lambda = 0.69, n_S = 9$, $1/\mu_S = 60$, $1/\mu_I = 20$, $1/\gamma_1 = 400$.

The reductions of mortality observed in the above experiments are obtained assuming zero service times at the imaging node and hence provide upperbounds for the actual benefits obtained in the case of limited imaging capacity and non-zero processing times. In the following section, we investigate the value of imaging with finite imaging capacity and positive and uncertain processing times.

4.4 Managing the Information-delay trade-off

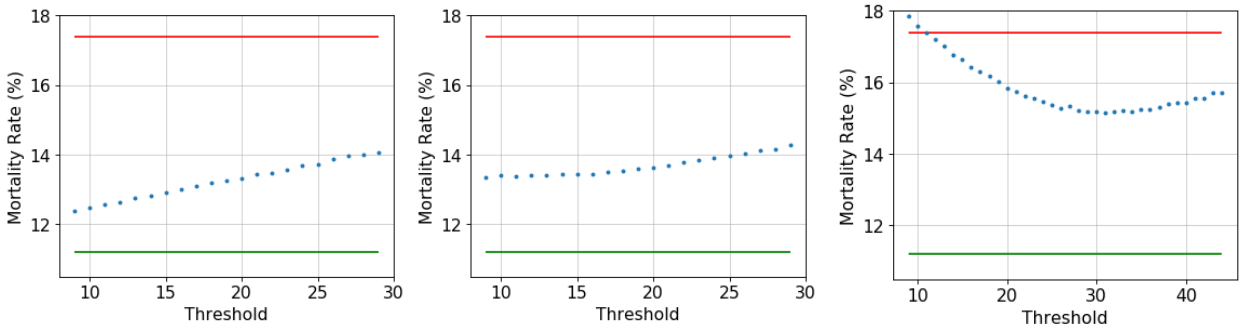
In this section we assume that the imaging node has a finite capacity n_I and that the processing times are random variables following the distribution specified in Section 4.1. In particular, the average imaging

time is assumed to be 20 minutes $\mu_I = 1/20$ which is relatively shorter than average surgery time of 60 minutes $\mu_S = 1/60$. In addition, imaging times are also less variable compared to surgery duration. This is reasonable to assume, since during an MCI imaging is done according to specific protocols aiming to standardize the process and minimize the duration (Mueck et al. 2016).

We use simulation experiments to estimate the expected mortality for different levels of imaging capacity. To this end, we fix the surgery capacity at $n_S = 9$ and consider three cases for the imaging capacity $n_I \in \{2,3,4\}$. Note that since $\mu_I/\mu_S = 3$, the three cases correspond to imaging capacity (i.e., maximum throughput) being less, equal, and above the surgery capacity, respectively. Other parameters are set to $\overline{q}_2 = 0.4$, $\underline{q}_2 = 0.1$, $\Lambda = 0.69$, $1/\gamma_2 = 800$. Note that from the results of Section 4.3, the mistriage probabilities correspond to a regime with high value of information.

The routing decision - whether to send class 2 patients to imaging - is made according to the policy described in Section 3.2. We vary the threshold value in the experiments. Note that when the threshold is equal to the surgery capacity, the policy is equivalent to routing class 2 patients to imaging whenever the surgery is at full capacity (i.e., all servers are busy). As the threshold increases, the utilization of the imaging node decreases and eventually converges to zero.

Figure 4 presents the results. In addition to the estimates, the plots also show the expected mortality for the full information case (green line) and no imaging (red line). We observe that when the imaging capacity is large $(n_I = 4)$ mortality is minimized at the smallest threshold value, whereby almost all patients are routed to imaging. When the imaging capacity is smaller, however, routing a smaller fraction of class 2 patients to imaging (i.e., higher threshold value) could lead to lower mortality. In the case of only 2 servers in the imaging node $(n_I = 2)$, routing all patients to imaging leads to higher mortality even compared to the scenario of performing no imaging at all. Nevertheless, using a large enough threshold value - which only sends patients to imaging when the surgery node is congested enough - leads to a significant reduction of mortality (from 18% to 15%) compared to no imaging.



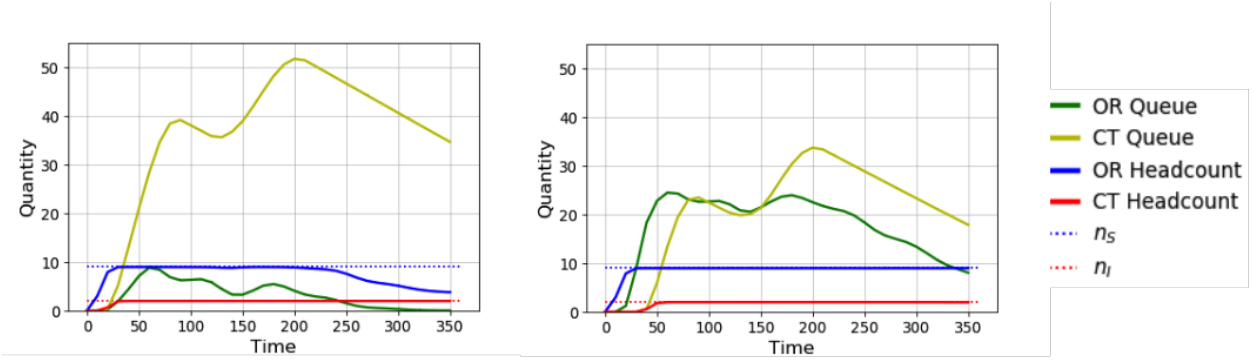
(a) n_I = 4 (b) n_I = 3 (c) n_I = 2

Figure 4: Estimated expected mortality under different imaging capacity levels and varying threshold values. Other parameters are set to $\overline{q}_2 = 0.4$, $\underline{q}_2 = 0.1$, $\Lambda = 0.69$, $1/\gamma_2 = 800$, and $n_S = 9$. The horizontal lines correspond to estimates of the expected mortality under full information (green) and no imaging (red).

The above observation can be explained noting that with limited CT capacity, routing too many class 2 patients to imaging may turn the imaging into the bottleneck capacity, leading to idleness at the surgery node. This can in particular occur with a time-varying arrival rate with periods of time when arrival rates are decreasing or are equal to zero as in our experiments.

To further illustrate the above observation, in Figure 5 we plot the average sample paths of the total number of patients in the surgery and imaging nodes of the network for the case with $n_I = 2$. More specifically, the plots are estimates of $\mathbb{E}[X_1(t) + X_2(t) + X_3(t)]$ and $\mathbb{E}[Y_1(t) + Y_2(t) + Y_3(t)]$ in time for two cases; (a) the case with threshold equal to 9 and (b) the case with threshold equal to 30 which results in the

lowest mortality rate. Observe that in case (a) the imaging (CT) queue can be very long whereas the surgery (OR) queue remains low. In addition, as the arrival rates start going down (before the arrival of the second wave, and at the end of the horizon) idleness is incurred in the surgery node (i.e., patients are waiting in CT whereas surgery capacity is idle). In this case, many patients who do need surgery are delayed in the imaging node overall leading to higher mortality than performing no imaging at all. In contrast, in case (b) no idleness is incurred as the arrival rate goes down. In this case, both the imaging and surgery maintain full utilization and incur no idleness, except the end of the horizon as the system starts emptying.



(a) Threshold = 9 (b) Threshold = 30

Figure 5: Estimated expected number of patients in the surgery (OR) and imaging (CT) nodes in time (for the first 350 time units) for a system with $\overline{q}_2 = 0.4$, $\underline{q}_2 = 0.1$, $\Lambda = 0.69$, $1/\gamma_2 = 800$, $n_S = 9$, $n_I = 2$, and two threshold values.

5 DISCUSSION AND FUTURE WORK

We propose a queueing network model of the ED during an MCI and use simulation experiments to quantify the value of identifying mistriaged patients through imaging. Imaging can correctly identify the triage level but takes a non-negligible time and the imaging capacity is also limited. Our main observations are as follows: (1) identifying mistriaged patients can significantly reduce the mortality rate, even for moderate (realistic) mistriage probabilities. In particular, identifying overtriaged patients can significantly reduce the surgery waiting time for real class 2 patients (and hence their mortality rates) by eliminating the need for surgery for patients who do not need it. (2) When the capacity of imaging is scarce, one needs to carefully manage the information-delay tradeoff that arises when deciding whether to send a patient to imaging before surgery or not. In particular, over-utilization of imaging can lead to idleness in the surgery and inflate the waiting time of patients who actually do need surgery. When this tradeoff is properly taken into account, using imaging can significantly reduce the mortality rate compared to the case of performing no imaging at all.

Our model clearly stylizes the operations of the ED during an MCI and we make certain modeling assumptions that may not hold in reality. For example, we assume that only T2 patients are subject to mistriage, whereas other types of mistriages do occur in reality. We assumed surgery times do not depend on the triage level and assumed fixed capacity (in time) for both imaging and surgery. Nevertheless, our experiments using other parametric assumptions, inputs and more detailed modeling of the processes confirm that the high-level observations from the model are fairly robust, although the quantitative values can change. Our proposed control policy assumes a strict rule for routing patients to imaging based on the congestion in the system. While this may not be practical in a real MCI setting, we believe that the qualitative message is still relevant: when routing a patient to imaging one should take its limited capacity into account and be cautious of possible idleness in the surgery rooms.

Our work can be extended in various practical and technical directions in the future. One important direction is to quantify and compare the value of other imaging procedures (e.g., X-ray) in the presence of congestion effects. Different imaging procedures provide different levels of information and vary in terms of the required resources and duration. Further, here we focus on a simple threshold policy to illustrate the importance of managing imaging capacity. As discussed, this policy is sub-optimal and can be improved by taking other state and arrival information into account. In doing so, however, one needs to take the practicality of the policy in mind, which is paramount in a MCI setting. Finally, in this work we rely on simulation experiments to obtain our results. Fluid models, which capture the transient dynamics of the system well, can provide accurate approximations as well as additional insights as in Cohen et al. (2014). In our model, the presence of mistriaged patients complicates the dynamics of the system and requires carrying more information. Developing tractable fluid approximations for our model is left for future work.

ACKNOWLEDGMENTS

We would like to thank the participants and organizers of the simulation workshop at the 2018 annual meeting of the American Society of Emergency Radiology (ASER). In particular, we are grateful to Ronald Bilow, Richard Bradley, Brandi Hicks, and Faisal Khosa.

REFERENCES

Alizamir, S., F. De Véricourt, and P. Sun. 2013. "Diagnostic Accuracy Under Congestion". *Management Science* 59(1):157–171.

Argon, N. T., and S. Ziya. 2009. "Priority Assignment Under Imperfect Information on Customer Type Identities". *Manufacturing & Service Operations Management* 11(4):674–693.

Argon, N. T., S. Ziya, and R. Righter. 2008. "Scheduling Impatient Jobs in a Clearing System with Insights on Patient Triage in Mass Casualty Incidents". *Probability in the Engineering and Informational Sciences* 22(3):301–332.

Berger, F. H., M. Körner, M. P. Bernstein, A. D. Sodickson, L. F. Beenen, P. D. McLaughlin, D. R. Kool, and R. M. Bilow. 2016. "Emergency Imaging After a Mass Casualty Incident: Role of the Radiology Department During Training for and Activation of a Disaster Management Plan". *The British Journal of Radiology* 89(1061):20150984.

Cohen, I., A. Mandelbaum, and N. Zychlinski. 2014. "Minimizing Mortality in a Mass Casualty Event: Fluid Networks in Support of Modeling and Staffing". *IIE Transactions* 46(7):728–741.

Frykberg, E. R. 2002. "Medical Management of Disasters and Mass Casualties From Terrorist Bombings: How Can We Cope?". *The Journal of Trauma* 53:201–212.

Huber-Wagner, S., R. Lefering, M. Kay, J. Stegmaier, P. Khalil, A. Paul, P. Biberthaler, W. Mutschler, K.-G. Kanz, and the Working Group on Polytrauma (NIS) of the German Trauma Society (DGU). 2009. "Duration and Predictors of Emergency Surgical Operations-basis for Medical Management of Mass Casualty Incidents". *European Journal of Medical Research* 14(12):532–540.

Hupert, N., E. Hollingsworth, and W. Xiong. 2007. "Is Overtriage Associated with Increased Mortality? Insights from a Simulation Model of Mass Casualty Trauma Care". *Disaster Medicine and Public Health Preparedness* 1(S1):S14–S24.

Levi, R., T. Magnanti, and Y. Shaposhnik. 2019. "Scheduling with Testing". *Management Science* 65(2):776–793.

Mueck, F. G., K. Wirth, M. Muggenthaler, U. Kreimeier, L. Geyer, K.-G. Kanz, U. Linsenmaier, and S. Wirth. 2016. "Radiological Mass Casualty Incident (MCI) Workflow Analysis: Single-Centre Data of a Mid-Scale Exercise". *The British Journal of Radiology* 89(1061):20150918.

Nelson, B. 2013. *Foundations and Methods of Stochastic Simulation: a First Course*. New York: Springer US.

Sun, Z., N. T. Argon, and S. Ziya. 2018. "Patient Triage and Prioritization Under Austere Conditions". *Management Science* 64(10):4471–4489.

Yom-Tov, G. B., and A. Mandelbaum. 2014. "Erlang-R: A Time-Varying Queue with Reentrant Customers, in Support of Healthcare Staffing". *Manufacturing & Service Operations Management* 16(2):283–299.

AUTHOR BIOGRAPHIES

JOSEPH SISMONDO is a senior undergraduate student in Industrial Engineering at the University of Toronto. His research interests include simulation and data analytics, with a focus on applications in healthcare and electoral politics. His email address is joe.sismondo@gmail.com.

VAHID SARHANGIAN is an Assistant Professor of Mechanical and Industrial Engineering, and the Associate Director of the Centre for Healthcare Engineering at the University of Toronto. His research focuses on modeling, simulation, and control of stochastic dynamical systems that arise in service and healthcare operations. His email address is sarhangian@mie.utoronto.ca.

EMMETT BORG is a Resource Strategist at Hoffmann-La Roche. He holds a Master's of Engineering in Operations Research from the University of Toronto. His research interests include data science and simulation applications in healthcare and drug development. His email address is emmettborg@gmail.com.

ERIC ROBERGE is an Assistant Professor of Radiology at the Uniformed Services University of the Health Sciences and the Chief of Radiology at Madigan Army Medical Center. His clinical work focuses on cardiac imaging and emergency radiology. His research focus is on the use of radiology in mass casualties and the fielding of radiology capabilities in austere environments. His email address is earoberge@gmail.com.

FERCO H. BERGER is a radiologist with dedicated training and experience in Emergency & Trauma radiology and specifically concepts of MCIs. He is a practicing radiologist and an Assistant Professor at the University of Toronto with research interests in optimizing imaging for patients requiring acute care in routine as well as MCI settings. His email address is ferco.berger@sunnybrook.ca.