Analyzing and modeling Mass Casualty Events in hospitals – An operational view via fluid models

MSc Research Proposal

Noa Zychlinski

Advisors: Dr. Cohen Izack, Prof. Mandelbaum Avishai

The faculty of Industrial Engineering and Management

Technion – Israel Institution of Technology

February 2011

Contents:

1. Introduction:

A Mass-Casualty Event (MCE) is an event in which the number of casualties exceeds the capacity for taking care of them. Such events, unfortunately, happen all the time. They may have either a world-wide effect (e.g., the 2004 Indian Ocean tsunami that killed over 200,000 people) or a local one (e.g., a bus accident that sends tens of casualties to a hospital) – in both cases it is an MCE if there is an imbalance between the number of casualties that need to be treated and the number of resources that should treat them. Therefore it is very important to prepare for such events. We concentrate on the healthcare aspects of MCE, particularly in hospitals. A central problem in establishing a hospital's emergency plan is the inability to forecast the performances of healthcare services. When an MCE occurs, suddenly and all at once, the demands on healthcare staff and facilities increase, and an emergency plan must be quickly implemented. Such a plan has clinical and operational components. In our research we focus on the latter.

The main approach for modeling MCEs has been through simulation. The purpose of our research is to develop a mathematical model (fluid model) that captures the operational performance of a hospital during and after an MCE. Such a model would enhance the understanding of MCEs effects and would allow for better preparation to the future ones. Possible outputs of the research will be capacity planning of the resources that are needed during an MCE (e.g., beds, medical staff, space etc.) and operational policies (e.g., allocation of equipment, identification of bottlenecks etc.)

The rest of this proposal is structured as follows. In Section 2 we present a literature review on MCEs and on Fluid Models. In section 3 we discuss our model and explain the methods that will be used in order to analyze it.

2. Literature Review

2.1 Mass Casualty Events

A mass-casualty event (MCE) is defined as an unusual situation in which at a certain moment, there are more casualties than the system is able to manage. During MCS, the hospital emergency services must treat a large number of patients that suddenly arrive [1]. When a disaster occurs, the number of patients who require treatment in Emergency Departments may increase 3–5 times the normal volume, which easily could overwhelm hospitals' resources [2].

The major challenges that hospitals face in an MCE include surge capacity issues, the fact that they are already at or near capacity for emergency and trauma services, a lack of on-call specialists and nurses, the need to coordinate between competing health care systems and incompatibilities in communications systems [3]

MCEs are classified into two categories: (1) those that result in an immediate or sudden impact and (2) those that result in a developing or sustained impact [3]. From a Service Engineering point of view, this classification is based on the arrival rate of casualties to receive medical service.

The first category of MCEs includes events such as the detonation of a conventional bomb, NBC (nuclear, biologic, chemical) attack, airplane or train crashes and natural disasters such as earthquakes or tsunamis. This immediate impact category is characterized by a large numbers of casualties at the outset of the event. In some cases there may be a second wave of casualties due to secondary exposure.

The second MCE category features events such as a massive exposure to anthrax or smallpox. Another example of this type of MCE is the potential case of an influenza pandemic, in which there would be a gradual increase in the number of people affected, possibly rising to a catastrophic number of patients. In this type of MCE, the number of cases may decline due to treatment and prophylactic efforts. This type of MCE would necessitate a more sustained response, as the impact would be felt over a much longer period than the immediate-impact MCE. In our research we will focus on the first category of MCEs with sudden immediate impact.

An MCE occurrence affects nearby hospitals, which have been working in steady state, treating a reasonable amount of patient, and now must start treating a large number of patients. The lack of adequate staff and equipment may cause injured patients to not receive the same level of medical treatment that would have been provided had they been treated as an individual rather than as one of multiple casualties arriving simultaneously. In order to overcome the temporary lack of qualified staff and resources, the hospital prepares itself to work under a triage strategy. Upon arrival, patients are triaged by the severity of their injuries. Severely injured must be treated immediately, since any delay can endanger their lives. The hospital allocates separate locations for each group so as to ensure that the majority of resources are allocated to the most severe injuries.

In MCEs that include toxic exposure (chemical, biological, radiological or nuclear) or contamination threat, the hospital activates its disaster plan and directs all arrivals through the emergency department [4], [5]. Adjacent to the ED entrance, the hospital establishes a decontamination zone. Casualties, entering the hospital must be decontaminated; only then, are they allowed entering the hospital. The most important step in decontamination is the speed of the removal of the agent [6]. Patients going through the decontamination procedure must remove their clothing as quickly as possible as part of their decontamination. Their personal items are carried with them in a sealed plastic bag. This procedure can be accomplished in most cases, though not always, by the casualties without assistance from medical personnel. The next step is rinsing the casualties with large quantities of water (high volume, low pressure). Gently scrubbing the skin with soap and a soft brush removes any remaining fat-soluble chemicals and solid materials [5]. Delaying or improperly conducting decontamination may inadvertently increase the dangers to the patient as well as to emergency (healthcare) providers [7].

Designing the decontamination zone can be done in several ways. Frezza et al [8] examine different designs. Their suggestion is to establish two decontamination lanes: one regular for stable patients and an emergency track for unstable patients.

When an MCE happens in a place with poor or no medical infrastructure, the local population and government are helpless and need assistance from other countries. An example of such an event is the earthquake that struck Haiti on January 2010 [9], [10]. The number of deaths directly and indirectly by the earthquake is estimated to be 230,000, plus approximately 250,000 injured people. Israel Defense Forces Medical Corps sent a delegation to Haiti, consisting of 121 medical personnel who established a field Hospital. The hospital was fully operational on site merely 89 hours after the earthquake struck and was capable of providing sophisticated medical care. During the 10 days that the hospital was operational, its staff treated 1111 patients, hospitalized 737 patients, and performed 244 operations on 203 patients [9].

Due to limited resources of the field hospital, not all casualties seeking help actually received it. A triage algorithm had to be implemented in order to decide which victim would be admitted to the hospital and which would not. The hospital staff was constantly required to deal with ethical dilemmas [10]. The Israel field hospital in Haiti managed to treat 100 patients a day despite the fact that its bed capacity was only 60 (later expanded to 72). The reason for this was efficiency and flexibility in resource allocation and staffing. For example, the distribution of injuries had changed during the time, and the hospital had to constantly re-balance its resources accordingly.

2.2 Modeling an MCE

Modeling healthcare systems or emergency departments in general and MCEs in particular is difficult due to the complexity, size and dynamic of the systems involved.

Most researches focus on modeling and simulating emergency department in their steady state.
However, the hospital model in disaster management differs from the model of normal operations in the characteristics of patient arrivals, which renders the system essentially transient in nature [11]. Modeling hospital operations has been done by several methods: linear and dynamic programming [12], [13], queueing models [14], system dynamics [15],[16],[17] and discrete event simulation [11], [18], [19]. Although linear and dynamic programming models allow quantitative analysis, these mathematical models are typically unable to accommodate either the random nature or transient nature of MCEs. Queueing models are capable of representing stochastic processes but they often only capture long-term system performance.
System Dynamics is a methodology that analyzes systems by modeling their components (sub-systems) and the internal connections (feedback loops) between them. A system dynamics model consists of an interlocking set of differential and algebraic equations. A completed model may contain hundreds of such equations along with the appropriate numerical inputs. Applications of system dynamics in healthcare systems exists either for solution-oriented systems (e.g. cancer scanning [15] or chronic disease prevention [16]) or for macroscopic system modeling (broad view of population health [16], total patient flow through the U.K. National Health Service in emergency and extended care [17]).
Simulation is widely accepted as an effective method for assisting management in evaluating different operational alternatives [18]. Marmor et el had developed a general flexible simulation tool for Emergency Departments, which provides estimates regarding the current operational state and enables short term operational planning. Aliyas et el [11], have developed a transient model for hospital ED using a simulation and a set of meta models. Their model enables them to predict patients' waiting times and estimate hospitals' capacities (for all the hospitals in the disaster region). The suggested model does not include estimation of arrival rate which keeps changing in a disaster event.
In another research, a discrete-event computer model of an emergency room and related hospital facilities was constructed and implemented, based on accumulated data from 12 urban terrorist bombing incidents in Israel [20].The simulation predicted that the admitting capacity of the hospital depends primarily on the number of available surgeons and defines an optimal staff profile for surgeons, residents, and trauma nurses. The researchers concluded that the major bottlenecks in the flow of critical casualties are the shock rooms and the computed tomography scanner but not the operating rooms. The simulation also characterized the number of support staff needed to treat noncritical casualties and showed that radiology is the major obstacle to the flow of these patients.

2.3 Fluid Models

Fluid models provide a useful approximation that support performance analyze and control of large systems (high arrival rate and large number of servers) that vary in time. For example, fluid models have been used to design and analyze Markovian service network, in which large demand is being supplied by a large number of servers. The approximation becomes more accurate as the system grows [21]. The basic fluid model (figure 1) refers to the system as a black box having an arrival rate function λ(t) and a departure rate function δ(t), t≥0.

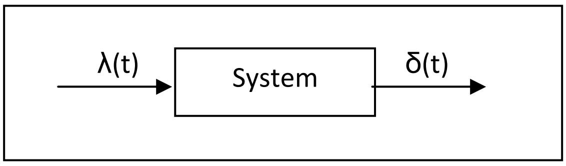


Figure 1 – a black box model

Let Q(t) represents the total amount of "fluid" in the system, it can be calculated by solving the differential equation:

$$\frac{\Delta Q(t)}{\Delta t} = \lambda(t) - \delta(t), \quad t \geq 0, \quad Q(0) = Q^0.$$

If the system consists of N(t) servers at time t, each one with service capacity μ, then we get the following differential equation on a refinement of the above:

$$\frac{\Delta Q(t)}{\Delta t} = \lambda(t) - \mu * \min(Q(t), N(t)), \quad t \geq 0, \quad Q(0) = Q^0.$$

Mandelbaum at el [22] used a fluid approximation for modeling a multi-server queue in a single service system station with abandonment and retrials. The model was proven accurate both in steady state and in transient state, the latter caused by a sudden peak in the arrival rate. Solving the model's equations yields an estimation of the total number of people in the system at any time (total number in queue, in service and the total number that will retry). The model allows for a prediction of the time until the system returns to steady state.

Green et el [23] implemented fluid approximation to an overloaded financial service call center. The approximation allowed estimating the waiting time and queue length (which vary in time) and the total time until the system recovers from congestion that occurs during rush hours.

Whitt [24] developed a deterministic fluid approximation for the *G/GI/s + GI* queueing model with large *s* (queueing model for a general multi-server queue with customer abandonment). In his work, he focuses on steady state behavior. Numerical examples show good results when compared to exact result *for M/GI/s + GI* models, obtained via computer simulations and numerical algorithms.

Yom-Tov [25] expanded the framework of Mandelbaum et al. [21] on time-varying queues and developed fluid and diffusion limits for the Erlang-R model. Erlang-R captures the behavior of Re-Entrant customers, who cycle between need of service and being content: such reentrant customers are prevalent in healthcare. In her work, Yom-Tov showed that fluid approximations are not only useful in analyzing time-varying systems, but also help understand the transient behavior of systems in steady-state. More specifically, the Erlang-R model is a stochastic queueing process which consists of 2-node state-dependent queues: $Q(t) = (Q_1(t),Q_2(t))$. $Q_1(t)$ represents the number of Needy patients in the system (i.e., those either waiting for service or being served), and $Q_2(t)$ the number of content patients in the system, at time t. The fluid model yields an approximation for $Q(t)$, $t \geq 0$.
The developed model is then used to analyze mass-casualty events in which the arrival rate changes rapidly during a short time. A numerical example, in which arrival rate is multiplied fivefold over two hours, was simulated and compared to the fluid and diffusion approximation. The comparison showed high accuracy, under the assumption that the time that the system is in critically-loaded is negligible.

Liu et el [26],[27] analyze a deterministic fluid model for systems which alternate between overloaded intervals and underloaded intervals, having time-varying arrival rate and staffing, exponential or non-exponential service and a non-exponential abandonment time. When the system is under-loaded, the total system content is less than its service capacity, so that there is no waiting and external input flows directly into service at time-varying rate $\lambda(t)$, $t \geq 0$. When the system is overloaded, there is no spare service capacity, so that the input is buffered in a queue, where abandonment occurs. Liu et el developed algorithms to describe time-dependent performance. They determined the time-varying potential waiting time, i.e., the virtual waiting time of an arrival at a specified time, assuming that it will not abandon. Simulations of queueing systems confirmed that the algorithm and the approximation were effective even when the number of servers was as low as 20. They also showed that non-exponential service distribution played an important role in the fluid dynamics.

3. Our proposed research

In our work, we shall focus on two types of MCEs. The first one is a conventional MCE, one that is caused by events such as a bomb, train/ bus crash or an earthquake. The second type of MCE that will be examined is an NBC (Nuclear, Biological, Chemical) MCE. The hospitals reaction to NBC MCEs differs from the reaction to conventional MCEs. The major differences between the two types are: (a) decontamination procedure that needs to be initiated to prevent staff and other patients' contamination. (b) The treatment in case of an NBC exposure is cyclic: medication is given to the casualties every predefined period. The casualty's condition dictates the number of cycles and their duration. (c) The functionality of healthcare staff (doctors and nurses) in NBC MCE is similar: both are allowed to administer medicines and decide whether an additional cycle is needed.

3.1 Objectives

The objectives of our research are to:

(1) Develop a fluid model for a hospital's Emergency Department (ED) during an MCE.
(2) Predict the performance of the ED in MCEs (e.g. total number in the system, average length of stay, average waiting time).
(3) Determine staffing and equipment allocations in order to provide certain service levels.

3.2 Research plan

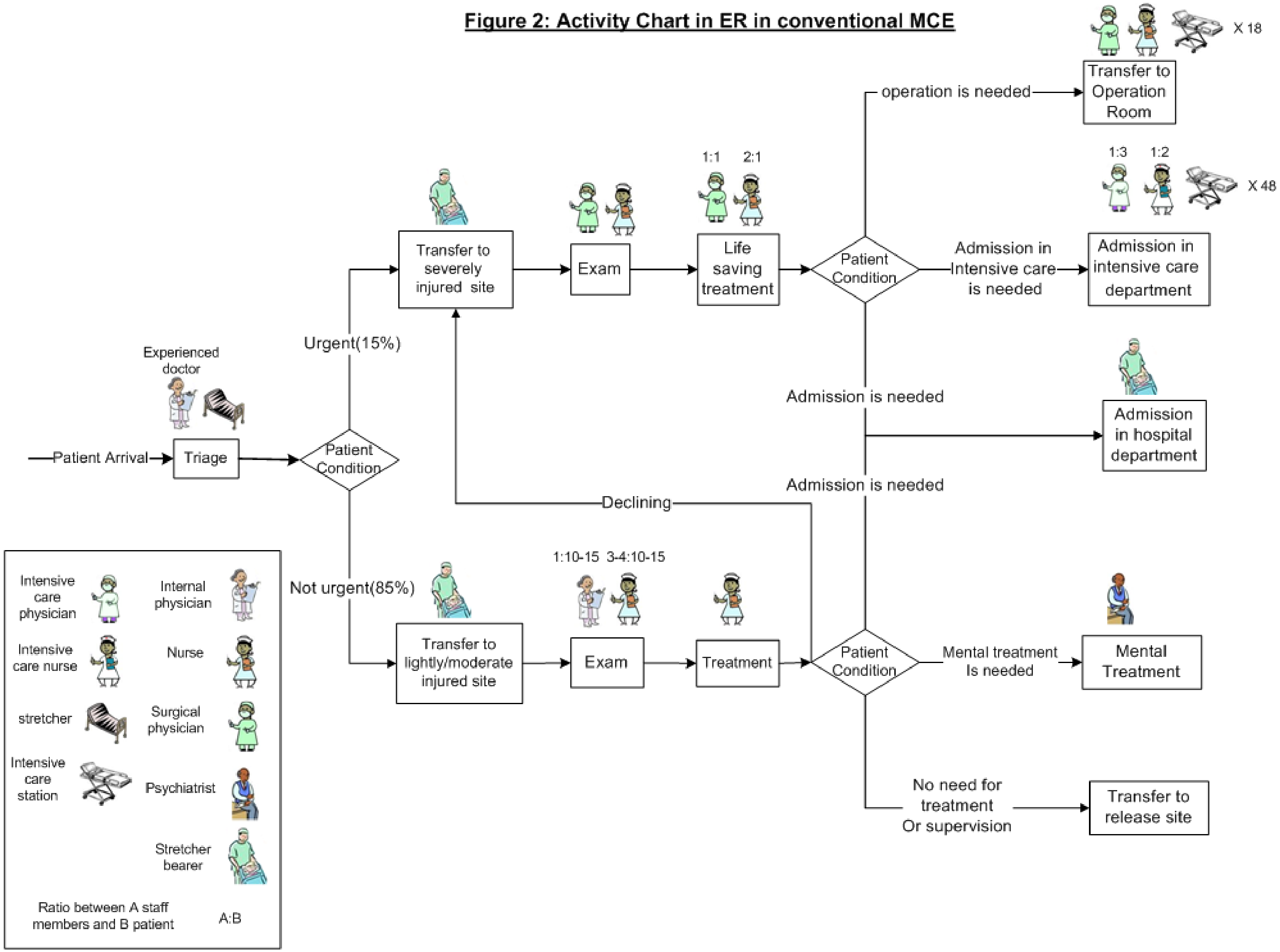
The plan to achieve the objectives consists of:

(1) Mapping and modeling the hospital operations in MCEs and in NBC MCEs in particular, considering activities and resource allocation.
(2) Development of fluid approximations to estimate performance measures.
(3) Testing the approximations using examples and data collected from a chemical warfare drill that was conducted at Rambam hospital on May 2010.

3.3 Conventional MCE

When an MCE occurs, the hospital usually receives an advanced notice. Then, it immediately activates its emergency plan. The activities (patients) go through in the ED as shown in Figure 2. The resources include staff and equipment. Healthcare staff includes (a) physicians: internal, surgical and intensive care, (b) nurses: ED nurses and intensive care (c) stretcher bearers. The equipment category covers stretchers, intensive care positions and operation room (OR).

Figure 2: Activity Chart in ER in conventional MCE



3.4 NBC (Nuclear, Biological, Chemical) MCE

The hospital emergency plan in NBC MCE is completely different from the plan in a conventional MCE. As already described, the former is activated, a decontamination zone is established. Entry to the hospital is allowed only after going through decontamination. After decontamination, patients go through another triage where they are directed to a treatment zone according to clinical condition (light injuries, moderate injuries, combined trauma injuries and severe injuries).

The treatment is given in cycles. The duration of each cycle is determined by the patient's clinical condition (10 minutes for severe injuries and 30 minutes for light and moderate injuries). Each cycle consists of 2 steps: the first one is when the medicine is given to the patient. The second is when the patient waits over the cycle duration according to clinical condition. At the end of each cycle, the medical staff decides whether another cycle is needed.

In the NBC MCE drill that took place at the Rambam hospital this year, an RFID (Radio Frequency Identification) system was used in order to track patients and staff. All entries and exits in the severely injured zone were monitored. We developed a preliminary fluid model according to the measured arrival rate. The comparison between the actual number of patients and the calculated number according to the fluid model is shown in Figure 3. The basic model shows relatively accurate results.

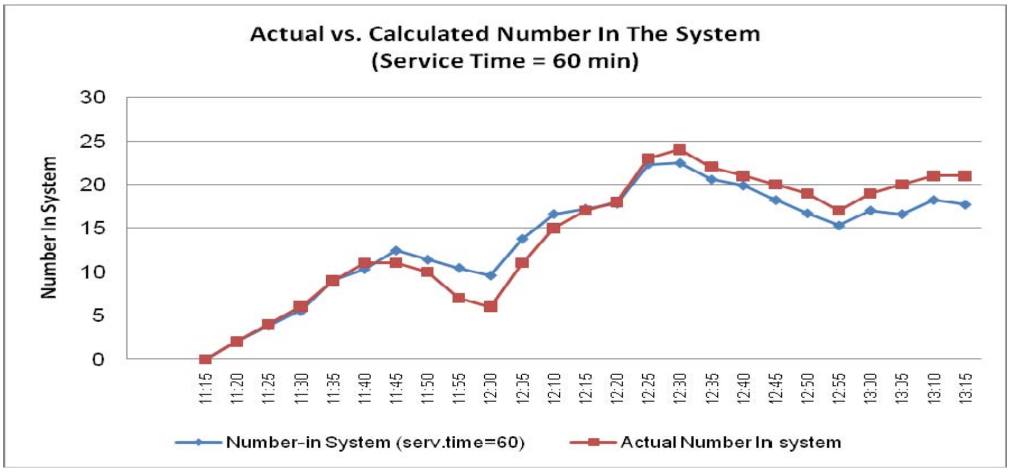


Figure 3: Comparison between actual numbers in the system vs. calculated number as derived from the fluid model

In our research, we shall improve this basic fluid model. We wish to expand it for the entire ED including the decontamination zone. In addition, we plan to develop a prediction for the event's dynamics, and finally analyze the resources required in order to identify optimal allocation and staffing.

4 References

[1] Levi, L., Michaelson, M., Admi, H., Bregman, D., & Bar-Nahor, R. (2002). National strategy for mass casualty situations and its effects on the hospital. Prehospital and Disaster Medicine, 17(1), 12-16.

[2] Henderson AK, Lillibridge SR, Salinas C, Graves RW, Roth PB, Noji EK: Disaster medical assistance terms: Providing health care to a community struck by Hurricane Iniki. Ann Emerg Med 1994;23:726–730.

[3] Phillips SJ, Knebel A, eds. Providing Mass Medical Care with Scarce Resources: A Community Planning Guide. Prepared by Health Systems Research, Inc. under contract no. 290-04-0010. AHRQ publication no. 07-0001. Rockville, MD: Agency for Health Research and Quality; 2006.

[4] Baker D. Civilian exposure to toxic agents: emergency medical response. emergency medical response.Prehospital Disaster Med 2004; 19:174–178.

[5] Levitin H, Siegelson H, Dickinson S, et al: Decontamination of mass casualties- Re-evaluating existing dogma. Prehospital Disast Med 2003;18:200–207

[6] Lake W: Chemical Weapons Improved Response Program, Guidelines for Mass Casualty Decontamination During a Terrorist Chemical Agent Incident.Domestic. Preparedness Program, US Soldier Biological and Chemical Command 2000.

[7] Institute of Medicine. National Research Council. Chemical and Biological Terrorism. Research and Development to Improve Civilian Medical Response.Washington, DC: National Academy Press, 1999. pp 97–109.

[8] Frezza, E. E., Fletcher, E., Flores, V., Popolo, E., & Placido, F. T. (2007). An engineer's mathematical approach to designing an emergency room decontamination area for weapons of mass destruction casualties. Journal of Social Sciences, 3(4), 223-31.

[9] Kreiss Y., Merin O., Peleg K., Levy G., Vinker S., Sagi R., et al. (2010). Early disaster response in haiti: The israeli field hospital experience. Annals of Internal Medicine 153(1), 45.

[10] Merin, O., Ash, N., Levy, G., Schwaber, M. J., & Kreiss, Y. (2010). The Israeli field hospital in haiti--ethical dilemmas in early disaster response. New England Journal of Medicine

[11] Paul, J.A., George, S.K., Yi, P., and Lin, L., 2006. Transient modelling in simulation of hospital operations for emergency response. Prehospital and Disaster Medicine,21 (4), 223–236.

[12] Boyd KM: The Ethics of Resource Allocation in Health Care. Edinburgh: University Press, 1979.

[13] Zon A, Kommer GJ: Patient flows and optimal health-care resource allocation at the macro-level: A dynamic linear programming approach. Health Care Manag Sci 1999;2:87–96.

[14] Bretthauer KM, Shetty B: The nonlinear knapsack problem—Algorithms and applications. Eur J Oper Res 2002;138:459–472.

[15] Royston G, Dost A,Townshend J,Turner H: Using system dynamics to help develop and implement policies and programmes in health care in England.System Dynamics Review 1999;15:293–313.

[16] Homer JB, Hirsch GB. System dynamics modeling for public health: background and opportunities. Am J Public Health 2006;96(3):452-8.

[17] Wolstenholme EF. A patient flow perspective of UK Health Services: exploring the case for new "intermediate care" initiatives. *System Dynamics Rev.* 1999;15:253–271.

[18] Sinreich D, Marmor Y (2004) "A simple and intuitive simulation tool for analyzing emergency department operations". Proceedings of the 2004 Winter Simulation conference.

[19] Lowery JC: Multi-Hospital Validation of Critical Care Simulation Model. In: 1993 Winter Simulation Conference Proceedings. IEEE, Piscataway.1993, pp 1207–1215.

[20] Hirshberg A, Stein M, Walden R. Surgical resource utilization in urban terrorist bombing: a computer simulation. J Trauma-Injury Infect Crit Care 1999; 47: 545–550.

[21] Mandelbaum A., W. A. Massey, M. I. Reiman. 1998. Strong approximations for Markovian service networks. Queueing Systems 30 (1-2), 149-201.

[22] Mandelbaum, A., W. A., Massey, M. I. Reiman, R. Rider. 1999. Time-varying multiserver queues with abandonments and retrials. In Proceedings of the 16th International Teletra±c Congress, P. Key, D. Smith (eds.), 355-364.

[23] Green L.V., Kolesar P.J. and Whitt W. (2005) Coping with time-varying demand when setting staffing requirements for a service system. Production and Operations Management, to appear. 2.4, 3

[24] W. Whitt, Fluid models for multiserver queues with abandonments, Oper. Res. 54 (2006), no. 1, 37{54}

[25] Yom-Tov G., Queues in Hospitals: Semi-Open Queueing Networks in the QED Regime, Ph.D. Proposal, IE&M, Technion (2007). 2.1

[26] Y. Liu and W. Whitt, A fluid model for a large-scale service system experiencing periods of overloading. working paper, Columbia University, New York, NY, 2010. Available at: http://www.columbia.edu/~ww2040/allpapers.html

[27] Y. Liu and W. Whitt, A Fluid Model for the Many-Server Gt/GI/st + GI Queue. Working paper, Columbia University, New York, NY, 2010. Available at: http://www.columbia.edu/~ww2040/allpapers.html