

Public Health Care As A Complex Adaptive System of Systems

**Thomas W. Moore, Patrick D. Finley, John M. Linebarger,
Walter E. Beyeler, Victoria J. Davey, Robert J. Glass**

Sandia National Laboratories
Albuquerque, New Mexico

[Complex Adaptive Systems of Systems \(CASoS\) Engineering Initiative](#)
tmoore@sandia.gov

Formulating policy for public health action can be characterized as designing solutions within a complex adaptive system-of-systems (CASoS). Public health systems are complex due to the large number of interdependencies and non-linear interactions among autonomous agents, such as individuals, health care organizations, and governmental agencies. They are adaptive in that the behaviors of individuals, organizations, and diseases are highly responsive to the behaviors of other such agents, as well as to hazards and natural disasters. System-of-systems structure is evident in their ability to be recursively decomposed into collections of interacting components, generally to an arbitrarily low level of detail. Ultimately, designed solutions (public health care actions) must be shown to be robust to the uncertainties inherent within the CASoS.

Adaptive systems, whether natural or artificial, evolve in response most strongly to proximal selective pressure. Inasmuch as the environments of adaptation contain stochastic components, these systems will tend to adapt in their ability to remain robust in the face of variable perturbations. However, these perturbations can vary in kind as well as in degree. A system that has evolved robustness to some kinds of perturbations can remain brittle against others, and the highly interconnected, interdependent nature of complex systems can respond to a brittleness-induced failure with a cascading collapse of functionality.

We implement a model of a health care system as an entity-based model consisting of actors (including staff and patients), diseases that progress according to a Markov process, resources distributed dynamically through the system and used in medical treatments, and organizational/environmental components, such as medical facilities. We adopt a flexible, component based model of the medical service unit (MSU), which can be used to represent facilities in a multi-scale, multi-network

architecture. The MSU can represent medical facilities ranging from field stations deployed to respond to an emergency, to emergency departments and intensive care units, to full hospitals, to whole health care networks. MSUs themselves can also be networked, allowing for the bottom-up construction of medical facilities, as when a health care network consists of multiple hospitals across a range of sizes, from local outpatient clinics to large hospitals serving major metropolitan areas.

Results will be given for a network of healthcare facilities subjected to two types of surges in demand: 1) a spike, representing a natural disaster which introduces patients with acute trauma requiring immediate treatment, and 2) a steady increase in demand imposed by the increasing prevalence of hepatitis C, which can result in a population with a significantly increased burden of end-stage liver disease and hepatocellular carcinoma. Policy actions are evaluated include striking the right balance between medical center-level policies and network-level policies (issues of multi-level selection), cost-effective resource utilization, and mediating resource conflicts. Our findings indicate that a critical determinant of policy success is a balance between network and medical center-directed policies.

1 Concepts

1.1 Highly Optimized Tolerance and Robust yet Fragile Functionality

Highly optimized tolerance (HOT) proposes a mathematical and logical framework that explains structural design and adaptation in response to experienced or anticipated perturbations [Carlson & Doyle, 2002].

Frequently, research on HOT considers perturbations that fall along a power law distribution in terms of size. Examples of such phenomena include damage and economic losses from forest fires, earthquakes, hurricanes, and tsunamis [Carlson & Doyle, 1998]. For some families of perturbations, however, it is also necessary to consider the shape of the perturbation, that is, whether the stimulus provided by the perturbation is a brief spike, a sustained surge, or cyclic fluctuations. Policies suited to address one family of perturbation shapes (say, using temporary resource redistribution to address immediate, short term needs created by an earthquake) might be inadequate to address another family of perturbations, for example a sustained increase in demand.

1.2 Origin of Structure in Organizations

Structures in organizations often exhibit imposed hierarchical relationships constituting a formal structure [Stevenson, 1990]. These structures facilitate and constrain the flow of information and resources through the network.

Within any given level of organization there may exist peer relationships between units. These peer relationships, consisting both of the network topology connecting the units and the policies governing resource and information exchanges between them, may be structured in such a way as to benefit or to hinder the

performance of the individual unit, with units that are more central and widely connected having a performance advantage [Kilduff & Tsai, 2003].

Conflict between levels of selection can occur when an adaptation favorable to one level creates a disadvantage at another level [Okasha, 2006]. One canonical example is the case of cancer, in which unregulated proliferation at the cellular level resulting from mutation and adaptation creates a cell line with relative reproductive advantage. At the level of the organism, however, cancer development can be highly maladaptive, encouraging the evolutionary development of multiple strategies operating at multiple levels to prevent carcinogenesis. Systemic functionality can only be developed and maintained by the regulation of the tendencies toward within group selection dynamics that favor the constituent units [Michod, 1999].

If we extend by analogy the biological paradigm of multilevel selection to analysis of the organization, we can begin to construct questions about the existence and nature of conflicts between organizational levels in the imposed hierarchical structure. What kind of selective pressures, originating from internal organizational dynamics or processes and environmental conditions external to the organization, lead to favor strategies from the constituent units that foster development at the expense of organizational peers, or at the expense of systemic functionality? If viewed from within the framework of designed and evolved tolerance against external perturbations, how do tolerance-based adaptations reflect potential conflicts between levels of organization?

To some extent, selective pressure on units in an organization is exerted by the organization through the imposed hierarchical structure, in terms of centrally set policies, resource and information flows, and decisions regarding the appointment of managers. If policies, resource distributions, and appointments are constructed with the goal of systemic improvement, there is still an opportunity for selection and reinforcement to occur at the level of the individual unit, which can possibly negatively impact other units or the function of the organization as a whole.

When selection is applied at the wrong level, for example through misidentification of the most desirable target of selection, systemically negative consequences can result. One example of this comes from the field of animal husbandry. Poultry farmers attempting to maximize egg production preferentially breed hens with the highest individual egg yields. Because relative egg yield positively correlated with aggressiveness, however, the modern practice of crated confinement lead to an increase in aggressive hens who exhibited violent tendencies towards cage-mates. The result was an overall decrease in egg production at the level of the crate, and at the level of the farm. The alternative strategy was preferential breeding of hens that exhibited positive tendencies in the crate environment, which reduced violent behaviors and resulted in an increased egg yield [Muir, 2003].

2 Architecture

Architectural elements of the system fall into three broadly defined domains: medical physics, governing disease progression and treatments; organizational physics, governing resource sharing between facilities and operational dynamics within

facilities; and social physics, governing interactions between individual agents within social networks. Findings presented in this paper integrate both the medical and organizational components.

2.1 Medical Physics

In this model, the domain of medical physics encompasses disease progression and treatments. Diseases are modeled generically as Markov processes, consisting of an arbitrary number of states and transition probabilities. Markov models are often used in medical science to build models of progression for diseases based on longitudinal studies, generating states pertinent to the particular diseases from medical observation, and deriving transitional probabilities from the longitudinal data sets. Models can incorporate the effects of different treatment regimens, building statistical models useful for prognosis and medical decision making [Beck & Pauker, 1983]. These models can then be applied in simulations. Diseases can be modeled across a wide range of fidelity, depending on the needs of the simulation and the availability of data.

In this simulation environment, treatments can be associated with disease and states. Treatments applied to a disease can alter transition probabilities, or as in the cases presented here, simply effect a state transition. Treatments are presented in the simulation as scaffolds that require the addition of resources, including both material components and labor, creating a resource-constrained environment.

2.2 Organizational Physics

The domain of organizational physics governs both the rules of operation within an MSU, including staffing decisions and medical decision making, and relationships between MSUs. Staffing decisions include minimal and maximal workload levels and decisions to add or remove staff in response to changes in demand. Medical decision making includes rules for prioritization in treatments, such as treating the most medically severe conditions first, or treating comorbidities in specific orders, as is the case when therapies for alcohol abuse must be initiated before treatments for liver disease are begun.

Relationships between MSUs govern policies that determine the transfer of resources, staff, and patients between facilities. A hospital represented by multiple MSUs may have a set of determinants for transferring patients between different specialized wards; a health care network has procedures for transferring patients between facilities depending on patient diagnosis, patient loads, and emergency conditions resulting from natural disasters. Facilities also have the ability to share resources between them, either as a result of policies designed to optimally distribute resources throughout the network, or in response to emergencies.

2.3 Diseases

Diseases are implemented using a general Markov framework with discrete states and probabilities of transitions between them. Markov models were proposed in a 1983

paper for use in medical prognosis due to their simplicity, ease of use, and fidelity in modeling disease progression [Beck & Pauker, 1983]. Since then, many diseases have been modeled using Markov techniques, including chronic obstructive pulmonary disease [Rutten-van Molken, 2006], multiple sclerosis [Confavreux & Wolfson, 1989], and HIV/AIDS [Farewell et al., 1997].

Diseases relevant to particular scenarios can be modeled to a relatively arbitrary level of fidelity, depending on data availability and the needs of the simulation. Using a Markov based approach, disease progression is decomposed into discrete states connected by state transition probabilities.

Each disease state can be associated with one or more effects, which can occur with a specified probability. Possible disease effects include changes to a patient's health status or the triggering of additional diseases, as in the case of hepatitis C progressing to hepatocellular carcinoma. Diseases may also occur as comorbidities, such as alcoholism co-occurring with hepatitis. Comorbidities can require distinct treatments, and organizational policies or medical necessity may require specific treatment orders, such as an acute, life threatening injury requiring treatment before a chronic condition.

2.4 Resources

Resources present the fundamental constraint under which the medical systems in this model operate. Resources are used in the instantiation of treatments; limited availability of resources at a given MSU can delay or prevent instantiation of a treatment.

Depending on the organizational configuration and MSU network topologies, resources such as medical supplies for acute traumatic injuries and medicine may be transferred between MSUs. Other resources, such as beds, are fixed properties of a given MSU. The organizational configuration may imply additional constraints on resource exchanges, such as limitations to the amounts of resources being shared, or a temporal delay in resource exchange imposed by geography or institutional procedures.

In addition to transferability, resources may be divided into those which are consumed after use, such as medicines, and those that become available again, such as beds. Consumable resources require storage and a constant resupply, while reusable resources act as a constraint on maximum load at a given time.

Labor is a special case of resource. Labor, including specialized labor, is available as produced by staff members. Labor production by staff members incurs individually based load, with load levels measured according to nominal and maximal constraints. Depending on organizational policies in the MSU, loads exceeding nominal can result in the addition of staff members to the available pool. Work levels exceeding the maximal level result in bottlenecking – the unavailability of the required labor resources.

2.5 Medical Service Units

Medical service units (MSUs) are the core component for the provisioning of medical services. MSUs bring together medical staff and patients with a supply of available resources and organizational rules.

Heterogeneity in medical facilities composing a network can be realized in the number of beds in each facility, number and types of staff present and available treatments. Large, metropolitan area hospitals can have upwards of several hundred beds, while smaller local clinics can have less than twenty. Staffing levels and composition, and availability of treatments, can vary similarly.

MSUs can be networked together to model modular health care facilities, such as individual hospitals with separate facilities for emergency care, cancer treatment, cardiac care, and surgical facilities. Similarly, hospitals can be networked to model integrated health care systems facilitating the transfer of patients and resources between them. This can be extended to networks of networks, such as the relationships between VA facilities, Department of Defense facilities, and private health services. Transfers between components can be constrained according to organizational rules, such as patients requiring particular types of care, or preferential sharing of resources within an organization but with restrictions or limitations outside of the organization.

3 Simulations

The scenarios considered here examine the behavior of a network of MSUs in response to perturbations described by different families of functions. The designed and evolved development of systemic robustness is usually conditioned on perturbations falling within anticipated or historically experienced bounds; perturbations that fall outside these bounds can cause the system to experience fragility, potentially resulting in a larger than average loss [Carlson & Doyle, 2002]. Overcrowding in medical services is correlated with increased risk of mortality, longer wait times for treatment, and higher probability of patients leaving the facility against medical advice or without being seen [Bernstein et al., 2009].

3.1 Disaster Scenario

The first series of simulations performed considered the effects of a single, large earthquake-based perturbation affecting a geographically concentrated portion of a health care network. Dynamics of interest included how normal operational levels of utilization of facility resources affected the ability to respond to large scale emergencies, and how the level of interoperability between facilities affected network-wide metrics.

This simulation uses a network of five MSUs conceptually arranged in three concentric rings. Pre-perturbation load levels and resource exchange policies are varied between runs.

For purposes of simplicity, these scenarios consider the interactions between two diseases. The first is an acute traumatic injury which drains patient health and

requires immediate treatment. The second is a chronic condition, chronic obstructive pulmonary disease (COPD), which requires maintenance treatments to reduce the risk of progression.

Both disease models are simplifications, intended to exercise system functionality and illustrate tradeoffs resulting from treatment prioritization in overloaded medical systems. The COPD model extends an existing four state COPD model [Rutten-van Molken, 2006] with additional sub-states to simulate within-state progression as a precursor to between state progressions.

Disaster medicine researchers have reported the phenomena of a “Double Spike” in patient surges in response to a major event [Miller & Arquilla, 2008]. The double spike occurs when treatment for patients with chronic diseases are delayed to enable personnel to treat disaster victims. The double spike is characterized by the expected spike in numbers of patients with acute trauma, followed by a broad surge of chronic patients who return to the facility sicker and in greater numbers than expected from the baseline load of chronic disease patients.

Figure 1 shows the baseline load of chronic COPD patients in a modeled facility. About 18 stage 1 patients and one stage 2 patient are admitted at any given time for treatment. Existing resources are sufficient to treat the baseline chronic patient load.

Figure 2 shows the surge in acute trauma patient population caused by the earthquake perturbation. Approximately 15 Stage 1, eight Stage 2, and two Stage 3 trauma patients occupy beds and personnel during the event which lasts from Time=25 to Time=65. The patient load illustrated in Figure 1 represents the first spike of the double spike phenomenon.

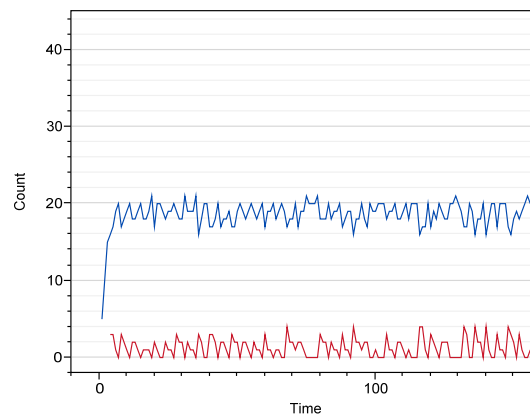


Figure 1: Baseline count of COPD Patients. Blue trace represents Stage1 COPD. Red TRACE represents Stage 2 COPD.

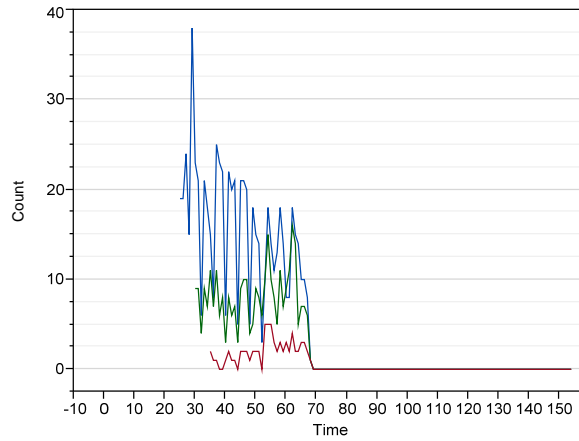


Figure 2: Disaster Scenario count of Trauma Patients. Blue trace represents Stage 1 trauma. Green trace represents Stage 2 trauma. Red trace represents Stage 3 trauma. Disaster perturbation occurs between Time=25 and Time=65

The population of COPD patients in the facility during and after the earthquake perturbation is shown in Figure 3. The graph shows that the hospital population of COPD patients falls precipitously at the start of the perturbation and remains low until all trauma patients have been treated and released. From Time=70 until the end of the simulation, the facility is dealing with the combined effect of the background levels of COPD patients seen in Figure 1, plus the COPD patients who had not been treated during the earthquake trauma treatment phase. This represents the second spike of the double spike phenomenon. In addition to the doubling of COPD patient counts over the baseline expectation, the chronic patients present in a more severe disease state as indicated by the large number of stage 2 trauma patients that show up at Time=70. Since these patients were not treated during the earthquake response, their conditions advanced to a more severe state. In summary, the second spike is characterized by a larger numbers of chronic disease patients who are sicker than would be expected from baseline experience.

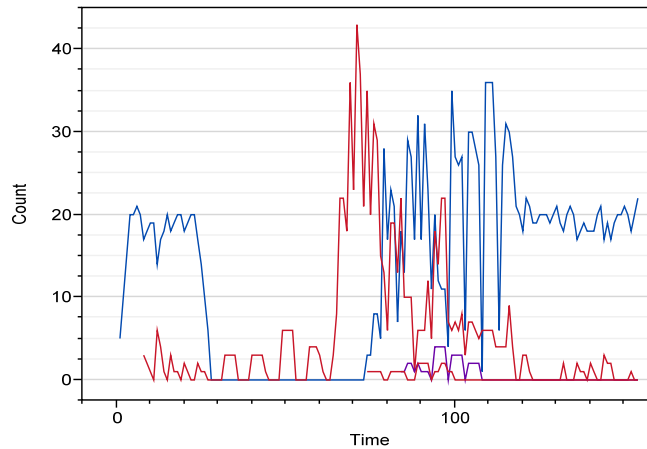


Figure 3: Disaster Scenario count of COPD Patients. Blue trace represents Stage1 COPD. Red TRACE represents Stage 2 COPD Disaster perturbation occurs between Time=25 and Time=65.

3.2 Long Term Resource Allocation Scenario

The second series of simulations considered the effects of a long term, prolonged increase in disease incidence in a network of health care centers.

Hepatitis C is a chronic liver condition resulting from infection with the Hepatitis C virus, frequently contracted via intravenous drug use or blood transfusions prior to screening practices adopted in 1990 [Remis, 2007].

The surge simulation assesses the effects of a constrained resource on patient outcomes. Hepatologists are medical doctors specializing in liver diseases such as hepatitis C. Often, healthcare systems such as the VA find it difficult to recruit and retain hepatologists. The projected increases in hepatitis C in the aging veteran population suggests that hepatologist availability may adversely affect hepatitis C treatment in years to come. The simulation examines the consequences of four different policies for deploying hepatologists within a networked healthcare system on three patient cohort groups. Patient cohorts with median ages of 50, 60 and 70 years living in a community seek treatment for liver diseases with wide a range of severity. Patients' health outcomes are tracked for different alternative modes of delivery of the constrained resource, hepatologists

Four policies govern hepatologist availability to the community:

1. Central Staffing: Specialist is only available at the central medical facility and not in outlying areas
2. Telehealth: Specialist expertise is delivered on demand to outlying clinics through electronic means such as video conferencing and telemetry
3. Circuit Rider: Traveling specialists visit local clinics on a routine basis

4. Local Presence: A specialist is assigned fulltime to each local clinic

The simulation is configured with patients going to medical facilities for treatment. The healthcare network is modeled as a 100-bed central facility and five 20-bed outlying clinics. Patients arrive at each treatment facility at regular intervals. A patient's disease state progresses until he is treated; patients who wait for treatment get sicker. If sufficient resources exist at the facility to treat a patient, the treatment is scheduled and delivered, and the patient is discharged. If resources are not available, the patient may wait until they become available or transfer to another facility where the treatment can take place.

Treatment delivery and outcome for patients vary between age cohorts (Figure 4). Since liver disease is progressive, older cohorts have more severe forms of disease which require the services of the hepatologist specialist. The less severe diseases exhibited by the younger cohorts can be treated locally with generalist medical personnel. Figure 4 shows that younger patients are treated and discharged at higher rates than older cohorts. Conversely, many older patients with more advanced disease must wait longer for treatment, or be transferred to the central facility where specialized care is available. Since the older patients are typically sicker, more of them die in treatment or awaiting treatment.

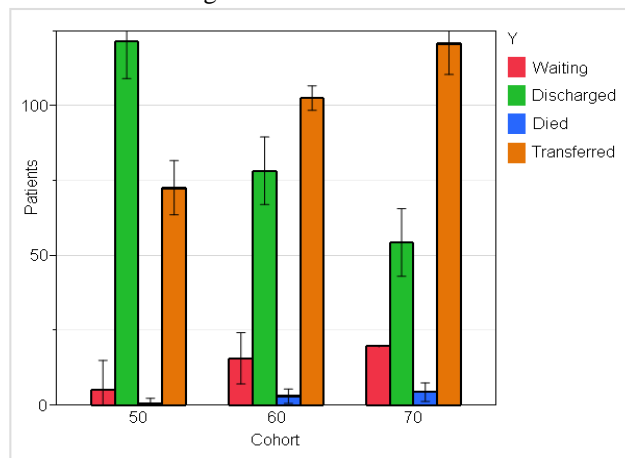


Figure 4: Disposition of 200-patient cohorts from an outlying clinic under Central Supply resource allocation policy. Error bars represent one standard deviation from mean of 25 model runs.

Figure 5 shows the how specialist allocation policy affects the number of patients who are treated locally at an outlying clinic. Individual bars on the chart represent the number of each 200-patient cohort which is treated locally, and not transferred for treatment. For each resource allocation policy fewer older patients can be treated locally. This reflects the more advance stage and severity of disease exhibited by the older patients. Across all age groups, the fewest number of patients were treated locally under the Central Staffing policy, followed by the Circuit Rider policy and the Telehealth policy. Local provisioning of specialists results in more

seriously ill patients being treated locally, and thus the highest overall levels of local treatment. Interestingly, Telehealth results in nearly as many patients being able to be treated at the community clinic as does Local Staffing.

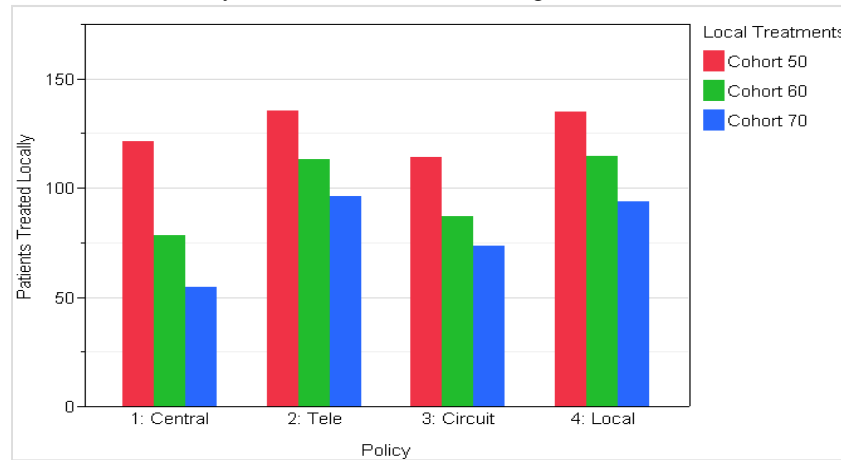


Figure 5: Patients treated locally at local clinic by policy. Each bar represents the median value of 25 model runs.

A corresponding pattern of emerges for the number of patients transferred to the central facility for treatment (Figure 6). For each treatment option, more of the sicker older cohort is transferred from the local clinic. Among resource allocation policies, Telehealth and Local Staffing are the most effective at reducing transferred, while Central Staffing and Circuit Rider policies require more transfers.

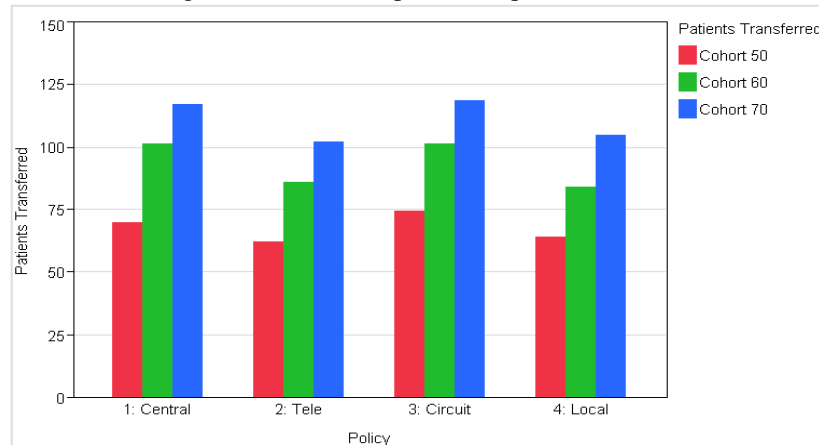


Figure 6: Patients transferred to central facility for treatment. Bars represent median counts from 25 simulations for each category

A third disposition of patients in the model is to be stuck in the queue awaiting the necessary resources for treatment or transfer (Figure 7). Both Telehealth and Local Staffing can provide sufficient local resources that the patient transfer process does not produce backups requiring patients to wait for extended periods. Central Staffing and Circuit Rider policies result in long patient wait times. Interestingly, the waiting profiles among the different age cohorts are different between Central Staffing and Circuit Rider policies. The waiting queue for the Central Staffing policy is biased toward older patients. This presumably reflects the increased severity of illness experienced by older patients, combined with the need to transfer all high-acuity patients to the central facility for advanced treatment. We propose that having a part time specialist in residence in local clinics allows him to treat a large number of the sicker older patients in the Circuit Rider scenario, reducing the number of older patients being queued for treatment or transfer.

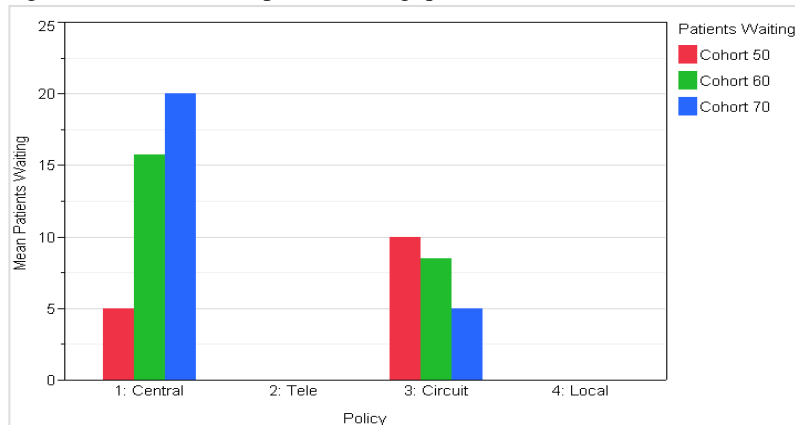


Figure 7: Number of patients in waiting queue by policy. Bars represent median patient counts from 25 simulations

Finally, patients with advanced liver illness may die while awaiting treatment or in transit. Figure 8 shows that while such deaths are rare, different resource allocation policies exert a control over them. The Local Staffing policy results in the fewest mean deaths, with $< 0.5\%$ deaths, and those being restricted to the oldest cohort. Next most effective policy for preventing deaths is Telehealth, which has nearly the same number of deaths, but spread over more age groups. Circuit Rider policy averages less than 2% deaths, weighted toward the older cohorts. Central Staffing policy exhibits the most deaths due to its extended waiting queues and transport requirements, with about 4% dying across all cohorts.

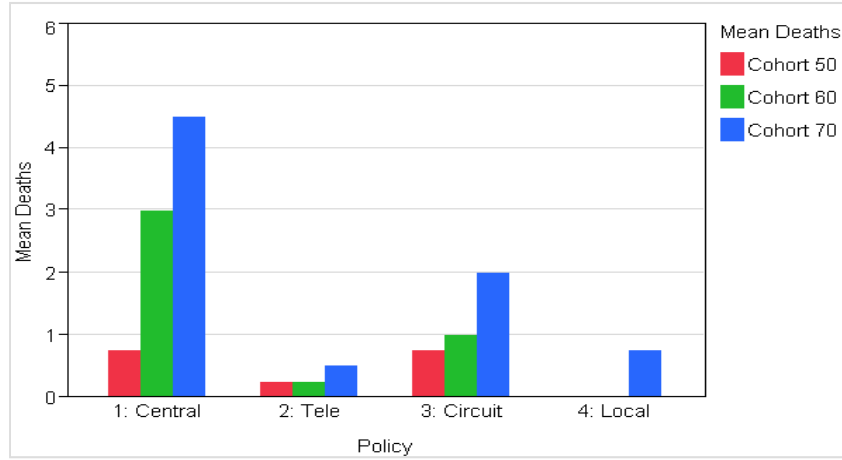


Figure 8: Number of patients that died awaiting treatment. Individual bars represent the mean number of patient deaths for each category from 25 simulations

This simulation suggests that patient outcomes are better when resources such as scarce specialists are deployed close to the patients, that is, in the local clinics. While not surprising, Local Staffing policy proved superior to Central staffing for all modeled metrics including transfers, time waiting, and deaths. Of the intermediate policies, Telehealth outperformed Circuit Rider. The effectiveness of the fractional hepatologist presence under the Telehealth policy suggests that a locally deployed specialist might be underutilized, and thus be a waste of a critical scarce resource.

4 Conclusions

Health care facilities that regularly operate near or at capacity levels will potentially have difficulties in addressing changes in demand for health services. Hospitals that reduce capacity as a cost savings measure may have difficulties responding to unexpected increases in demand or loss of capacity, even under the best of circumstances. The less spare capacity an individual unit has, the greater the difficulty it will have in addressing changes in patient loads.

The form of the perturbation can be as important as its size – it is possible to design a system that exhibits robustness to spikes in demand by temporarily relocating resources or transferring patients, operating under the assumption that normal operations can be resumed and that backlogged work can be processed after the emergency ends. Long term surges in demand, on the other hand, require alternative resource distribution schemes, including both the integration of technological innovations to better distribute resources throughout an integrated health care system, and an expansion in general capacities.

Some of these fragilities can be made more robust if the network is viewed as the target of optimization, rather than the individual facility. If the spare capacities of

all individual institutions are made globally available, systemic capacity can scale to meet demands more efficiently and under a greater range of perturbations. This can happen if organizations providing medical services are sufficiently integrated, such that capacities can be easily shared and selection can act at the appropriate network level. Integration of services, with an organizational framework that provides for rule enforcement and suppressing within group competition, makes this focus of selection possible.

Bibliography

- [1] Beck, J. R., & Pauker, S. G. (1983). The Markov Process in Medical Prognosis. *Medical Decision Making*, 3(4), 419-458.
- [2] Bernstein, S. L., Aronsky, D., Duseja, R., Epstein, S., Handel, D., Hwang, U., McCarthy, M., et al. (2009). The Effect of Emergency Department Crowding on Clinically Oriented Outcomes. *Academic Emergency Medicine*, 16(1), 1-10. doi:10.1111/j.1553-2712.2008.00295.x
- [3] Carlson, J. M., & Doyle, J. (2002). Complexity and robustness. *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl 1), 2538.
- [4] Carlson, J. M., & Doyle, J. (1998). Highly optimized tolerance: Robustness and power laws in complex systems. *Arxiv preprint cond-mat/9812127*.
- [5] Confavreux, C., & Wolfson, C. (1989). Mathematical models and individualized outcome estimates in multiple sclerosis. *Biomedicine & Pharmacotherapy*, 43(9), 675-680.
- [6] Farewell, V. T., De Angelis, D., Day, N. E., Gill, O. N., & Aalen. (1997). A Markov model for HIV disease progression including the effect of HIV diagnosis and treatment: application to AIDS prediction in England and Wales. *Statistics in medicine*, 16(19), 2191-2210.
- [7] Kilduff, M., & Tsai, W. (2003). *Social Networks and Organizations*. Sage Publications Ltd.
- [8] Michod, R. (1999). *Darwinian Dynamics*. Princeton University Press.
- [9] Miller, A. C., & Arquilla, B. (2008). Chronic diseases and natural hazards: impact of disasters on diabetic, renal, and cardiac patients. *Prehospital and Disaster Medicine*, 23(2), 185.
- [10] Muir, W. (2003). Indirect Selection for Improvement of Animal Well-being. *Poultry Genetics, Breeding and Biotechnology* (pp. 247-255). CABI Publishing.
- [11] Okasha, S. (2006). *Evolution and the Levels of Selection*. Oxford University Press.

[12] Remis, R. (2007). Modeling The Incidence and Prevalance of Hepatitis C Infection and its Sequelae In Canada 2007: Final Report. Community Acquired Infections Division Centre for Communicable Diseases and Infection Control Infectious Disease and Emergency Preparedness Branch Public Health Agency of Canada.

[13] Rutten-van Molken, M. (2006). Economic Modeling in Chronic Obstructive Pulmonary Disease. *Proceedings of the American Thoracic Society*, 3(7), 630-634. doi:10.1513/pats.200603-095SS

[14] Stevenson, W. B. (1990). Formal structure and networks of interaction within organizations. *Social Science Research*, 19(2), 113–131.