1 Introduction

The health sector accounts for a significant portion of gross domestic product (GDP) around the globe, particularly in developed countries. According to the Organization for Economic Cooperation and Development (OECD)’s Health Statistics (2015), the U.S. spent 16.4% of its GDP in 2013 (excluding capital expenditure) in the healthcare sector, while the spending average of the thirty-four OECD countries in the same sector was 8.9%. This has remained unchanged since 2009 as health-spending growth matched economic growth. OECD estimates the per capita health spending in the U.S. was $8,713 in 2013. In Canada, where the per capita health spending is about half of that of the United States’, hospitals account for 30% of the overall costs. Hospitals are then followed by drugs and physicians, each with about a 15% share (Canadian Institute for Health Information (CIHI) 2015). As a result of hospital consolidation and bed closures as well as sustained transition from inpatient to outpatient care, there has been a significant reduction in hospital spending since the ’90s—now stabilized at its current level. This suggests that there may not be any more “low-hanging fruit” for further cutting hospital costs via macro-level policies. Although continued efforts at the governmental level for improving the health sector are still necessary, they are not sufficient for achieving tangible success on the ground. To increase the efficiency of care delivery processes and the quality of care, healthcare providers and decision makers must implement health policies through concerted and sustained efforts at the operational level. Healthcare processes are dynamic and complex systems. Their design and improvement often involve pushing the frontiers of research in engineering and management.

There is a consensus among the stakeholders concerning the overarching objectives of healthcare: health outcomes, accessibility, quality, and cost of care. Although the plethora of stakeholders may have different individual prioritizations of these objectives, no one would find a health system that performs worse than a certain threshold in any of these attributes. There is, however, a widespread disagreement as to how the goals mentioned above can be attained, since the preferred path to these targets depends on the organization of the health system in a country or state (i.e., the extent of private/public ownership, the insurance and reimbursement systems in effect, single/multiple payers, and the various access barriers to healthcare).

For the purposes of this chapter, it is important to also highlight the difference between the care providers’ and POM experts’ viewpoints. The physicians and nurses are predominantly...
trained in a disease-oriented fashion, which constitutes their mindset for providing the best care to the patients. On the other hand, the POM experts—who may be the administrators or the consultants to the administrators—are trained in a process-oriented fashion that tends to downplay the differences with regard to the disease condition of the patients in an effort to come up with generalizable managerial insights. It is evident from our and other scholars’ experiences that POM can make a positive impact for the care providers, patients, and administrators to the extent that the attention to health outcomes and process performance are balanced and the tradeoff between generalizability and individuality in patient care is treated carefully at the operational level (Ibrahim et al. 2016; Hulshof et al. 2012).

From a process perspective, the journey of each individual through the healthcare continuum involves one or more care phases, each with differentiating characteristics. A non-exhaustive list of these phases includes preventive care, primary care, pre-hospital (i.e., ambulance) care, acute (i.e., hospital) care, rehabilitative care, community-based care, chronic care, home care, long-term care, and palliative care. We will discuss the features of three of these phases (i.e., preventive, primary, and emergency care) and their implications for process design and improvement in the remainder of this chapter. At the outset, however, we must point out the significance of the interaction among these phases.

For example, the lack of access to primary care physicians increases the burden on the emergency departments (EDs), since EDs are the only facilities where the patients have a legal right to access healthcare. Recently, Sarnak and Ryan (2016) found out that about one of five patients (19% of all patients) aged 65 and older with at least three chronic conditions in the U.S. and Canada reported an avoidable ED visit that could have been treated by a family physician. Interestingly, the causes for this lack of access are quite different in the U.S. and Canada. In the U.S., the major challenge is formed by financial barriers due to the unaffordability of insurance, whereas in Canada the limited availability of primary care resources is a fundamental access barrier. Nevertheless, the overall impact of a problem at the upstream primary care phase on the downstream emergency care is quite similar from a patient-centric viewpoint.

In this chapter, we will focus on some of the process related challenges currently faced by the health sector and demonstrate how POM can be helpful. Notwithstanding the fact that healthcare systems vary around the world, POM methodology takes the macro system structure given and aims at supporting the strategic, tactical, and operational decisions concerning healthcare delivery. The lack of collaboration across the healthcare continuum is a significant impediment to patient care. A number of large-scale organizations such as Kaiser Permanente, the Cleveland Clinic, and the Mayo Clinic in the U.S. and King’s Health Partner’s in the U.K. have made great strides toward providing integrated care to their patients under one roof. Nonetheless, there remains a lot more work to be done for the public at large to benefit from such initiatives. In Quebec, for example, Bill 10 has reorganized the healthcare landscape in order to improve integration as recently as April 2015. So far, the implementation outcomes have yet to be seen.

There is a rich series of POM literature (Hulshof et al. 2012) offering a number of methods for designing integrated health networks as well as for improving the effectiveness at the implementation stage. Another major challenge is the disentanglement of the internal versus external determinants of performance at each phase of the care continuum. POM methodology can be also effective in identifying the most promising interventions through internal factors that are under the jurisdiction of the managers at each phase, whereas having a quantitative handle on the impact of external factors on their own system (e.g., the impact of lack of access to primary care on ED, as discussed above) would increase their ability to convince their superiors in releasing some of the burden.
This chapter is not intended to be a comprehensive literature review on healthcare operations management. There are a number of significant challenges in this domain that are out of the scope of the chapter.

POM can make a positive contribution concerning these issues as well. For example, the misalignment of the incentives provided to the care providers and the overall system objectives constitutes a significant challenge for the managers. It is well established that whether the physicians are reimbursed on a “fee for service” or on a “capitation” basis influences the type and volume of patients they serve. Supply chain coordination literature presents alternative ways to align the incentives through the healthcare service chain as well as for the different stakeholders. We include only the issues directly related to the configuration of the care delivery processes in this chapter and provide a narrative review that aims at demonstrating the potential value of POM. A comprehensive literature review on healthcare operations management would require multiples of the space allocated to a single book chapter. To make this point and to demonstrate the increasing scholarly activity in this domain, Figure 23.1 depicts the proliferation of the number of papers published in the most relevant journals during a twenty-five-year time horizon.

The potential value of POM for process design and improvement at two of the upstream layers of the healthcare continuum constitutes the focus of this chapter. This is due to the attention preventive care and emergency care received from POM scholars. This attention has led to the development of a rather solid methodology and application base. In contrast, the POM literature on primary care is fledgling, despite the significance of this upstream phase—something we discuss briefly in the last section. We provide a detailed discussion of the prevailing POM methodology
for preventive care in Section 2. Section 3 presents the current challenges in emergency care, discusses a case study, and directs the reader to the core papers within the abundance of literature in this domain. Section 4 discusses the implications for managers, and Section 5 concludes the paper with an outlook for the needed developments in POM methodology concerning the upstream phases of the healthcare continuum.

2 POM for Preventive Care

We start by providing a fundamental understanding of the preventive care processes. We continue with a basic formulation in Section 2.2, and discuss its extensions as well as other modeling frameworks in Section 2.3

2.1 Preventive Care Processes

Preventive care is based on the premise that it is easier and more humane to prevent health conditions, such as many forms of cancer, rather than to treat them. With preventive care, it is also more likely that the patient will fully recover as a result of early diagnosis. Thus, there are three types of preventive care programs: (i) primary prevention reduces the likelihood of diseases in people with no symptoms, such as flu shots; (ii) secondary prevention identifies and treats people with risk factors or are at very early stage of diseases, such as colonoscopy to detect early forms of colon cancer; and (iii) tertiary prevention treats symptomatic patients in an effort to decrease complications or severity of disease, such as sugar control in a diabetic in order to mitigate vision and nerve problems.

The differentiating characteristics of preventive care are as follows: (i) it is a user-choice environment in terms of the allocation of clients to facilities, where the regulators typically do not control the patronage of the facilities; (ii) the clients’ willingness to participate in preventive programs could decrease significantly when either access is inconvenient or there is congestion at the facilities leading to unreasonable wait times; and (iii) each facility needs to capture a certain volume of clientele to satisfy a minimum workload requirement that ensures the quality of the service.

The substantial savings in the costs of diagnosis and therapy as well as the relatively lower capital investment associated with preventive care programs have been recognized for a long time (Walker 1977). Although it is well established that preventive care programs can save lives and contribute to a better quality of life by reducing the needs for radical treatments, they do not constitute a priority for very early stage of diseases, such as colonoscopy to detect early forms of colon cancer; and (iii) tertiary prevention treats symptomatic patients in an effort to decrease complications or severity of disease, such as sugar control in a diabetic in order to mitigate vision and nerve problems.

Zimmerman (1997) found through a survey that the convenience of access to the facility was a very important factor in a client’s decision to have prostate cancer screening. Facione (1999) revealed that the perceptions of lack of access to services were related to the decrease of mammography participation.

From a POM perspective, the underlying problem involves the design of a preventive care facility network in a way that would maximize the level of participation. The number of facilities to be established as well as the location, capacity, and service mix of each facility constitute the primary structural decisions. Of course, there are also infrastructural decisions to be made in this...
context, particularly concerning the skillset of the care providers at each facility. The prevailing POM literature uses the time spent to receive preventive services as a proxy for accessibility of healthcare facilities (Zhang et al. 2012). This includes the time spent in transportation to the facility and the total time spent at the facility (i.e., waiting and care).

2.2 A Basic Formulation for Designing Preventive Care Networks

We will use the model by Zhang et al. (2009), where each facility is represented by an $M/M/1$ queue, as the basis for illustration. To formulate the problem, they used the following notation:

- $y_j = 1$ if a facility is located at demand node $j$, 0 otherwise
- $x_{ij} = 1$ if clients from node $i$ require service from facility $j$, 0 otherwise
- $N = \text{set of demand nodes, indexed by } i$
- $X = \text{set of alternative facility sites, indexed by } j$
- $\lambda = \text{overall rate of demand for the preventive service (assuming Poisson distribution)}$
- $1/\mu = \text{average service time (assuming exponential distribution)}$
- $h_i = \text{fraction of clients residing at node } i$
- $a_{ij} = \text{fraction of clients from node } i \text{ who would patronize facility } j$
- $R_{\text{min}} = \text{minimum workload at each operational preventive care facility}$

The implicit assumption is that all individuals from the same node request service from the same facility. Also, in the long run, the clients will gather sufficient information about the total time required to obtain preventive healthcare services at the facilities in their vicinity, and hence, each user will choose the facility by which they would receive the service the fastest. To this end, Zhang et al. (2009) denote the travel time $t_{ij}$ and the expected waiting time for the service $W^j_i$. Assuming a linearly decreasing participation function, $a_{ij} = A_{ij} - \gamma (t_{ij} + W^j_i)$, where $W^j_i = 1/(\mu - \lambda \sum_{i \in N, j \in X} h_i x_{ij})$ and $A_{ij}$ and $\gamma$ are the intercept and slope of the participation function, respectively. The resulting nonlinear binary formulation is as follows:

Maximize $\lambda \sum_{i \in N} h_i \sum_{j \in X} a_{ij} x_{ij}$ \hspace{1cm} (1)

Subject to

$$\sum_{j \in X} x_{ij} = 1 \hspace{1cm} i \in N$$ \hspace{1cm} (2)

$$x_{ij} \leq y_j \hspace{1cm} i \in N, j \in X$$ \hspace{1cm} (3)

$$R_{\text{min}} y_j \leq \lambda \sum_{i \in N} h_i a_{ij} x_{ij} \leq \mu \hspace{1cm} j \in X$$ \hspace{1cm} (4)

$$x_{ij}(t_{ij} + W^j_i) t_{ik} \leq W^k_i + M(1-y_k) \hspace{1cm} i \in N, j, k \in X$$ \hspace{1cm} (5)

$$a_{ij} x_{ij} \geq 0 \hspace{1cm} i \in N, j \in X$$ \hspace{1cm} (6)

$$x_{ij}, y_j = 0,1 \hspace{1cm} i \in N, j \in X$$ \hspace{1cm} (7)

The objective function (1) maximizes the total number of people expected to participate in the preventive care program. Constraints (2) impose the single-facility service requirements on each demand node. Constraints (3) ensure that clients can require service only from open
facilities. Constraints (4) guarantee the stability of the queue and impose minimum workload on the open facilities. Constraints (5), where $M$ represents a big number, are the closest facility assignment constraints in terms of minimum total time to receive service, whereas constraints (6) prohibit negative values for participation.

The above model was used for determining which breast cancer screening centers in Montreal, Canada, should be included in the Health Ministry’s program to subsidize mammograms for women between the ages of 50 and 69 (originally discussed in Verter and Lapierre 2002). Zhang et al. (2009) demonstrated that the incorporation of congestion would considerably increase the accuracy of the estimates for the participation level to the program. They also showed that centralizing system capacity at the locations preferred by the clients would be a better policy than pushing for decentralization by accrediting a larger number of small facilities. Their methodology can be used in making decisions concerning the total system capacity as well as the investments in raising public awareness with regard to preventive healthcare programs. That is, the allocation of some resources on health promotion projects targeted to potential clients can be considered as an effective means of increasing participation. This should, of course, be considered as a complementary to the optimization of the facility network configuration.

In a follow-up paper, Zhang et al. (2010) extend the basic model by also optimizing the capacity (i.e., the number of servers) at each facility and allowing for the people from the same population zone to receive preventive care at different facilities, as long as the total (travel + wait + service) times required for receiving care are the same. They developed a bi-level formulation, where the lower level represents the user choices, whereas the upper level determines the structural decisions pertaining to the preventive care facility network. The lower level aims at reaching user equilibrium, where all clients are content with the facility choices. The lower-level problem is incorporated in the upper level via a variational inequality. Using their multi-server model on the Montreal case mentioned above, Zhang et al. (2010) were able to demonstrate that capacity pooling (i.e., centralizing the system capacity in a few large facilities) may raise participation by reducing the mean waiting time. Note that the minimum workload requirement also favors centralization, since it may not be feasible to accredit many single-server facilities. Interestingly, they also show that Zhang et al. (2009)’s findings using a single server model, pertaining to the value of the centralization strategy, hold true only when the number of servers is not too restricted. For the Montreal case, when there are less than ten mammography workstations available, however, rather than striving for capacity pooling by adding one server to an existing facility, it is better to increase the spatial coverage by locating a new facility in another high-density area.

### 2.3 Extended Models for Preventive Care

The stream of literature presented in the previous section has been extended in many ways. In particular, Aboolian et al. (2016) studies the problem of determining the overall capacity at each site rather than the number of servers. They reformulated the nonlinear model as a mixed-integer program by replacing the service rates with the waiting times that are approximated using tangent line approximation. By studying Toronto’s twenty-two-hospital network (Berman et al. 2007), they observed that the capability of the Ontario government to increase participation in its services by simply increasing accessibility is limited. They also showed that a gradual capacity expansion strategy could be robust as long as the system is originally conceived with an overall capacity that is above a threshold level. Zhang et al. (2012) incorporated probabilistic user choice to account for the facility attributes that attract patients other than accessibility. Using
the Montreal case, they demonstrated the similarity between the optimal-choice and probabilistic-choice models concerning the trade-off between capacity pooling and spatial coverage. The resulting network configurations obtained from the two models, however, are quite different. This emphasizes the significance of accurately representing the user choice in preventive care facility network design.

One of the challenges that is not explicitly taken into account by the above papers is that many screening facilities also provide diagnostic services. Therefore, the capacity of such facilities needs to be allocated between the urgent diagnostic needs of a group of patients and the preventive needs of the asymptomatic patients to be screened. In the context of colonoscopy procedures, Güneş et al. (2015) studied the impact of preventive care on reducing the future demand for diagnostic services. By integrating the disease progression and operations perspectives in a system dynamics model, they were able to assess alternative resource allocation policies. They showed that if the diagnostic service capacity could be set to keep the average waiting time for diagnosis below two weeks, then the remaining capacity can be dedicated to screening. Örmeçi et al. (2016) analyzed the long-term impact of preventive care, mentioned above, through a partially endogenous random environment. They explored the operational-level colonoscopy scheduling policies using a Markov decision process (MDP). Interestingly, Örmeçi et al. (2016) and Güneş et al. (2015) both confirm the current practice of prioritizing the diagnostic services, although under certain—albeit less often—circumstances patients with lower waiting costs (i.e., screening) need to be given priority.

In an effort to personalize the screening decisions, a stream of papers utilize the partially observable Markov decision process (POMDP) framework, where the true state of the cancer being screened for is not observable. Erenay et al. (2014) incorporated age, gender, and risk of having colorectal cancer in a POMDP to optimize colonoscopy screening policies so as to maximize the patient’s total quality-adjusted life years. Ayer et al. (2012) and Maillart et al. (2008) developed a POMDP for breast cancer screening, whereas Zhang J. et al. (2012) used the same framework for prostate cancer screening. All of these studies provide a favorable comparison between the proposed screening policy and the current clinical guidelines and highlight the effect of aging on the optimal screening policies. There is a long road for these personalized screening policies to be implemented by practitioners in a wide scale, since this would require their validation through extensive randomized control trials. Nonetheless, POM can certainly claim a significant contribution towards more effective and efficient design of the needed control trials as well as towards empowering the individual physicians (whose patient-centric decisions can override the guidelines) with more accurate knowledge concerning the potential benefits of personalized screening policies.

### 3 POM for Emergency Care

We first provide a brief account of the key process challenges emergency department (ED) managers face in providing timely and quality care to patients. The literature on the use of POM in the context of ED process design and improvement is vast. A recent comprehensive review based on 350 papers can be found in Saghafian et al. (2015). Due to the complexity of the ED processes, a vast majority of the prevailing studies resort to simulation for the purposes of modeling, analysis, and improvement. Thus, Section 3.2 is devoted to the simulation of ED process and a new case study at the ED of a large tertiary hospital is presented in Section 3.3.

The recent trend in using POM for emergency care involves the use of optimization methods by taking advantage of stylistic models. In particular, Mandelbaum et al. (2012) studied the patient admissions from an ED to the inpatient wards by means of a queuing model with a single
centralized queue and several server pools. They proposed a randomized most-idle routing policy for patient admissions. Under the proposed policy, a patient is assigned to one of the available wards, with probability that equals the fraction of available beds in that ward out of the total number of available beds in the system. Saghafian et al. (2012) studied the impact of streaming ED patients based on predictions concerning the disposition decision after receiving care (i.e., whether they will be discharged or admitted to the hospital). They underline the importance of sharing ED resources across streams and not earmarking them to the patient streams. Saghafian et al. (2014) showed that estimating the complexity of the required care during the ED triage could lead to considerable reductions in both the risk of adverse events and the average length of stay. These two papers also determine the conditions under which the virtual streaming and complexity-augmented triage policies would be effective.

3.1 Key Challenges in ED Management

The ED constitutes the point of entry to a hospital for all patients—except those who are scheduled for elective surgeries. According to Schuur and Venkatesh (2012), ED is the first point of contact for nearly half of all hospital admissions in the U.S. Nonetheless, emergency care is sufficient for the significant majority of the patients who present at the ED and eventually get discharged home. ED managers are faced with a multitude of challenges including overcrowding, resource reductions, limited bed capacity, long waiting times, and low staff morale. These issues are becoming more pronounced due to the increasing number of ED visits and never-ending budget cuts from the health sector at large.

The Canadian Association of Emergency Physicians (CAEP) (2013) defined overcrowding as “demand surpassing the ability of an ED to provide quality care within acceptable time frames”. This definition implies that ED overcrowding corresponds to an “access block”, to use POM terminology. The most significant impact of overcrowding is on the ability to deliver emergency care in a timely manner (Derlet and Richards 2000; Derlet et al. 2001). Overcrowding often leads to an increase in the number of interruptions the caregivers endure while caring for a patient. This makes them feel overextended and has the potential to cause adverse patient outcomes. Another important effect of overcrowding is patient morale and satisfaction. Delays in patient treatment may unnecessarily increase their time spent on stretchers in the ED. Patients who experience disheartening waiting times, are often forced to lie in crowded corridors which robs them of both privacy and a sense of dignity. Overcrowding has also led to an increase in patients leaving without being seen by medical personnel as well as ambulance diversions whereby critical patients may be denied access to the closest ED, and rerouted to other centers (Schull et al. 2003).

The optimization of staffing decisions is another key challenge for the ED managers, where the prevailing POM methodology can be very helpful. The visits of the ambulance and walk-in patients are unscheduled, whereas most EDs are staffed through a fixed schedule. The number of physicians and nurses is predetermined and does not fluctuate depending on patient volume. This often results in either the under or over staffing of units. Many EDs are thus left scrambling to meet the demands of a crowded and hectic unit. Staffing the ED with the appropriate type and number of personnel needed to meet these demands is essential. The ED managers must also take into consideration the efficient utilization of existing personnel. Non-medical tasks, often performed by highly trained physicians and nurses, are well known to decrease efficiency. Examples of such tasks include looking for charts, tracking down laboratory results, arranging for admission, or rewriting information. It is well known that these additional tasks cause
unnecessary disruptions in patient care and frequently result in increased job stress and risk for medical errors (Wears and Leape 1999).

### 3.2 Simulation of ED Processes

A process view of the patient flow through the ED is depicted in Figure 23.2. The first step for all patients (irrespective of the mode of arrival) is triage, where a triage nurse evaluates the patient’s medical condition to identify the need for an ED bed as well as the priority for seeing a physician. The next step is registration upon which patients wait in the waiting area until an ED bed (or an ambulatory treatment space, if appropriate) becomes available. ED treatment starts with the initial physician and nurse assessments. An ED physician may request one or several lab tests (i.e., blood/fluids or imaging), keep the patient under observation, or decide to consult a specialist. Such consultations often include a clinical examination of the patient by the specialist as well as additional lab tests. The last step is the disposition decision by which the patient is either admitted to the hospital by a specialist for further treatment or discharged home. Many of the admitted
patients wait in the ED for an inpatient bed, which is called “ED boarding” (i.e., an access block for the newly arriving ED patients).

There are number of process challenges concerning the flow of patients through an ED. First, the triage and assessment operations are based on priority queues: At triage, ambulance arrivals have priority over walk-in arrivals, and at the first assessment, patients are often prioritized according to their triage code. Second, it is entirely possible for a new patient arrival to pre-empt the current task an ED physician is performing. For example, in the event of the arrival of a patient requiring resuscitation, the physician will have to pause caring for the current patient, which may require restarting the entire task upon performing the resuscitation. Third, and perhaps most importantly, ED physicians depend on their specialist counterparts for consultation and, if necessary, hospital admission of the patients. In some hospitals, the consultation requests from ED may not constitute the top priority for specialists who need to also provide care at the inpatient wards as well as the outpatient clinics and (in surgical specialties) operate in the surgical theatres. In addition, the specialist who is sent to the ED may need to consult someone with more experience or different specialty. Last, but not least, many EDs do not have dedicated laboratories and imaging facilities to conduct the necessary blood/urine tests and diagnostic scans a patient may require. It is due to these differentiating characteristics of the ED that patient flow is a rather complex phenomenon, one where POM has ample opportunity to bring value to management.

Computer simulation lends itself naturally for modeling the ED processes with such complexity. A major advantage of simulation in modeling EDs, over analytical techniques (which we will discuss later in this section) is its ability to represent variability in individual patient characteristics. Note that patient visits to EDs are mostly unscheduled and present irregular peaks and troughs in the number of patients as well as in the acuity of their illness or injury. For tractability, analytical methods require the aggregation of patients into homogenous groups. This thereby ignores the individual differences within each group. In our experience, such stylized representations seem to hinder the usefulness of analytical techniques in detailed operational modeling of EDs; at least for the decision makers with a medical background.

As addressed by Sinreich and Marmor (2005), an important aspect of ED simulation is the structure and collection of the necessary data as well as its level of detail. Saunders et al. (1989) is one of the earliest ED simulation models to reflect the underlying complexity. They built a rather sophisticated model (relative to the state of technology at the time) that assigned an individual nurse to each patient; incorporated the patients’ movements on several pathways simultaneously; and represented tests, treatments, and procedures in detail. They used sensitivity analysis in studying the factors that determine the patient’s waiting time and used severity as the only patient characteristic. In contrast, McGuire et al. (1994) categorized patients on both severity and broad category of diseases. McGuire used simulation to test several operating alternatives in a Sun Health Alliance Hospital, with the objective of reducing patient length of stay (LOS). The most comprehensive framework, to date, is developed by an Israeli team of scholars, who made significant strides toward simulation-based real-time decision-support systems for EDs (Zeltyn et al. 2011).

### 3.3 A Case Study in ED Triage

In order to demonstrate the potential value of simulation, we report on the triage module of a comprehensive simulation platform that was developed by our multi-disciplinary research team for the ED of a tertiary hospital in Montreal, Canada (Verter et al. 2012). This ED has about 66,000 patient visits annually. Patient triage, performed by a registered nurse (RN) immediately
upon arrival, aims to rapidly identify patients with urgent (possibly life-threatening) conditions and to determine the most appropriate treatment area in ED. Both the U.S. and Canada use five-level triage codes, where a lower code indicates a higher level of urgency. The Canadian triage scores are based on the target time the patient needs to be first seen by a physician, whereas the U.S. emergency severity index is associated with the expected resource utilization of the patient in the ED.

For the purposes of this project, the triage was observed over a 15-week period during weekday shifts (8:00 to 16:00) for an average of 8 hours/day. During this period, 537 ambulance and 3,205 walk-in patients were observed. We also extracted data from the ED administrative database: socio-demographic, patient arrival patterns, and triage severity. During the data collection period, the one full-time RN and a second RN available for about 5 hours throughout the day were assigned to triage. Figure 23.3 depicts the observational data concerning the distribution of arrivals, service times, and wait times depending on the mode of arrival.

The triage module was built using the ARENA software (www.arenasimulation.com), and the simulation model was validated through comparing the simulated wait time with those depicted in Figure 23.3. Two remarks are warranted here: (i) at the time of data collection, the triage performance was far below the Canadian Triage Acuity Standards (CTAS), which stipulate that 95% of the code II patients should be seen by a physician within 15 minutes of arrival, and the first physician assessment for 90% of code III patients should occur within half an hour; and (ii) the bimodal shape of the triage wait times for the walk-in patients (i.e., with a kink around two hours) is due to the multiple ambulance arrivals within short time windows.

As noted earlier, our baseline involves 1.5 RNs, each triaging both ambulance and walk-in patients, where the former group has priority. The ED chief was interested in increasing the triage staffing to 2 RNs and wanted to know the impact of the following two possible interventions: (i) dedicating one RN to ambulance patients and one RN to walk-ins (i.e., un-pooling) and (ii) performing a 1-minute pre-triage screening to fast track eligible patients to a rapid assessment zone. Table 23.1 shows the mean and standard deviation of the triage wait times as well as the nurse utilization levels for the four scenarios we have studied.

The benefits of capacity pooling are well established in the general POM literature (Cachon and Terwiesch 2012). Although we knew the direction of change when the RNs are dedicated, the simulation model informed the ED chief as to the level of reduction in the wait times for the ambulance patients and the level of associated increase in the mean and variability of the wait times to be experienced by the walk-in patients. Clearly, pre-triaging all arrivals is beneficial for both patient groups. Although combining pooling with pre-triage (i.e., the fourth scenario) brings the triage wait times to levels that would enable the ED to achieve the CTAS standards, the expected nurse utilization of 39% constituted a sticking point in an environment plagued with limited resources.

After a lengthy discussion period, the ED management was convinced to keep the nurses pooled and to implement the pre-triage as well as the rapid assessment zone ideas. By depicting how the problematic right-hand tale of the wait time distributions would be curtailed as a result of these interventions, Figure 23.4 was instrumental in achieving the administrators’ buy-in.

We also studied the impact of a dynamic nurse staffing policy for ED triage, where the second RN is called in only when the triage waiting line reaches a predetermined threshold level. Although this strategy further reduces the wait times (albeit not significantly) while keeping the triage nurse utilization at high levels, it was not acceptable to the ED chief due to the perceived complexity of implementation.
Figure 23.3 Triage Data: Patient Arrivals, Service Times, and Wait Times

<table>
<thead>
<tr>
<th>Time period</th>
<th>Walk-ins</th>
<th>Ambulances</th>
</tr>
</thead>
<tbody>
<tr>
<td>7:00–8:00</td>
<td>30</td>
<td>60</td>
</tr>
<tr>
<td>8:00–9:00</td>
<td>36</td>
<td>90</td>
</tr>
<tr>
<td>9:00–10:00</td>
<td>18</td>
<td>120</td>
</tr>
<tr>
<td>10:00–11:00</td>
<td>18</td>
<td>150</td>
</tr>
<tr>
<td>11:00–12:00</td>
<td>12</td>
<td>180</td>
</tr>
<tr>
<td>12:00–13:00</td>
<td>9</td>
<td>180</td>
</tr>
<tr>
<td>13:00–14:00</td>
<td>1</td>
<td>180</td>
</tr>
<tr>
<td>14:00–15:00</td>
<td>5</td>
<td>180</td>
</tr>
<tr>
<td>15:00–16:00</td>
<td>18</td>
<td>180</td>
</tr>
</tbody>
</table>

Average Number of Arrivals per Hour

Number of arrivals

### Table 23.1 Scenario Analysis for Triage Process Design

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Triage Wait Time (Min.)</th>
<th>Nurse Utilization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ambulance</td>
<td>Walk-in</td>
</tr>
<tr>
<td>Baseline (1.5 Pooled RNs)</td>
<td>3.6 ± 5.9</td>
<td>18 ± 29</td>
</tr>
<tr>
<td>2 Dedicated RNs + Regular Triage</td>
<td>1.5 ± 3.8</td>
<td>68 ± 108</td>
</tr>
<tr>
<td>2 Dedicated RNs + Pre-triage</td>
<td>1.4 ± 3.7</td>
<td>9 ± 13</td>
</tr>
<tr>
<td>2 Pooled RNs + Pre-triage</td>
<td>0.68 ± 1.66</td>
<td>2.25 ± 3.7</td>
</tr>
</tbody>
</table>

#### Figure 23.4 The Impact of Triage Scenarios on the Distribution of Wait Times

4 Implications for Managers

Individuals with medical or nursing training and experience have been managing healthcare facilities for a very long time. Notwithstanding the increasing number of management professionals recruited to positions with decision-making authority, the management and policymaking positions in the health sector are still largely occupied by people without any formal exposure to the fundamentals of management. Hence, the health sector is plagued by the prevalence of “management by intuition” and/or “management based on experience”. However, the previously discussed complexities associated with the dynamics of healthcare delivery operations often render the intuitive solutions sub-optimal. In addition, the “low hanging fruit” often does not result in the envisioned benefits in such complex systems. This is why it is essential that the managers of healthcare facilities take advantage of detailed analyses through the use of POM methodology in improving healthcare delivery. In facilities of sufficient size, we would advocate for an in-house team of POM experts, rather than resorting to one of the reputable consulting.
companies. The advantage of this approach, in our experience, is that the POM team will own
the process improvement problem as seen through the continuous improvement of the delivery
process.

5 Conclusions and Future Research Directions

This chapter highlights the usefulness of POM in designing and improving preventive and emer-
gency care processes. In closing, we comment on some of the remaining research challenges in
these two domains and point out the use of POM for primary care as a fruitful research avenue
that is largely untapped.

Concerning preventive care, the prevailing studies cited in Section 2 represent accessibility in
terms of the patient’s distance to or time in the facility. There is empirical evidence that there are
additional factors that attract patients to a preventive care facility, such as the friendliness of the
staff and the availability of parking. In addition, these studies are not applicable to representing
preventive care services that work with appointments (i.e., they are geared toward walk-in ser-
vice). One of the main challenges associated with modeling appointment systems is the fact that
wait times for appointments are often not commensurate with the travel times to and the service
times at the facilities. Research on overcoming these issues would be a welcome contribution,
particularly from the perspective of practical applicability.

There are a number of challenges facing researchers concerning the use of POM for emer-
gency care. Although an abundance of ED simulation models have been reported in the scholarly
literature (Saghafian et al. 2015), their uptake by the ED managers has not been stellar. To the
best of our knowledge, a very significant majority of the simulation models have been used for
decision-making episodes only once or twice rather than becoming an integral part of the toolkit
of ED managers. The suite of ED simulation models developed by Mandelbaum and his team
(Zeltyn et al. 2011) (which are being used in Israel at multiple hospitals) is the only exception
to this that we know of. There is a need for a concerted effort to understand the reasons behind
this low uptake in practice. In addition, top-tier POM journals are typically not interested in
publishing such application-oriented simulation work, unless there is a methodology contribu-
tion. This certainly constitutes a concern. Thus, the recent trend we mention at the beginning
of Section 3 involving the dual use of analytic and simulation approaches for tackling specific
processes at the ED rather than the entire ED process seems to be the best approach to push
frontiers of research at this time.

An important upstream phase of the healthcare continuum that we did not discuss in this
chapter is primary care. This is not an oversight but mainly a reflection of the state of using POM
for designing and improving primary care processes. The first point of contact for a patient with
the health system should be primary care. Despite its significant impact on the overall population
health, primary care has been largely unstudied by POM researchers. To this end, the recent
paper by Güneş et al. (2014) on matching patient and physician preferences and the research by
Graber-Nadich et al. (2015) on the regulator’s perspective concerning primary care are merely
initial steps. There is a lot of work that remains to be done in this important domain.

References and Bibliography

Aboolian, R., O. Berman and V. Verter (2016) “Maximal accessibility network design in the public sector.”


