Lean Transformation Frameworks for Hospital Departments

Ting Yu

A Thesis
In
The Department
Of
Mechanical, Industrial and Aerospace Engineering

Presented in Partial Fulfilment of the Requirements
For the Degree of
Doctor of Philosophy (Industrial Engineering) at
Concordia University
Montreal, Quebec, Canada

June 2020

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Ting Yu

Entitled: Lean Transformation Frameworks for Hospital Departments

and submitted in partial fulfillment of the requirements for the degree of

Doctor Of Philosophy (Industrial Engineering)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Ahmed A. Kishk	Chair
Dr. Jeannie Haggerty	External Examiner
Dr. Yong Zeng	External to Program
Dr. Mingyuan Chen	Examiner
Dr. Onur Kuzgunkaya	Examiner
Dr. Kudret Demirli	Thesis Co-Supervisor
Dr. Nadia Bhuiyan	Thesis Co-Supervisor

Approved by

Dr. Ivan Contreras, Graduate Program Director

August 21, 2020

Dr. Mourad Debbabi, Dean
Gina Cody School of Engineering and Computer Science

Abstract

Lean Transformation Frameworks for Hospital Departments

Ting Yu, Ph.D.
Concordia University, 2020

Critical issues in hospitals include prolonged wait times and soaring expenditures. While Lean methods appear to be a solution to these issues, Lean implementations are difficult to replicate in different healthcare settings because of a lack of well-established methodologies. Therefore, this thesis applies a unified Lean approach to create transformation frameworks for three major types of hospital departments that are facing different challenges attributed to their underlying characteristics.

The first part of this thesis discusses a framework for schedule-based and treatment-oriented outpatient departments. The framework identifies the best patient mix, schedule, and staffing levels to reduce wait times and improve the utilisation of critical resources. After implementing the framework, an oncology department can run under a one-day regime while reducing patient visit time by 36.3% and increasing the number of daily chemotherapy treatments by 38.6%. The second part of this thesis describes a framework for shared processes in outpatient departments in which scheduled and unscheduled patients coexist. The framework analyses system capability and devises efficient patient and staff schedules to meet demand with supply in a timely fashion. Implementing the framework in a radiology department results in a 21.4% increase in the number of elective patients and 77.1% and 37.2% decreases in the lead times of emergency patients and elective patients, respectively. The final part of this thesis demonstrates a framework for emergency departments receiving unscheduled patient visits. The framework manages the discharge process and plans staffing and materials to alleviate crowding. The estimated results in a case emergency department are a 50.8% reduction in patient length of stay and 41.5% decrease in occupancy rate of stretchers.

The major theoretical contribution of this thesis is its adaptation of several tools, which are organised into a series of structured activities towards Lean systems for different types of hospital departments. In practice, each framework is a comprehensive and step-by-step guideline for hospital managers to follow in their application of Lean to serve more patients, decrease wait time, and use their resources more efficiently. Furthermore, the frameworks are expected to foster a culture of continuous process improvement and facilitate a hospital-wide Lean transformation.

Acknowledgements

First, I would like to express my deepest gratitude to my supervisors, Dr. Kudret Demirli and Dr. Nadia Bhuiyan for their financial and academic support through my graduate studies. Without them, it would have been impossible to complete my Ph.D. programme.

Second, my committee members are greatly acknowledged, Dr. Kudret Demirli, Dr. Nadia Bhuiyan, Dr. Yong Zeng, Dr. Mingyuan Chen, Dr. Onur Kuzgunkaya, and Dr. Jeannie Haggerty, for their valuable comments as well as their precious time in reading my work and attending my defence.

Additionally, I would like to express my special thanks to the radiology and the emergency teams and St Mary's Research Centre under centre intégré universitaire de santé et de services sociaux de l'Ouest-de-l'Île-de-Montréal (CIUSSS ODIM) for their support. I would like to thank Dr. Jeannie Haggerty, Professor and Research Chair at McGill University and St Mary's Research Centre, Dr. Renée Proulx, Delegated Director of Academic Affairs, Teaching and Research Directorate at CIUSSS ODIM, Mr. Pierre Lemay, Chief of Performance and Statistics Department at CIUSSS ODIM, for their helpful insights and professional advice. Furthermore, I would like to express my deepest gratitude to the CIHR Healthcare System Impact Fellowships and CIUSSS ODIM for offering me the grants to conduct the emergency department project.

Finally, I owe my gratitude to my family and friends who have provided the strongest support throughout my studies and my life in general. My deepest thanks to my beloved partner, Ruofei Ma, and my dearest friend and mentor, Rick Mah.

Contents

List of Figures

List of Tables

List of Acronyms

AWT	Average Weighted Treatment Time
AWLT	Average Weighted Lead Time
AWCT	Average Weighted Cycle Time
BC	Balance Chart
BMI	Body Mass Index
C/T	Cycle Time
CT	Computerised Tomography
CCT	Cell Cycle Time
DMAIC	Define, Measure, Analyse, Improve, Control
DES	Double Exponential Smoothing
DI	Discharge Interval
ED	Emergency Department
ECT	Effective Cycle Time
FIFO	First-In-First-Out
FSS	Finished Stretcher Supermarket
HQC	Health Quality Council
LOS	Length of Stay
LWBS	Leave Without Being Seen
LT	Lead Time
LTF	Longest Treatment First
LAF	Lowest Acuity Level First
NHS	National Health Service
OPD	Outpatient Department
OBC	Operator Balance Chart
PDCA	Plan, Do, Check, Act
PCT	Planned Cycle Time
RIE	Rapid Improvement Event
RICE	Rapid Improvement Capacity Expansion

SI	Scheduling Interval
SES	Single Exponential Smoothing
TT	Takt Time
TWC	Total Work Content
VMPS	Virginia Mason Production System
VSM	Value Stream Map

List of Notations

Chapter 3

Lean Transformation Framework for Treatment-oriented Outpatient Departments

c_t	Effective capacity
x	Number of a studied resource
T_x	Effective working time of the x^{th} resource
c^{PM}	Pacemaker capacity
n	Number of patient groups
ω_n	Ratio of the n^{th} patient group
t_n	Treatment time of the n^{th} patient group
α	Human factor
m	Number of resources
c_e	Effective capacity
d^{US}	Daily upstream patients
d^{DS}	Daily downstream patients
d^{HD}	Daily home-downstream patients
I	Number of scheduling intervals
d_i^{PM}	Pacemaker patients in scheduling interval i
d_i^{HD}	Home-downstream demand in scheduling interval i
d_i^{US}	Home-upstream demand in scheduling interval i
d_i^{DS}	Downstream demand in scheduling interval i
d_i^{PD}	Pacemaker-downstream demand in scheduling interval i
k	Number of patient groups that go to the downstream process through the pacemaker in scheduling interval i
d_{k_i}	Number of patients in the k^{th} patient group per scheduling interval i
d_k	Number of patients in the k^{th} patient group per day
d_i^{UD}	The upstream-downstream demand in scheduling interval i
c_i^{US}	Upstream capacity in scheduling interval i

Δ_i^{PM}	Pacemaker patients that exceed the upstream capacity in scheduling interval i
TT_s	Takt time (or the highest ECT) of the succeeding cell
TT_p	Takt time (or the highest ECT) of the preceding cell
d_i	Number of patients in scheduling interval i
ρ	Efficiency factor

Chapter 4

Lean Transformation Framework for Shared Processes in Outpatient Departments

c_i	Interval capacity
x_i	Urgent demand in interval i
N	Number of critical resources
z_i	Standard score for interval i
μ_i	Mean demand forecast for urgent patients in interval i
σ_i	Standard deviation for urgent demand in interval i
μ	Mean demand forecast for urgent patients
σ	Standard deviation for urgent demand
T_{n_i}	Net available time of the n^{th} critical resource in interval i
P	Probability of the pacemaker meeting urgent demand without delay
t_u	Exam time of urgent patients
d^P	Daily planned urgent demand
d_i^E	Number of electives in scheduling interval i
θ_i	Percentage of patients in scheduling interval i
t_e	Exam time of electives
k	Desired service level
m	Number of non-critical resources
P_j	Probability given by a standard score $z = j$
ρ	Efficiency factor
y_t	Actual observation at time t
$\hat{y}_t$	Fitted value at time t

Chapter 5

Lean Transformation Framework for Emergency Departments

t_{la}	Time to start the blood work
t_{con}	Processing time of consulting
t_{im}	Processing time of imaging
t_{twc}	Total work content
m	Number of resources
α	Human factor
i	Time index; I denotes the total hours per day
j	Shift index; J denotes the number of shifts per day
k	Option index; K denotes total number of mid-shift options
x_{kj}	Number of operators assigned to shift j under the k^{th} option
y_{ki}	Number of operators during the i^{th} hour of the day under the k^{th} option
z_k	Square of the sum of the discrepancy between the demand and effective capacity throughout the day under the k^{th} option
d_i	Demand in the i^{th} hour of the day
β_{ki}	1 if mid-shift is effective in the i^{th} hour under the k^{th} option; otherwise 0
t_i	Net available time in the i^{th} hour of the day
Q	Number of replenishment cycles
Δ_i	Number of untreated patients in the i^{th} hour of the day
c_i	Resource capacity in the i^{th} hour of the day
μ	Mean forecast demand
σ	Standard deviation of the demand
β	Admission rate
t_B	Boarding time
$\hat{y}_t$	Fitted value at time t
y_t	Actual observation any time period t

1. Introduction

1.1 Background

Hospitals aim to provide excellent patient care with reduced wait time, minimum healthcare expenditure, and high quality service (*10 IHI Innovations to Improve Health and Health Care*, 2017). While making progress towards more advanced technology-care practices, hospitals are under pressure to enhance performance by reducing patient wait times, efficiently allocating a scarce healthcare workforce, and meeting an increasing demand through better healthcare scheduling strategies (Gleeson et al., 2016; Leggat et al., 2015; Naiker et al., 2018). Facing these issues, many healthcare researchers have begun experimenting with various improvement approaches, one of which is Lean.

Lean is a process improvement concept that strives to improve value, as defined by the customer, through continuous process improvements with minimised financial input (Shook & Marchwinski, 2014). The foundation of Lean is based on five principles: value, value stream, flow, pull, and perfection (Womack & Jones, 1997). The Lean principles not only provide a comprehensive guideline of how to improve a system but also ensure the effectiveness of the improvement process by identifying customer values. Specifically, Lean systematically examines steps, identifies the components that add value, and eliminates those that do not (i.e. waste) from the perspective of customers. Once the waste is removed, the process is improved by ensuring that the remaining steps are efficient and as integrated as possible, allowing for flow. As flow is introduced, Lean enables customers to pull value through the system and finally pursues continuous improvement.

Rooted in the automotive industry, Lean, known as Lean Healthcare in health systems, was introduced to healthcare sectors in the 1990s because of its strong resonating success in the manufacturing environment (Ohno, 1988). The first Lean application in healthcare was possibly a work published two decades ago by the National Health Service (NHS) Modernisation Agency, one of the biggest healthcare organisations implementing Lean in the UK (De Brandao, 2009). More celebrated Lean examples in the NHS can be found in the articles published by Silvester et al. (2004) and Matthias and Brown (2016). Lean Healthcare was introduced later to North America, and by 2009, 53% of US hospitals reported having implemented Lean to some extent (*Hospitals*

See Benefits of Lean and Six Sigma, 2009). The Virginia Mason Institute, one of these US health institutions, established a Lean system known as the Virginia Mason Production System on the foundation of Kaizen, one of the most important Lean tools (Kenney, 2012). In Canada, having realised the urgency of health system improvements and the promising future of Lean, the Conference Board of Canada commissioned a series of briefings on Lean in Canada with a goal of promoting Lean implementation nationwide (Fine et al., 2009). In 2010, a government-committed system-wide Lean transformation was initiated in Saskatchewan, Canada (Marchildon, 2013). The Health Quality Council of Saskatchewan concluded that Lean increases patient safety by eliminating errors, increases patient satisfaction, reduces cost, and improves patient health outcomes (Bath et al., 2016). Moreover, MacKenzie and Hall (2014) sent online surveys to the senior management of all health regions in Canada with questions regarding their Lean experience. The results show that Lean is effective in improving efficiency, reducing cost, and increasing productivity (number of patients treated) through the redesign of the workflow and resource schedule. Apart from these positive outcomes, approximately three-quarters of physicians and staff reported that their attitudes towards Lean are positive, particularly in organisations in which Lean has been practiced for years.

Lean brings healthcare organisations both tangible and intangible benefits. The tangible benefits include wait time and cost reduction, and the intangible benefits involve a better understanding of the system, cross-team collaborations, and increased patient and employee satisfaction. However, soon after Lean was implemented, healthcare researchers raised concerns. Spear (2005) underlined that Lean Healthcare has not yet been fully institutionalised to Toyota's level, which has the ability to continuously and systematically eliminate waste. Spear's contention is largely supported in the literature over the last decade. Young and McClean (2008) concluded that Lean Healthcare practices may be pragmatic but fragmented because people simply mimic how Lean works in manufacturing, while in fact, health systems are different from this industry. Mazzocato et al. (2010) reported that in 112 reviewed articles, most organisations limited themselves to adopting specific Lean techniques to solve a target problem within one small and specific area. A review of Lean applications in the world-famous Lean organisation, the English NHS, indicates that Lean still tends to be isolated rather than systematic (Burgess & Radnor, 2013). A recent paper demonstrates that Lean Healthcare implementation in Quebec, Canada, shares the same challenges, being unable to transition from using tools and techniques into system-wide change (Fournier & Jobin, 2018). Hallam and Contreras (2018) stated that empirical studies of Lean are still highly localised with small successes that cannot be easily sustained or reproduced. Additionally, researchers in other countries described similar issues in Lean implementation, such as a high heterogeneity in the studied healthcare areas and the imitation of simple techniques of knowledge in the manufacturing area (Costa & Filho, 2016; Parkhi, 2019).

The Lean journey in health systems is continuing. Common contexts in health systems (e.g. oncology department, radiology department) interacting with different components of Lean interventions (e.g. Kaizen, value stream) trigger different change mechanisms and various project scopes. All these factors make Lean a patchy, fragmented, non-reproducible, and hard to sustain method. On the other hand, past Lean practices are wide-ranging but can be categorised either based on organisational settings, such as the types of departments, or clinical fields, which includes clinical specialties, diagnostic services, and support services (Mazzocato et al., 2010). However, this classification is proposed solely based on the structures of health systems and clinical practices without any connection to the Lean system.

Overall, on the one hand, people have acknowledged the promising future of Lean in improving healthcare performance; on the other hand, an unsolved issue impeding the use of Lean in healthcare is the lack of structured methodologies that demonstrate how to transform health systems into Lean systems.

1.2 Research objectives and thesis outline

In response to the existing research gaps and the pressing need for healthcare improvement, the primary goal of this thesis is to extend Lean implementation in hospitals beyond small practices to a system level transformation with step-by-step guidance. We therefore introduce a unified Lean approach following three major steps. This approach primarily identifies patient demand in a given healthcare setting. To meet this demand, Lean resource planning is applied to determine an appropriate number of operators in every step by synchronising the operation rate (i.e. cycle time, CT) with the pace of patient arrivals (i.e. takt time, TT). Then, flowing patients throughout the entire value stream entails devising an effective patient scheduling strategy or managing buffers by allocating an appropriate level of the material. The foundation of this Lean approach is Lean principles and a series of adapted Lean tools for healthcare contexts.

As the proposed approach begins with the identification of demand, we classified hospital departments into three types based on the features of their patient demand: treatment-oriented outpatient departments (OPDs) refer to appointment-based departments (i.e. known demand); shared processes in OPDs refer to departments in which scheduled and unscheduled patients compete for resources (i.e. demand is partially known); and emergency departments (EDs) representing departments in which patients are not scheduled (i.e. unknown demand). The specific objectives of this thesis are as follows:

1) To explore the main characteristics and challenges of each type of department;
2) To use a unified Lean approach to create a Lean transformation framework that incorporates proper Lean tools to identify patient demand, devise better schedules for resources and/or patients, and manage wait times in each type of department; and
3) To demonstrate the applications of these Lean frameworks in real clinical settings.

The structure of this thesis is organised as follows. After the introduction, Chapter 2 presents the current state of Lean in health systems by analysing the relevant literature. Lean frameworks for treatment-oriented OPDs, shared processes in OPDs, and EDs are presented in Chapters 3, 4, and 5, respectively. Chapter 6 concludes the thesis and envisions future research.

2. Literature Review

The last two decades have witnessed the advance of Lean implementation in health systems and an increase in Lean healthcare research in different countries. However, many researchers have implied that Lean applications are quite different due to the fact that every Lean study is a result of two variables: scale and scope (Fournier & Jobin, 2018; Hallam & Contreras, 2018; Mazzocato et al., 2010; Radnor et al., 2012). The scope of Lean intervention varies from the implementation of specific Lean tools, such as Kaizen, to the system-level application of Lean principles, while the research scale of a Lean study can vary from local process improvements, e.g. the scheduling process, to hospital-wide transformations. This chapter reviews Lean studies from both contextual and methodological aspects.

The process of literature review involves two steps. Papers published in peer-reviewed journals up to May 2020 were selected by researching the Engineering Village, Scopus, Google Scholar, and Pubmed databases, using the combination of the keywords such as ‘Lean’, ‘Lean methodology’, ‘Lean thinking’, ‘Kaizen’, ‘pull system’, ‘Rapid improvement event’, ‘hospital’, ‘acute care’, ‘healthcare’, ‘hospital department’. We included only articles that are published in the English language. We then scrutinized the potential works to remove duplications and identified additional papers by reading papers included in the first round. All selected papers are studied in terms of scope (i.e. Lean tools/principles applied) and scale (i.e. in which context Lean was used).

2.1 Scope of Lean healthcare applications

The scope of Lean healthcare applications can be categorised into three levels as shown in the Lean Pyramid (Figure 2.1). At the base of the pyramid are studies focused on specific tools/activities of Lean, while the mid-level of Lean implementation refers to the studies that concentrated on Lean principles. With the support of these tools and principles, a system would reach the level of Lean, where Lean does not occur as a single intervention but instead requires ongoing efforts and interventions aimed at improving the process.

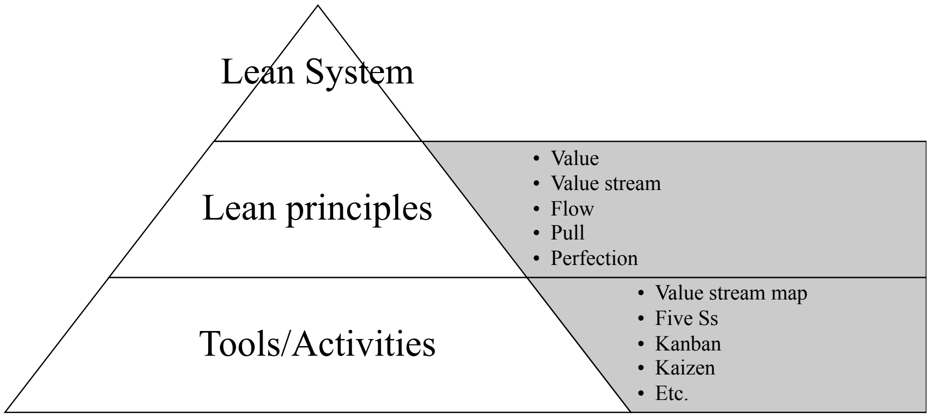


Figure 2.1 Three Level Lean Pyramid

2.1.1 Base of the Lean pyramid

Lean healthcare is often perceived as a set of tools and techniques for improving processes. This is called the nominal “Lean” phenomenon by Radnor et al. (2012). Some of the Lean tools are easy to teach, and results can be obtained quickly after applying these tools; however, the tool level applications show localised impacts, and fall short of their intended impact on the overall performance of the system. Table 2.1 lists the tools used in 59 Lean healthcare articles in the recent decade.

Table 2.1 Lean Tools in Health Systems

Name	Brief Definition	References	# Articles (%)

Process mapping	A tool that describes a series of activities in a process	Aakre et al.(2010), Abouzahra and Tan(2014), Bal et al.(2017), Baril et al.(2016), Brett Benfield et al.(2015), Bhat et al.(2016), Carter et al.(2012), Chadha et al.(2012), Chan et al.(2014), Cheung et al.(2016), Cookson et al.(2011), Damle et al.(2016), Doğan and Unutulmaz(2016), Duska et al.(2015), Ford et al.(2012), Gill(2012), Gonzalez et al.(2014), Hitti et al.(2017), Hydes et al.(2012), Improta et al.(2018), Kruskal et al.(2012), LaGanga(2011), Martin et al.(2013), Mazzocato et al.(2012), McDermott et al.(2013), McDermott et al.(2015), Moore and Arthur(2019), Murrell et al.(2011), Ng et al.(2010), Rico and Jagwani(2013), Robinson et al.(2012), Sanz et al.(2019), Skeldon et al.(2014), Smith et al.(2011), Teichgräber and de Bucourt(2012), Vashi et al.(2019), Wang et al.(2015), White et al.(2014), Wong et al.(2012)	36 (63%)
Kaizen/RIE	Small process improvement activities	Baril et al.(2016), Beck et al.(2016), Collar et al.(2012), Deanna Belter et al.(2012), Holden and Hackbart(2012), Jayasinha(2016), LaGanga(2011), Lamm et al.(2015), Mazzocato et al.(2016), McDermott et al.(2013), Murrell et al.(2011), Naik et al.(2012), Piggott et al.(2011), Skeldon et al.(2014), Smith et al.(2012); White et al.(2014), Yousri et al.(2011)	16 (28%)
PDCA/DMAIC	Plan, do, check, act /define, measure, analyse, improve, control	Aakre et al.(2010), Al- Araidah et al.(2010), Baril et al.(2016), Bhat et al.(2014), Bhat et al.(2016), Cheung et al.(2016), McCulloch et al.(2010), Ng et al.(2010), Niemeijer et al.(2010), Rutledge et al.(2010)	10 (18%)
5Ss	Sort, Set In order, Shine, Standardise and Sustain	Al- Araidah et al.(2010), Improta et al.(2018), Rutledge et al.(2010), Shah et al.(2013)	4 (7%)
Standardisation	Work elements are efficiently organised in a way in which they can easily be repeated with the least effort	Kruskal et al.(2012), Migita et al.(2011), Simons et al.(2014)	3 (5%)
5whys/fishbone	A way of identifying root causes of a problem	Al- Araidah et al.(2010), Carter et al.(2012), Smith et al.(2011), Cheung et al.(2016)	4 (7%)
Kanban	A scheduling system	Ng et al.(2010), Newell et al.(2011)	2 (4%)
Cell	A series of processing steps running close to	Chadha et al.(2012), Wang et al.(2015)	2 (4%)

	one another to have a continuous flow		
A3	Structured problem-solving and continuous-improvement approach	Simons et al.(2014)	1 (2%)
Takt time	Time between demand arrivals	Wang et al.(2015)	1 (2%)

Lean healthcare strongly relies on the use of Lean tools. As this table indicates, 11 tools are identified in these articles, in which two-thirds of the Lean applications use process mapping and one-third mention Kaizen or rapid improvement event (RIE). Articles with respect to plan, do, check, act (PDCA) and define, measure, analyse, improve, control (DMAIC) account for 16% of the total. Major tools in Lean Manufacturing, such as cell, Kanban, takt time (TT), and supermarket, are rarely seen in healthcare systems (less than 10%). Almost all the articles mentioned waste elimination to some extent.

Process mapping is the most common tool in Lean healthcare research. The value stream map (VSM) is one of the most powerful process mapping tools in Lean. It is a simple diagram of every step involved in the healthcare workflow needed to treat a patient. A complete set of VSMs for a project has at least two VSMs, one for the current system analysis and the other for a future state (sometimes, an interim VSM can be added in between). The importance of the VSM depends on its ability to capture the overview of a system and guide the Lean implementation (Rother & Shook, 2003). The future VSM can even be a structural approach to create process improvement events. For example, Gonzalez et al. (2014) used a VSM in a urology department to identify opportunities to improve the system. Two maps were created, one for the current system and another for the proposed value stream. The current VSM shows how the steps are logically connected to one another, how the information and/or the patient (product) flows along the process, and what the required and available resources at each step are. The proposed VSM, called the future VSM, demonstrates the improvements after waste reduction. Current and future VSMs are of great importance in the success of Lean transformations; one helps in understanding the existing facts and the other assists in the systematic transition process of Lean. However, the literature implies that most researchers applied only current VSMs (Hydes et al., 2012; Improta et al., 2018; McDermott et al., 2013) or simply utilised flow charts instead (Baril et al., 2016; Vashi et al., 2019). Indeed, researchers were able to look at the overall system and identify waste by using only the current VSM; however, Lean healthcare implementation is either too ambitious or too conservative without the proposed VSM.

Kaizen, also called a rapid improvement event (RIE), is the second most frequently used method. It is “a focused and structured improvement project, using a dedicated cross-functional team to improve a targeted work area, with specific goals, in an accelerated timeframe” (Glover et al., 2014). Kaizen continues to gain popularity among many healthcare organisations in that it 1) provides a faster return, which is more visible and does not challenge existing management controls (Holden, 2011), and 2) is favoured by the staff as they feel engaged in an improvement process that quickly demonstrates results and requires their input (Mazzocato et al., 2016). In many circumstances, Kaizen was used after completion of the current VSM to develop solutions to eliminate waste and improve flow. In an outpatient service sector, to adopt an RIE in a Lean process improvement project in an outpatient sector, the manager and researcher convened a “rapid improvement capacity expansion” (RICE) cross-functional team (LaGanga, 2011). After the workflow was identified, the RIEs were conducted to assist in process improvements and the elimination of waste. As one of the most powerful Lean tools, there are two types of Kaizen: system/flow Kaizen and process Kaizen. These are all process Kaizen events; flow Kaizen has not yet been identified in the field of healthcare systems.

2.1.2 Second level of the Lean pyramid

The second level of the Lean Pyramid includes the five Lean principles. Although many large healthcare organisations have widely implemented Lean, evidence of the application of all five principles in healthcare is not sufficient. To investigate the extent to which Lean healthcare has been implemented in healthcare systems, we again reviewed the literature listed in Table 2.1. Studies that fall into the range of a Lean principle used or at least intended to use the corresponding principle.

All the reviewed articles meet the first principle of value to a certain degree. In the reviewed papers, the goals of Lean implementation cover the topics of patient satisfaction (Hydes et al., 2012; Jayasinha, 2016), employee satisfaction (Mazzocato et al., 2016), hospital length of stay (LOS, patient arrival to departure) (Murrell et al., 2011; White et al., 2014), patient wait times (Ng et al., 2010; Vashi et al., 2019), and patient safety (Newell et al., 2011; Simons et al., 2014), all of which targeted improving the value to patients and associated healthcare stakeholders.

Approximately two-thirds of the reviewed papers (all that utilised process mapping in Table 2.1) fall into the second category. The value stream identifies every step involved in the production of products or services and is a starting point in understanding a process. Value steam mapping has become a well-known procedure in Lean healthcare and assists healthcare management and decision makers in understanding the information and patient flows. Teichgräber and de Bucourt (2012) applied a VSM to identify the work and patient flows and further reduced non-value-added activities along the value stream. Gonzalez et al. (2014) drew a VSM of a urology department for the purpose of identifying the waste and reducing the total patient LOS. There are also many other studies that anecdotally describe every step in the health system without using a scientific method of process mapping.

Establishing flow means that, after waste has been eliminated, steps for manufacturing a product are adjacent to each other so that materials can be processed in very nearly continuous flow, either one at a time or in small batch sizes (Shook & Marchwinski, 2014). This is mostly supported using a cellular design. There is only one article in which a cellular design was adopted in a Lean healthcare application (Wang et al., 2015). The authors converted the current design of an ED to certain U-shaped cellular workstations by rearranging the process elements adjacent to each other and loading the cellular workstations with a calculated number of resources. For other studies, despite the fact that many state Lean practitioners investigated the processes and reduced identified waste in their systems, there is not sufficient evidence that their waste elimination activities supported the creation of a continuous flow. Mazzocato et al. (2016) evaluated 186 structured Kaizen documents containing improvement suggestions that were produced by employees at a hospital. The authors indicated that a substantial number of Kaizen documents addressed symptomatic problems and that there is a need to combine Kaizen practices with a goal of creating a flow throughout the entire system.

The first three principles pave the way for creating a pull system. Ideally, the customer can pull the value from the most downstream process to assure that the products or services are as close as possible to the customer requirements. In manufacturing, supermarkets are used to create a pull system. A supermarket is a location where a predetermined standard inventory is maintained to supply downstream processes (Shook & Marchwinski, 2014). In a health system, patients are not only customers but “products” as well, which makes the pull principle more challenging, and therefore literature regarding this principle is sparse. In 2009, a book titled *Making Hospitals Work*

(Baker et al., 2009) anecdotally described a pull system in which empty beds instead of patients become the products. Unoccupied beds are placed in a supermarket, and whenever a patient arrives, he/she pulls a bed and continues to be treated by downstream resources without waiting for a bed to be allocated. This is an excellent example of a pull system with minimised patient wait times; however, the book does not explicitly discuss the pull mechanism, and no follow-up study can be found on this topic. There are studies that discuss the pull system based merely on anecdotal stories: Abouzahra and Tan (2014) created a simple pull system between the local and main pharmacy; Ng and others (2010) in their articles described a pull system in which a Kanban box is used to discharge patients; Newell and his colleagues (2011) also applied a Kanban system to control the use of medications. In contrast, two of these studies are in the pharmacy in which the process is akin to a manufacturing production line with medications as inventories.

The final principle of Lean is perfection. Lean includes cyclic activities, resulting in more efficient practices in the continuous pursuit of perfection. This principle is the most difficult to trace as it requires follow-up studies. Abouzahra and Tan (2014), Ng et al. (2010), and Newell et al. (2011) claimed that their studies followed the five Lean principles. There is one available evaluation study regarding the work of Ng and others (Vermeulen et al., 2014). The results show that Lean appeared to be effective in wait time reduction at the beginning; however, the benefits started fading away over the years. This phenomenon occurred because Lean principles were actually not fully applied in the study owing to the absence of establishing pull and seeking perfection.

2.2 Scale of Lean healthcare applications

Healthcare delivery involves numerous organisational units, ranging from hospitals to units providing a single service. The variations in the scale of Lean studies are therefore vast. Table 2.2 summarises the implementation areas of the reviewed empirical studies. The literature shows that there are five levels of Lean studies in terms of scale: hospital, cross-departmental, departmental, process, and step levels, among which departmental level studies comprise the majority.

Table 2.2 Scale of Lean Healthcare Studies

Scale	Reference	#articles (%)
Hospital (or cross-department)	Ford et al.(2012), Yousri et al.(2011)	2 (4%)

Department	Abouzahra and Tan(2014), Al- Araidah et al.(2010), Bal et al.(2017), Baril et al.(2016), Beck et al.(2016), Bhat et al.(2016), Chadha et al.(2012), Chan et al.(2014), Cookson et al.(2011), Collar et al.(2012), Damle et al.(2016), Doğan and Unutulmaz(2016), Duska et al.(2015), Gill(2012), Gonzalez et al.(2014), Hydes et al.(2012), Holden and Hackbart(2012), Improta et al.(2018), Jayasinha(2016), Kruskal et al.(2012), LaGanga(2011), Mazzocato et al.(2012), McDermott et al.(2013), Migita et al.(2011), Moore and Arthur(2019), Naik et al.(2012), Ng et al.(2010), Rico and Jagwani(2013), Rutledge et al.(2010), Sanz et al.(2019), Skeldon et al.(2014), Smith et al.(2011), Vashi et al.(2019), Wang et al.(2015), White et al.(2014)	26 (46.4%)
Process	Aakre et al.(2010), Brett Benfield et al.(2015), Bhat et al.(2016), Carter et al.(2012), Cheung et al.(2016), Deanna Belter et al.(2012), Hitti et al.(2017), Lamm et al.(2015), Martin et al.(2013), McCulloch et al.(2010), Newell et al.(2011), Piggott et al.(2011), Teichgräber and de Bucourt(2012), Shah et al.(2013), Simons et al.(2014), Wong et al.(2012)	15 (26.8%)
Step	Mazzocato et al.(2016), McDermott et al.(2015), Murrell et al.(2011), Niemeijer et al.(2010), Robinson et al.(2012), Simons et al.(2014), Smith et al.(2012)	7 (13%)

Hospital and cross-department level studies are rarely seen in the reviewed articles. Despite a media-reported hospital-wide (and provincial) Lean transformation in Saskatchewan, Canada (Kinsman et al., 2014), no evidence has been found to support this claim. Cross-departmental work was discovered in papers that studied the flows of stroke patients and fractured necks of patients (Ford et al., 2012; Yousri et al., 2011). In both studies, patients came from EDs, required surgeries, and were then discharged to hospital units. Although Lean was mentioned in the papers, information on how Lean was utilised in the improvement of patient flow was missing.

Approximately half of the reviewed studies implemented Lean at departmental levels. Classifying these departments on the basis of healthcare organisational structures, we noticed that more than half of the studies were in EDs (Bal et al., 2017); hospital departments where specialised services are offered, e.g. outpatient oncology clinics (Duska et al., 2015); ancillary departments where services are provided to support the work of a primary physician, e.g. pharmacy (Al-Araidah et al., 2010); a few operating theatres (Simons et al., 2014); and a very few administrative departments, e.g. an information technology department (Holden & Hackbart, 2012). We also grouped these studies with respect to the characteristics of patients arriving with or without appointments. Of the reviewed studies, 33.3% were in appointment-based departments in which each patient is informed with a timeslot according to a doctor's availability, as for example, in an oncology clinic (Baril et al., 2016). In 47.9% of the studies, non-appointment-based departments were studied, in which patients present themselves without prior appointments, for example, in the ED (Beck et al., 2016). The remaining 12.5% of the studies are in departments that serve both appointment and non-appointment patients, for example, in the radiology department (Martin et al., 2013). Lean activities at the hospital department level were numerous; however, they were tailored to the department under study and were substantially differentiated from one another.

Almost one-third of the studies were conducted at the process level. A process level study concentrates on a series of interrelated steps. These steps, in most cases, constitute an independent process within a department in which multiple services are provided. For example, Shah and others (2013) applied Lean to improve the mammography screening workflow; mammography is one of the modalities in the radiology department, and it has a particular process that is not intertwined with other modalities with respect to resources. Similarly, Teichgräber and Bucourt (2012) adopted VSM techniques to eliminate non-value-added waste for the workflow of another modality in their radiology department. However, there are situations in which the selected process is a sub-process that includes a couple of steps but is not an independent process, e.g. the medication delivery process, which is a sub-process in providing inpatient services (Newell et al., 2011).

The last type of study is a step level study that focuses on one small component of a process in a department. These types of studies are quite small in terms of scale. Murrell and others (2011) worked on a triage step in an ED and developed a rapid triage and treatment system. They claimed that by improving the triage step, the average patient LOS and the rate of patients who leave without having been seen (LWBS) were reduced. The Lean implementations aimed at step level improvement mostly showed positive results because they were easier to sustain and monitor. This is also the reason why most Kaizen events in healthcare target small components of a system. However, these random and small improvements do not ensure a broad positive effect on the overall value creation for the system.

2.3 Summary

The disconnected Lean tools and the absence of major Lean tools in healthcare settings reinforce our belief that healthcare organisations are still in their infancy in adopting Lean. The use of Lean tools in healthcare institutions is a critical challenge. Its success largely rests on understanding that Lean is a system, not simply a toolbox or a trouble-shooting kit that includes fragmented activities for addressing pre-existing problems, and that health systems do differ from traditional manufacturing systems, with adaptations being necessary in implementing these tools.

Lean principles enable Lean to be contextualised in healthcare; however, the concept of the last three principles are easily overlooked, which also explains the absence of Lean tools used in the last three stages. Undeniably, previous studies showed promising outcomes such as reduced emergency patient wait times (Ng et al., 2010) and increased patient satisfaction (Hydes et al., 2012). Nonetheless, there are two associated consequences in adopting incomplete Lean principles: 1) isolated consequences that may or may not contribute to improvements in the entire system; and 2) the benefits obtained may not be sustainable and may diminish or disappear over time. The second consequence may not be as easy to resolve as it relates to the improvement culture in the organisation; however, developing a framework that is replicable might assist in cultivating this culture.

There are variations in the scale of Lean implementation, from a hospital to a single step. Departmental level studies are the most popular for two reasons. From a clinical perspective, process improvement is easier within departments than it is across departments because any activity that crosses two departments has no formal management procedure. Given that a hospital consists of interconnected and interdependent but sometimes self-contained departments, it would be easier for the hospital management to conduct departmental level studies and integrate the studies at some point. In addition, for hospital departments in which several processes operate independently, such as radiology, process-level studies may be more pragmatic.

3. Lean Transformation Framework for Treatment-oriented Outpatient Departments

3.1 Introduction

Treatment-oriented OPDs are a set of hospital departments in which consultations and treatments are given to scheduled patients. Some examples are medical and radiation oncology departments for cancer patients. The level of demand for the services offered by such OPDs is known, given that patients are scheduled in advance. The demand comprises patient quantity and type. The quantity is determined by the daily capacity of specialists, since OPDs typically prioritise the time of expensive specialists (Gupta & Denton, 2008). It is noteworthy that the daily number of specialists and their working hours vary based on their availability. The type refers to patients in different treatment stages. Because treatments are repeatedly given to patients who may not need consultations at each treatment stage (Liang et al., 2015), patients classified into the different types undergo different processes within the value stream.

Considerable and increasing demand for outpatient services exists due to the pressure to reduce costs and improve health service accessibility (Vogenberg & Santilli, 2018). Using OPD resources efficiently while reducing wait times and proposing better scheduling strategies have become the primary foci of hospital administrators (Armony et al., 2015; Gijo & Antony, 2014; Liang et al., 2015). However, it is not unusual for patients to stall in their treatment process following their consultation, thus creating longer wait times. This occurs because the compositions of patient types (defined as patients in different treatment stages) within critical processes (i.e. the treatment and consultation processes) are inconsistent due to the specialist-centred scheduling strategy found in OPDs. In addition, an uneven distribution of arrivals throughout the day, due to specialists' daily schedules, imposes an imbalanced workload on non-specialists, thus lowering the utilization of and generating fatigue for operators. If a high level of demand occurs at the end of the day, non-specialists have no time to catch up and may thus prompt the misperception that resources are lacking.

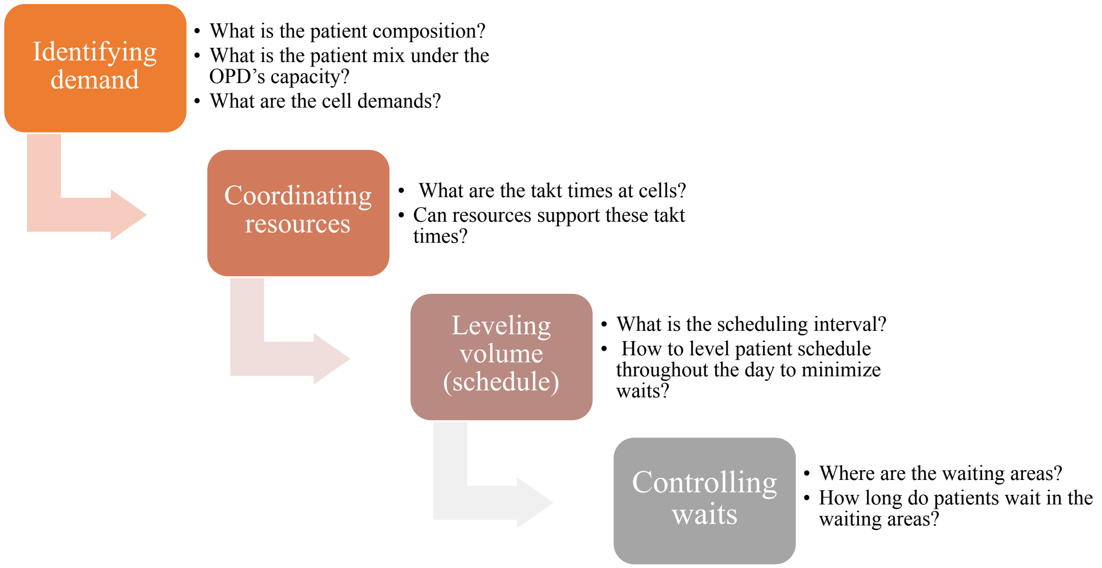
Given the development of Lean in healthcare systems, hospital managers have also implemented Lean in the context of OPDs. Gonzalez and colleagues (2014) applied a VSM in an outpatient urology department to eliminate waste based on better understanding of patient flow; hence, patient visit time was reduced in the department. After Lean implementation, the new process efficiency (defined as time spent with a healthcare provider versus total patient visit time) reached 60.7%, which constituted a 13% increase when compared with that of the actual situation. In outpatient uro-oncology clinics, Skeldon et al. (2014) carried out values stream analysis and a rapid improvement event to reduce waste in the patient flow. The result revealed that patient wait times were reduced while the time patients spent with healthcare providers versus that of the total patient visit doubled, therefore resulting in increased process efficiency. Another Lean study was conducted in a haematology-oncology clinic (Baril et al., 2016). The authors applied Kaizen to find solutions that could improve processes and satisfy stakeholders (doctors, nurses, etc.), thereby prompting a reduction in wait time of 74%. Although successfully implemented, Lean implementations in treatment-oriented OPDs are also tool-oriented and lack of a clear theoretical basis like other implementations reviewed in Chapter 2.

Given the challenges in treatment-oriented OPDs and the promising outcomes of Lean, in this chapter, we propose a Lean framework for use by treatment-oriented OPDs to meet the growth in demand while reducing patient wait times and improving their utilization of resources. This Lean framework incorporates Lean tools that are adapted to healthcare and elaborates on a process whereby OPDs are systematically transformed following Lean principles. The application of the proposed framework is demonstrated in an outpatient department at a local community hospital in Montreal, Canada.

3.2 Lean framework for treatment-oriented OPDs

A treatment-oriented OPD comprises clinical cells, such as those related to registration, consultation and treatment. Different patient types pass through different cells; the patients passing through a particular cell constitute its demand. Among the cells, the consultation cell serves as the pacemaker for such a Lean system, as all remaining cells should support and coordinate with the specialists. Furthermore, the treatment and consultation are the two critical processes as they perform the critical functions and determine the demand in treatment-oriented OPDs.

A systematic framework for Lean transformations in treatment-oriented OPDs is developed and presented in Figure 3.1. The framework is designed to reduce wait times and increase resource utilization by creating a smooth patient flow. It includes the following steps: demand identification, resource coordination, volume (schedule) levelling, and wait control, in alignment with Lean principles. The patient demand (i.e. the best patient mix) is obtained based on the patient compositions within the two critical processes (treatment and consultation) and the capacity of the OPD (determined by the specialists). Therefore, the demand of each cell is known and must be supported by an appropriate number of resources running at a particular speed of the TT, which constitutes the daily capacity of the non-specialists. An uneven distribution of arrivals is then prevented by assigning patient appointments in scheduling intervals based on the capacity of non-specialists. A patient queue, which is monitored through the first-in-first-out (FIFO) lane, is formed where two cells are running at different speeds. The performance is evaluated by an efficiency factor ρ, which is the total cycle time (C/T) versus the total patient visit time (i.e. the lead time, LT) (Shook & Marchwinski, 2014). The details of this framework are discussed below.



- What is the patient composition?
- What is the patient mix under the OPD’s capacity?
- What are the cell demands?

- What are the takt times at cells?
- Can resources support these takt times?

- What is the scheduling interval?
- How to level patient schedule throughout the day to minimize waits?

- Where are the waiting areas?
- How long do patients wait in the waiting areas?

Figure 3.1 Lean Transformation Framework for Treatment-oriented OPDs

3.2.1 Identifying patient demand

In the treatment-oriented OPDs, the treatment and consultation processes are not always synchronised, as mentioned in the introduction. This results in an imbalanced system in which some new patients (called excess new patients) are placed on the waiting list after their consultation, thus creating longer wait times. Hence, it is necessary to identify the real patient demand by identifying the best composition of patient types and adjusting the number of patient types based on the capacity of the OPD.

To this end, we first identify the patient compositions found at the pacemaker and treatment cells. Eliminating these excess new patients who would otherwise be placed onto the waiting list gives the OPD the current number of new patients that should be accepted. Because the number of new patients joining the system is proportional to that of patients in the remaining patient types, the current number of patients in the remaining patient types is obtained. Thus, we obtain the patient compositions at the pacemaker and treatment cells necessary for achieving a balanced treatment-oriented OPD.

The next step is to find the capacity of the pacemaker, or the maximum number of patients that can be scheduled in the OPD. The theoretical capacity, c_t, of a given resource is calculated as follows:

$$c_t = \frac{T_1+T_2+\cdots+T_x}{\mathrm{TWC}}, \tag{3.1}$$

where x is the number of a studied resource, T_x is the effective working time (the available time minus paid breaks) of the x^{th} resource, and TWC is the total work content (the amount of time that is required to process a patient by one operator). The pacemaker capacity in treatment-oriented OPDs is, therefore, calculated as follows:

$$c^{PM} = \frac{T_1+T_2+\cdots+T_m}{t}, \tag{3.2}$$

where m is the number of specialists, T_m is the effective working time of the m^{th} specialist and t is the treatment time. In Equation (3.2), the daily number of specialists and their working hours vary, as stated in the introduction; therefore, the daily capacity of the pacemaker varies.

If the patient treatment times are different, an average weighted treatment time (AWT) can be applied.

$$\text{AWT} = \omega_1 t_1 + \omega_2 t_2 + \cdots + \omega_n t_n, \tag{3.3}$$

where $\omega_1 + \omega_2 + \cdots + \omega_n \; = 1$; n is the number of patient types, ω_n is the ratio of the n^{th} patient type, and t_n is the treatment time of the n^{th} patient type. It is assumed that the same type of patient has the same treatment time.

Given the pacemaker capacity and the patient composition, we find the number of patients for each of the patient types, that is, the best patient mix. The total number of patients among all patient types that go through the pacemaker should be less than or equal to the pacemaker's capacity; the patients in the patient types should align with the patient compositions at both the consultation and treatment cells.

At the end of the stage during which demand is identified, a demand map is created to show the flow of patient types along with the demand mix in the value stream. Figure 3.2 provides a sample demand map containing five cells and five patient types in the patient value stream that is the subject of this evaluation; d_x denotes the number of patients for patient type x.

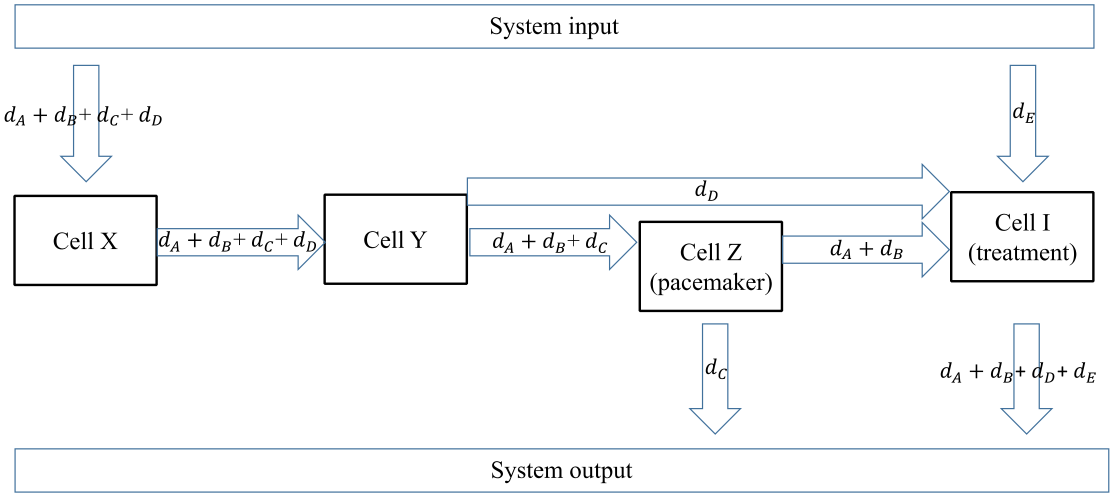


Figure 3.2 Sample Demand Map

As shown in Figure 3.2, the processes preceding the pacemaker (Cell Z) are upstream processes and those following it are downstream processes. Patients coming from home to Cell X and Cell I are home-upstream patients and home-downstream patients, respectively. The home-upstream patients are distributed into pacemaker patients entering the pacemaker, and the upstream-downstream patients are positioned at the start of the downstream process (Cell I). The pacemaker patients can then go home or go to Cell I, and are classified as pacemaker-discharged and pacemaker-downstream patients, respectively. The upstream-downstream and pacemaker-downstream patients join the home-downstream patients at Cell I as the latter are beginning their downstream processes. It is evident that among these patients, only the home-upstream and home-downstream patients are direct inputs for the entire system.

3.2.2 Coordinating resources

Resource coordination entails assigning the proper number of resources at the non-pacemaker cells to support the specialists and create patient flow. Because treatment-oriented OPDs are specialist-centred, the non-specialists typically lack coordination, thus creating isolated cells that result in underutilized resources and longer wait times. This section will discuss how to identify the appropriate number of operators in different cells in the treatment-oriented OPDs using the Lean techniques of C/T, TT, and planned cycle time (PCT).

TT essentially represents the time between the demand arrivals and is calculated as follows:

$$\mathrm{TT} = \frac{\text{Effective working time}}{\text{Total demand during that time}}. \tag{3.4}$$

All patients in a cell use the same resources, so all resources working within the cell are required to respect the same TT. However, the TT per cell could be different since the demand may vary in different cells within the value stream (as they serve different patient types).

The C/T is a measure of the frequency with which a patient is treated and reflects the operation speed. It can be discussed at resource-level or at process level. At resource level, the C/T ≤ TT is necessary to ensure that the process is running at the right speed to meet the demand. More specifically, this relationship is established as follows:

$$\mathrm{C/T} = \mathrm{TWC} = \alpha\mathrm{TT}, \tag{3.5}$$

where $\alpha<1$ is a factor which is used to compensate for operator fatigue, minor interruptions, and other minor uncertainties. A good target for α in healthcare is 85-90%. A gap of over 15% between C/T and TT indicates low resource utilization and one of less than 10% places a strain on resources. In Equation (3.5), increasing the number of resources m will increase the effective working time, thereby bringing the TT to be higher than the C/T. It should be noted that when AWT is used in Equation (3.5), the corresponding C/T becomes the average weighted cycle time (AWCT).

At process-level, in a cell with m resources, the speed of the cell's operation becomes m times faster. The objective is then focused on identifying the appropriate number of operators, m, in a cell to reduce the cell cycle time (CCT) below TT and meet the demand as shown in Equation (3.6):

$$\mathrm{CCT} = \frac{\mathrm{TWC}}{m} = \alpha\mathrm{TT}, \tag{3.6}$$

It should be noted that the calculated m should be an integer number, which therefore leads to an effective cycle time (ECT) that is less than the CCT.

In addition, by using Equations (3.4) and (3.5), we can obtain the maximum number of patients that can be processed by a resource or the effective capacity of a resource, c_e, as follows:

$$c_e = \frac{\alpha\times\text{Effective working time}}{\mathrm{TWC}}. \tag{3.7}$$

It should be noted that, the effective capacity is less than the theoretical capacity established above, as $\alpha < 1$ (due to various losses mentioned above).

The variations in demand along the value stream require different cells to move patients along at different speeds, which is accomplished by setting the appropriate staffing levels. To do so, the relation of TT, C/T, and PCT is demonstrated in a BC. In the BC of a cell, the C/Ts of all operators are represented as bars and are compared to a line of its TT (Duggan, 2018). We look for C/Ts that do not fall within 85-90% of the TT and recalculate the number of operators using Equation (3.6).

3.2.3 Levelling volume (schedule)

Non-specialists suffer from an imbalanced patient load throughout the day due to variations in specialists' daily schedules, as mentioned in the introduction. In this situation, their utilization tends to be lower and patients must wait longer. By levelling the volume, this challenge is overcome, as patients are then scheduled into scheduling intervals as evenly as possible, based on the daily capacity of the anticipated number of non-specialists throughout the day. In the treatment-oriented OPDs, we simply use the consultation time in the pacemaker process as the natural choice for the scheduling interval.

We adopted the policy of keeping the number of patients constant at the treatment cell of each scheduling interval. For other cells in the treatment-oriented OPDs, we prioritise the patients going directly to the treatment cell while filling in the cell's remaining capacity with the rest of the patients. Among the patients who remain, the pacemaker patients are scheduled first based on specialist availability. If some pacemaker patients in a scheduling interval cannot be scheduled because of upstream capacity issues, these pacemaker patients are then shifted to a time slot that is earlier in the day so their consultations with specialists will not be delayed. The strategy that is proposed to obtain the patient schedule is described below.

Given the number of scheduling interval I, the daily upstream patients, d^{US}, the daily downstream patients, d^{DS}, the daily home-downstream patients, d^{HD}, and the pacemaker patients of each scheduling interval, d_i^{PM}(see Figure 3.3), we must identify:

i. The home-downstream demand in scheduling interval i, d_i^{HD}.
ii. The home-upstream demand in scheduling interval i, d_i^{US}.

The steps to obtain these variables are as follows.

Step 1) Divide the daily downstream demand d^{DS} from the demand map into scheduling intervals and calculate the downstream demand in scheduling interval i, d_i^{DS}, as follows:

$$d_i^{DS} = \frac{d^{DS}}{I}. \quad (3.8)$$

Any fractional numbers are carefully managed by rounding the numbers up and down during scheduling.

Step 2) Divide the daily home-downstream demand d^{HD} and compute the home-downstream demand in scheduling interval i, d_i^{HD}, as follows:

$$d_i^{HD} = \frac{d^{HD}}{I}. \tag{3.9}$$

Step 3) After computing d_i^{DS} in Step 1) and d_i^{HD} in Step 2, obtain the pacemaker-downstream demand in scheduling interval i, d_i^{PD}, as follows:

$$d_i^{PD} = \sum_k d_{k_i} \tag{3.10}$$

$$\sum_i d_{k_i} = d_k \tag{3.11}$$

$$\sum_k d_{k_i} \leq d_i^{DS} - d_i^{HD}, \tag{3.12}$$

where k is the number of patient types that go to the downstream process through the pacemaker in scheduling interval i, d_{k_i} is the number of patients in k^{th} patient type in scheduling interval i, and d_k is the number of patients in k^{th} patient type per day based on the demand map. Per the policy we applied, the pacemaker-downstream and home-downstream patients are prioritised in the schedule.

Step 4) Calculate the upstream-downstream demand in scheduling interval i, d_i^{UD}, as follows:

$$d_i^{UD} = d_i^{DS} - d_i^{HD} - d_i^{PD}, \tag{3.13}$$

where d_i^{DS}, d_i^{HD}, and d_i^{PD} are obtained via Steps 1 to 3.

Step 5) Compute the upstream capacity of each scheduling interval c_i^{US} by

$$\mathrm{c_i^{US}} = \frac{\text{Effective working time}}{\text{ECT}}, \tag{3.14}$$

where ECT is the effective cycle time in the studied cell, and $\mathrm{c_i^{US}} = \max\{\mathrm{d_i^{US}}, \forall\ \mathrm{i}\}$.

Step 6) identify the pacemaker patients that exceed the upstream capacity in scheduling interval i and need to be shifted, $\Delta_\mathrm{i}^{\mathrm{PM}}$, as follows:

$$\Delta_\mathrm{i}^{\mathrm{PM}} = \max\{\mathrm{d_i^{PM}} - (\mathrm{c_i^{US}} - \mathrm{d_i^{UD}}), 0\}, \tag{3.15}$$

where $c_i^{US} - d_i^{UD}$ represents the remaining upstream capacity after the upstream-downstream demand is satisfied. When the pacemaker demand exceeds the remaining upstream capacity, $\Delta_i^{PM} > 0$, and these patients are then shifted to an earlier scheduling interval.

Step 7) Compute the upstream demand, d_i^{US}, in scheduling interval i as follows.

$$d_i^{US} = d_i^{UD} + d_i^{PM} - \Delta_i^{PM}. \tag{3.16}$$

At the scheduling interval that accepts the shifted demand, the Δ_i^{PM} is a positive number; whereas, at the earlier scheduling interval to which patients are shifted, the Δ_i^{PM} is a negative number.

Figure 3.3 summarizes the procedure described herein. The home-downstream demand, d_i^{HD}, and the home-upstream demand, d_i^{US}, are the two determining factors for the patient schedule in an treatment-oriented OPD and $(d_i^{PM} - d_i^{PD})$ represents the pacemaker-discharged patients. If the critical side of the process is that of the upstream, a similar scheduling logic will be applied; however, instead of repairing the downstream demand, d_i^{DS}, the upstream demand, d_i^{US}, is constant for each scheduling interval.

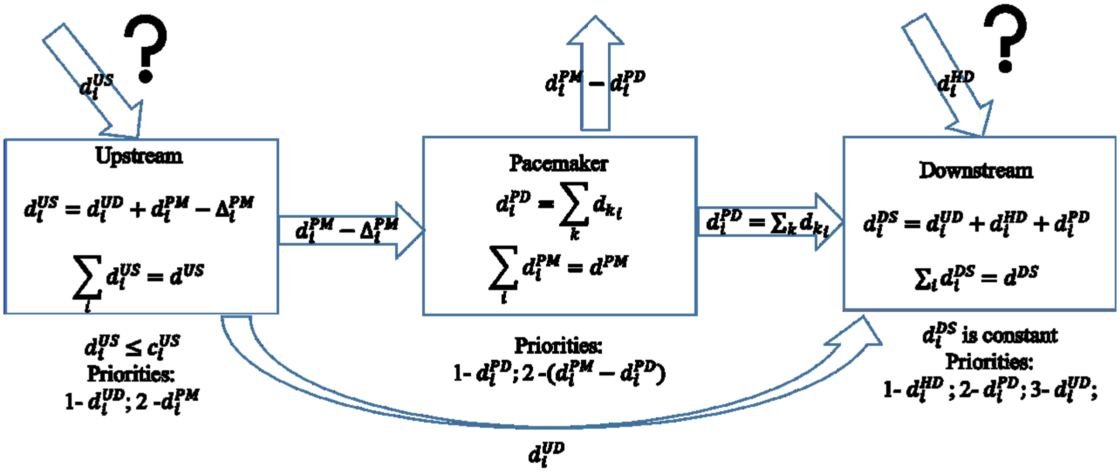


Figure 3.3 Patient Scheduling Logic

The patient schedule designed here serves as a template that treatment-oriented OPDs may use to assign appointment times for their patients. After a treatment-oriented OPD receives patient requests, starting with the first scheduling interval, patients who meet the criteria (patient type) in the template are scheduled with an identical appointment time. The treatment-oriented OPD then moves to the next scheduling interval, and schedules patients in the same manner. The procedure is repeated until all available times in the daily schedule are occupied.

3.2.4 Controlling waits

The patients themselves are the resources that are transformed in the healthcare context (Roemeling et al., 2017). In treatment-oriented OPDs, since cells typically run differently, the patient wait time, i.e. the time buffer, is created when a cell works more slowly than that which proceeds it. This time buffer requires control mechanisms to perform monitoring functions. In Lean, one control mechanism is FIFO, which is a controlled time buffer that serves as a sort of waiting area between two processes. FIFOs represent a temporary wait for a patient. According to Duggan (2018), the size of a FIFO, measured in units of time, is calculated as follows:

$$FIFO\ size = \left(TT_s - TT_p\right) \times d_i, \tag{3.17}$$

where TT_s is the TT (or the highest ECT) of the succeeding cell, TT_p is the TT (or the highest ECT) of the proceeding cell, and d_i is the number of patients of each scheduling interval. It should be noted that C/T (ECT) in such equations is always the C/T of the process/cell. FIFO lanes control the patient wait time and keep patients in the right sequence.

3.2.5 Future state of Lean treatment-oriented OPDs

The resulting VSM is shown in Figure 3.4, which demonstrates a future state of a Lean treatment-oriented OPD. The daily patient schedule includes two components: the patient list and arrivals. These components communicate with the department through a scheduling method known as Offset Sequencing (Duggan, 2018). Since patients must arrive earlier to receive pre-consultation services, such as bloodwork, there is a gap of time between the offset point (first process) and the pacemaker, known as the offset time, which must allow patients to go through all pre-processes before they reach the pacemaker.

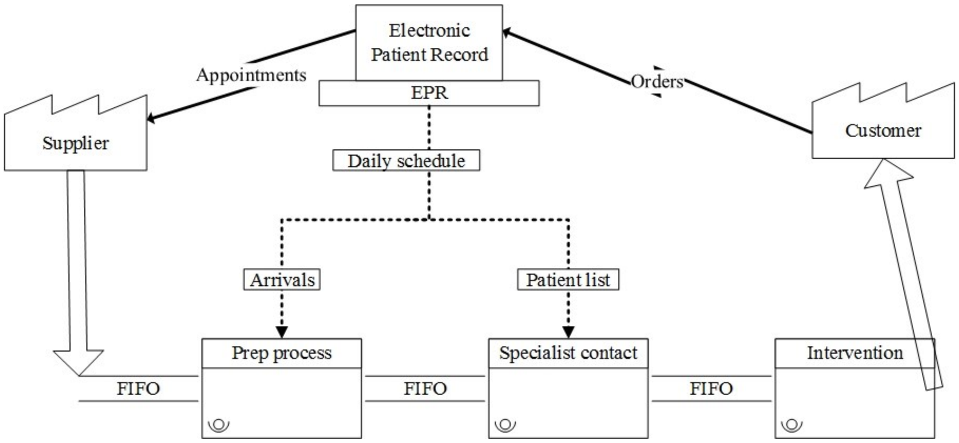


Figure 3.4 Future VSM for Treatment-oriented OPDs

3.3 Case Study

A case study was conducted in an outpatient oncology department at a local community hospital, which services approximately 300,000 outpatients per year. The department facilitates 100 to 250 appointments for oncologists each day and provides 60 to 80 chemotherapy treatments. The department is under pressure to meet the growing demand for cancer care while also maximizing its resource utilization and minimizing patient wait times.

3.3.1 Current system analysis

The current patient flow is described as follows. A patient first registers at the front desk with a receptionist. The receptionist then refers the patient to an adjacent room for bloodwork, at which point a second registration is required. As the blood sample is being analysed, the patient's weight and height (body mass index, BMI) are measured. It is not unusual for the patient to be called by an oncologist before the results of the bloodwork are available. Upon the completion of the consultation, this patient then goes to the pharmacy and registers for chemotherapy medications. Finally, once the medications are ready, a nurse calls the patient, asks him or her to register at the chemotherapy station, and then administers the medications. The current system is a two-day system in which all activities preceding the visit to the pharmacy are completed on day one and the rest are carried out on day two. Three types of patient visits may occur as follows: a first visit, a follow-up visit, and a cyclic visit. A new patient visiting the department for the first time is expected to undergo the entire process. A follow-up visit is typically complete after the patient has seen an oncologist. A cyclic patient has booked a series of appointments, mainly for chemotherapy treatments, but this patient may or may not have bloodwork done or see an oncologist during every visit.

The process flow of patients is drawn in a VSM to allow further system diagnosis (Figure 3.5). The information contained in this map was collected through direct observations, interviews with staff, and data extraction from historical patient records. The data collected were analysed using Arena Analyzer, the averages of which are shown in Figure 3.5.

As Figure 3.5 illustrates, major delays happen before patients go to the lab, have their oncologist consultation, visit the pharmacy, and receive their chemotherapy treatment. A blood sample is sent to the lab through a pneumatic tube and over two hours elapse before results are available. Without the results of their bloodwork, patients sometimes wait longer before seeing an oncologist. In the VSM, the most critical issues in the system are the indeterminate wait icon following the consultation and the overnight icon before chemotherapy registration. The indeterminate wait icon corresponds to a patient waiting list. Patients on this waiting list are new patients who cannot initiate their chemotherapy treatments because the patient mix in the consultation process is imbalanced. This imbalance is due to a disproportionately high number of new patients entering the system. These excess new patients are placed on the waiting list, where they must remain for an undetermined length of time, hence the indeterminate wait icon. The actual patient mix that can be supported by the current system is calculated by eliminating these excess new patients. However, the pharmacy, which is the current bottleneck, is still unable to support the correct patient mix and must schedule overtime to cope with the daily demand. The necessity of overtime combined with a lack of prioritisation in the pharmacy to support the patient chemotherapy schedule cannot guarantee same-day treatment. Hence, the oncology department opted for a two-day treatment regime, in which patients with consultations on day one are scheduled for treatment on day two, hence the overnight icon. The total AWCT is 147.4 minutes versus the total average weighted lead time (AWLT) of 264.4 minutes, which leads to an efficiency factor ρ of 55.7%. The overnight waiting is not counted.

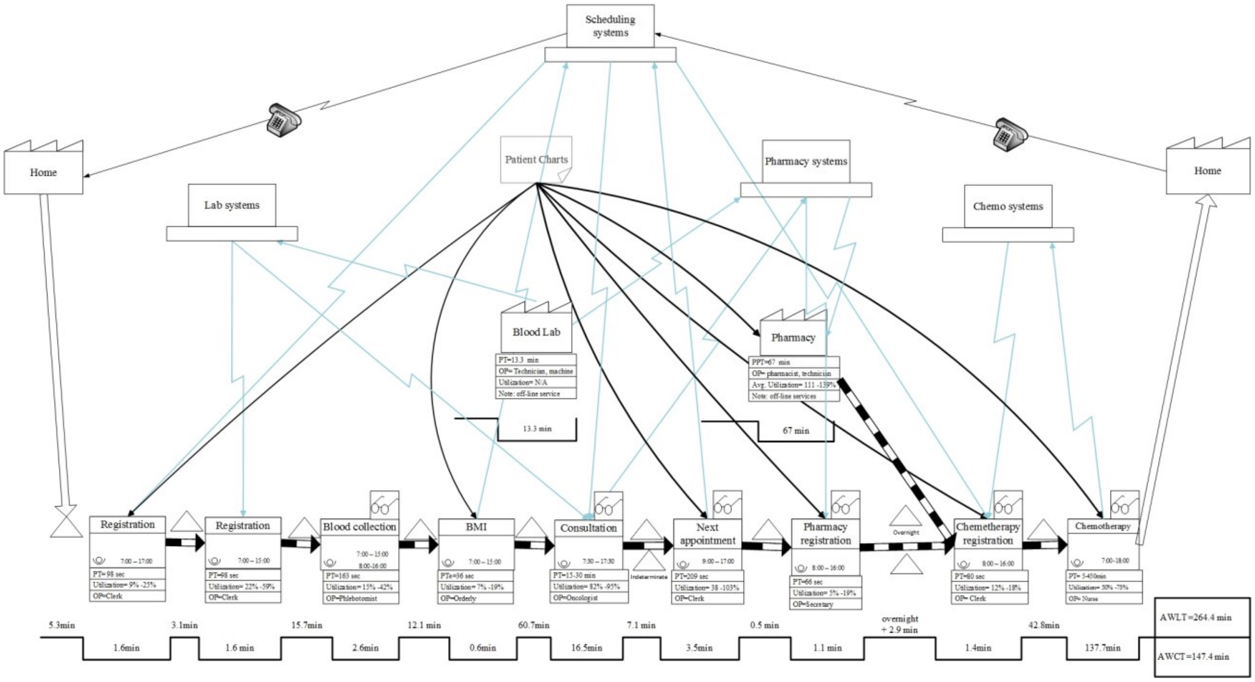


Figure 3.5 Current VSM of the Oncology Department

Given the above-mentioned issues in this oncology department, we apply the proposed methodology to identify the best patient mix, schedule, and resource levels to reduce patient wait times and improve resource utilization.

3.3.2 Lean interventions

3.3.2.1 Identifying demand

Based on our proposed framework, we analyse the current demand to identify the patient composition in the consultation and treatment cells and propose a new patient mix.

There are five clinical cells and five patient types in this oncology department. The cells comprise the registration (which includes all registrations), the bloodwork (which includes blood collection, BMI measurements, and bloodwork analysis), the consultation, the pharmacy, and the chemotherapy cells. The five patient types are new, follow-up, cyclic-A, cyclic-B, and cyclic-C patients.

To find the patient composition, we start with the chemotherapy cell. Currently, an average of 70 patients is scheduled for chemotherapy treatments per day, of which 16.7 patients (23.9%) are cyclic-A combined with new patients; 37.4 patients (53.3%) are cyclic-B patients; and 16.0 patients (22.8%) are cyclic-C patients. The daily number of new patients that are accepted for chemotherapy treatment was unavailable; however, as only discharged patients free up the chemotherapy resources, the daily number of new patients equals that of discharged patients. The number of discharged patients is calculated by dividing the total number of chemotherapy patients by the average weighted length of chemotherapy treatment, and considering the type of chemotherapy, number of chemotherapy cycles, and number of treatments per cycle. The number of discharged patients is calculated as 4.4 patients. Hence, the number of new patients is 4.4 and the number of cyclic-A patients is 12.3. As such, the composition of chemotherapy patients is 6.2% new patients, 17.7% cyclic-A patients, 53.3% cyclic-B patients, and 22.8% cyclic-C patients, on average.

A total of 493.5 patients are currently scheduled for oncology consultations per week, 130 of whom are new patients; 26 are seen per day, on average. The group of 26 new patients is five times larger than that of 4.4 new patients, the latter of whom have received all their chemotherapy treatments and are discharged. Thus, 21.6 patients are placed on the waiting list for chemotherapy. This causes a continuous accumulation of patients in the system, which leads to long wait times for these patients. This is represented by the indeterminate wait icon following the consultation process in Figure 3.5. The current oncology department is therefore imbalanced in terms of the number of patients scheduled per patient type.

Once patients finish their chemotherapy treatments, these patients are then classified as follow-up patients and revisit the department for periodic oncologist consultations. Now, we trace the demand back from the chemotherapy cell to that of the pacemaker. As mentioned above, 22 (4.4 daily) new patients and 61.5 (12.3 daily) cyclic-A patients go from the pacemaker to chemotherapy on a weekly basis. Combined with 302 follow-up patients, we find that the total weekly adjusted pacemaker demand that is synchronised with that of chemotherapy is 385.5 patients. We then find that the adjusted composition of pacemaker patients is 5.7% new, 16.0% cyclic-A, and 78.3% follow-up patients, respectively. The 21.6 new patients are excess patients and are not included in the calculation, as they have not yet initiated their chemotherapy treatments.

After identifying the patient composition at the pacemaker and chemotherapy cells, to identify the new mix of patients, we adopt a policy to keep the daily number of new and cyclic-A patients in the pacemaker constant and fill in the remaining pacemaker capacity (on different days of the week) with the follow-up patients. We adopt the policy of keeping the daily number of new and cyclic-A patients in the pacemaker constant primarily to maintain a constant demand on each day at the chemotherapy cell to ensure that the staffing level can remain consistent throughout the week. The stepwise procedure used to accomplish this is as follows.

Step 1) Find the weekly pacemaker capacity per Equation (3.2) and compute the number of weekly new and cyclic-A patients using the weekly pacemaker patient composition. According to the policy we adopted, we divide these numbers equally into five days to find the daily number of new and cyclic-A patients at the pacemaker process; the same number of new and cyclic-A patients will be served by the chemotherapy cell on a daily basis.

Step 2) Find the weekly number of follow-up patients from the number of new patients calculated above and the pacemaker patient composition. Given the weekly follow-up patients, we obtain the daily number of follow-up patients by filling the remaining capacity after scheduling the new and cyclic-A patients to be seen by oncologists first and distributing the remaining follow-up patients into each day of the week according to the daily pacemaker capacity.

Step 3) Obtain the daily number of cyclic-B and cyclic-C patients to be scheduled for chemotherapy from the daily number of new patients and the patient composition in the chemotherapy cell. Since we kept the daily number of new patients constant, the daily number of cyclic-B and cyclic-C patients will also be constant.

We now demonstrate the procedure used to identify the new patient mix: following Step 1, we calculate the pacemaker capacity using Equation (3.2) based on oncologist availability in terms of the number of consultation hours per day (1470 minutes, 2850 minutes, 2280 minutes, 2130 minutes, and 960 minutes from Monday to Friday) and a treatment time of 18.8 minutes. This gives us 78, 151, 121, 113, and 51 patients from Monday to Friday, which totals 514 for the week. We then use the pacemaker's weekly capacity to obtain the weekly number of new and cyclic-A patients, by which we compute the weekly number of follow-up patients using the chemotherapy patient composition. The weekly new and cyclic-A patients are divided by five, as stated in Step 1, while the follow-up patients are distributed according to Step 2, which gives us six new and 17 cyclic-A patients per day at the pacemaker. Finally, following Step 3, we calculate the daily number of cyclic-B and cyclic-C patients at the chemotherapy cell by chemotherapy patient composition and the daily number of new patients.

The resulting patient mix is demonstrated in the suggested demand map, as shown in Figure 3.6, which illustrates the flow of patient types and their demand mix. The arrow with a dashed line following the consultation cell indicates the discharged follow-up patients; the arrow with a dashed line following the chemotherapy cell indicates the discharged chemotherapy patients; and the solid arrows indicate the flow of the remaining patients.

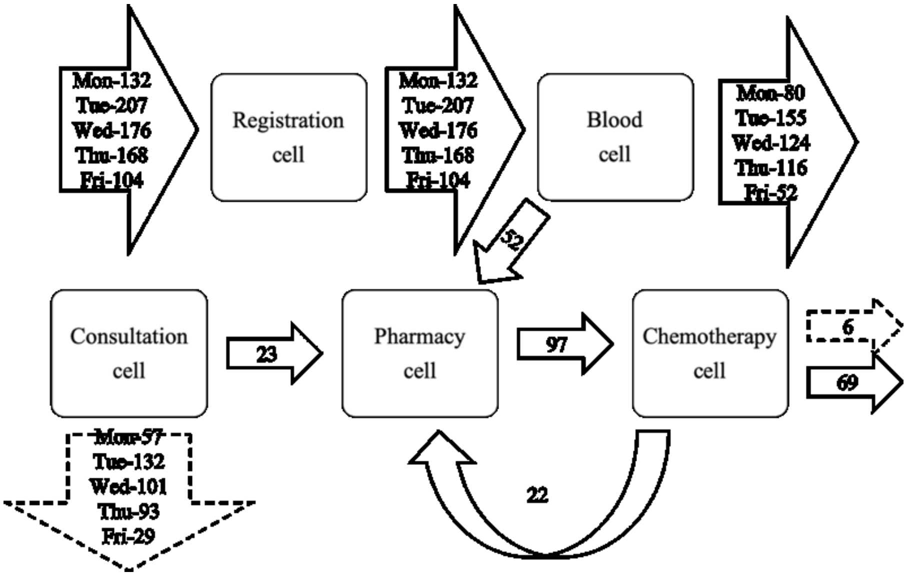


Figure 3.6 Demand Map of the Oncology Department

3.3.2.2 Coordinating resources

Given the daily demand, to verify whether the number of resources in the oncology department supports the proposed demand mix in the value stream, we create an intermediate BC and a future BC for each cell. The intermediate BCs use the new balanced patient mix identified in Figure 3.6 but keep the same staffing levels, whereas the future BCs reflect the proposed staffing levels identified by closing the gap between the C/Ts of resources and the TT of a cell, as explained in Chapter 3.2.2.

Figure 3.7 shows the intermediate and proposed BCs of the registration cell. The five receptionists in the intermediate BC currently work at three different locations to conduct registrations for consultations, bloodwork, and chemotherapy. The digitalization of the registration process renders the integration of the receptionists' tasks possible at one site, and we therefore present all five receptionists in the same registration cell. In the intermediate BC, all five receptionists have a much lower C/T than the TT of the cell. This indicates that the cell could perform with a fewer number of receptionists by combining their tasks. After combining the clerical tasks, the TWC is 8.1 minutes. Furthermore, based on the consultation with the involved staff, following digitalization, receptionists could save at least 30 percent of their total time, which leads to a new TWC of 5.7 minutes. Because the TT of this cell varies daily due to variations in demand, the ECT is set to the lowest TT of the week, since the staffing level is considered constant for non-specialists. According to Equation (3.6), the number of receptionists needed is 2.7, given the lowest TT of 2.3 minutes on Tuesday. By rounding 2.7 receptionists down to two receptionists, we obtain an ECT of 2.9 minutes. This ECT exceeds the TT of the cell on Tuesday, Wednesday, and Thursday. Thus, we round the 2.7 receptionists up to three receptionists and again compute the ECT, which is 1.9 minutes. The proposed BC in Figure 3.7 shows that the ECT is below the TT on every day of the week. While the ECT approximates the TT on Tuesday, it is well below the TT for Monday and Friday.

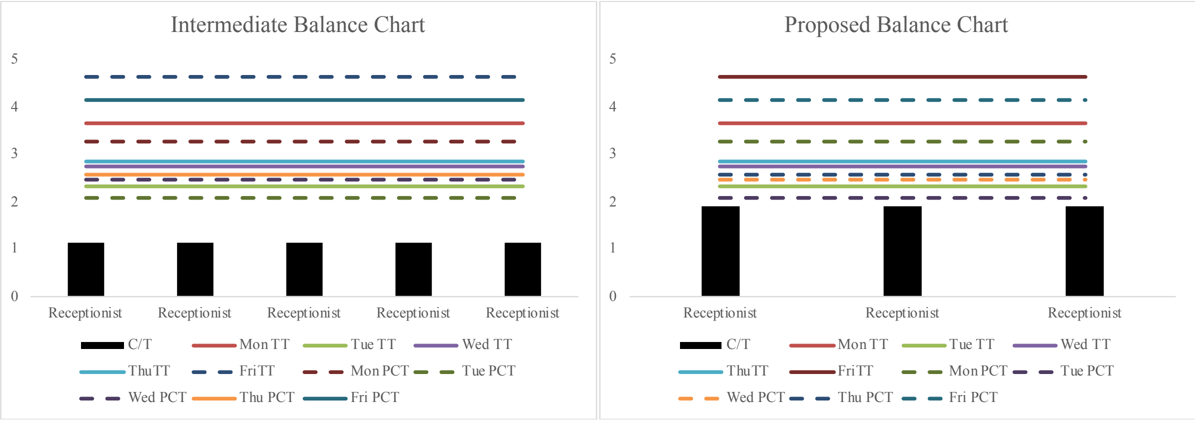


Figure 3.7 Balance Charts of the Registration Cell

The balance chart of the consultation cell, or the pacemaker process, is shown in Figure 3.8. Because the daily number of patients and the daily patient mix are different, Oncologists' AWCT, PCT, and the TT on each day of the week are different. As it can be observed in Figure 3.8, the gap between the TTs and the C/Ts is 86.9 percent on average ranging from 85 percent to

90 percent from Monday to Friday; this gap is left to absorb minor interruptions and/or system variations.

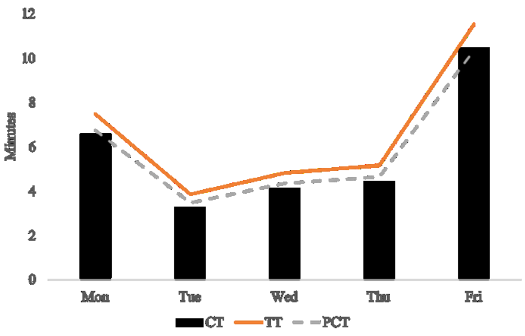


Figure 3.8 Weekly Balance Chart of Consultation Cell

Figure 3.9 illustrates the intermediate balance chart of the pharmacy. Previous TT is calculated using 70 chemotherapy patients. As it can be seen in Figure 3.9, the pharmacy cell is a bottleneck in the current system because the C/T of technicians B is higher than the previous TT. The oncology department handles this bottleneck by working overtime as we mentioned in Chapter 3.3.1 but clearly, to support the new patient mix, resources at the pharmacy cell needs to be re-evaluated.

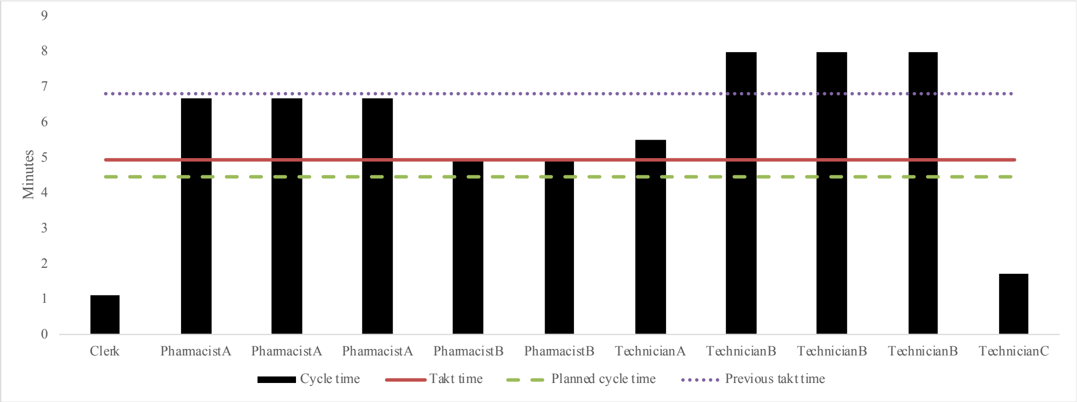


Figure 3.9 Intermediate Balance Cell of Pharmacy

With the new patient mix, the C/Ts of pharmacists A, pharmacists B, technicians A and technicians B are all higher than the TT of the cell. In order to bring the C/Ts below the TT, the number of required resources was recalculated according to Equation (3.6). The results are shown in Table 3.1 where m' is the feasible number of resources. Since the number of required pharmacists A, B, and technicians A were rounded up, they should all be targeted for Kaizen to eliminate waste and reduce these staffing levels. As for technicians B, since the number of required technicians was rounded down, reduction in C/T through Kaizen is needed to bring the effective C/T below the planned C/T. The proposed balance chart is shown in Figure 3.10. Next, we discuss the chemotherapy cell.

Table 3.1 Resource Summary-Pharmacy Cell

	TWC	**PCT**	m	m'
Pharmacist A	20 minutes	4.5 minutes	4.5	5
Pharmacist B	10 minutes		2.2	3
Technician A	5.5 minutes		1.2	2
Technician B	24 minutes		5.4	5

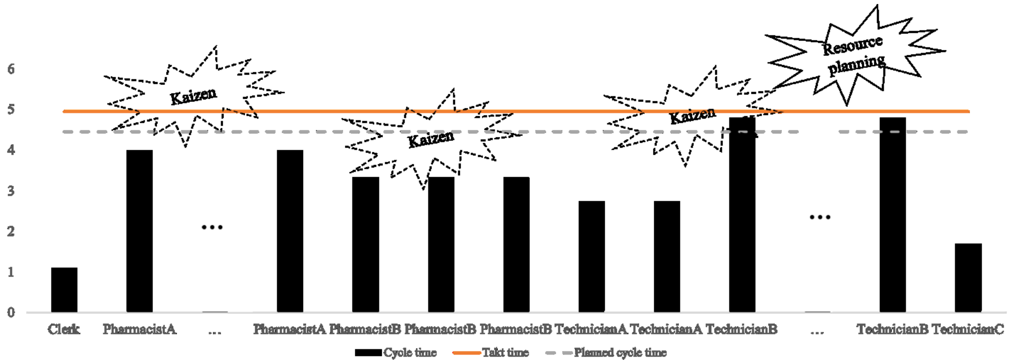


Figure 3.10 Proposed Balance Cell of Pharmacy

The improved patient mix increases daily chemotherapy treatments from 70 patients to 97 patients. The TT becomes 6.2 minutes and the average weighted treatment time is 137.7 minutes. With 35 chairs, the chemotherapy cell has a C/T of 3.9 minutes per Equation (3.6). As shown in Figure 3.11, the C/T is about 63.7 percent of the TT. The gap between the C/T and the TT indicates that precious chemotherapy capacity is being wasted and more patients can be treated in the chemotherapy cell. However, the consultation cell, which is the most critical resource in this oncology department, cannot accept more patients due to shortage of oncologists.

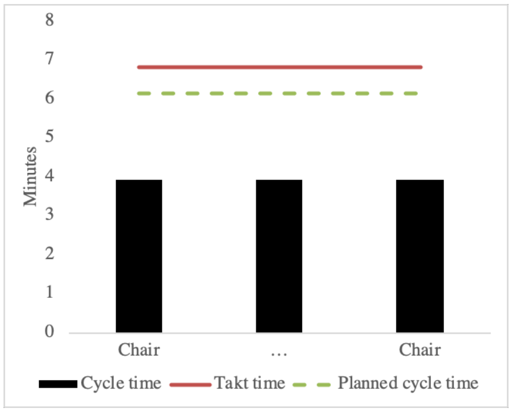


Figure 3.11 Balance Chart of Chemotherapy Cell

3.3.2.3 Levelling patient schedule

After the proper number of operators is obtained to meet the daily demand, we establish a patient schedule in which the daily workload of operators is levelled out. We select a scheduling interval of 15 minutes, which is close to the AWT at the pacemaker. The total pacemaker capacity fluctuates due to oncologists' schedules. To obtain the patient schedule, we apply our levelling strategy, as described in Chapter 3.2.3.

Figure 3.12 to Figure 3.14 show the patient schedules after levelling is performed. According to our scheduling policy applied in Chapter 3.2.3, chemotherapy patients with/without consultations are prioritised in the schedule. Figure 3.12 shows the daily patient schedule for each day of the week for the registration and bloodwork cells (upstream processes) in each scheduling interval after levelling is performed. For example, from 7:00 am to 8:00 am on Tuesday, two patients are scheduled at the beginning of each scheduling interval, followed by a scheduling interval of three patients; the remainder of the day is scheduled with seven patients in each scheduling interval, except for 11:30 am to 12:00 pm, during which only six patients are scheduled to avoid exceeding the capacity (7 patients) of the upstream processes. Figure 3.13 shows a detailed upstream patient schedule, which includes the patient quantity as well as the patient type for Tuesday. For example, as shown in Figure 3.13, at 7:00 am on Tuesday, we schedule two patients, one cyclic-B patient and one consulting patient classified as a new patient, a cyclic-A patient, or a follow-up patient. Similarly, at 8:00 am on Tuesday, we schedule three consulting patients; one of the patients is a pacemaker patient who has been shifted from his or her original appointment time of 9:30. Figure 3.14 shows the detailed patient schedule that the pharmacy and chemotherapy cells (downstream processes) will receive in each scheduling interval after levelling is performed on Tuesday. In Figure 3.14, in each scheduling interval, the appointments are only given to patients coming directly from home (the home-downstream demand), and the remaining patients are expected to arrive at the downstream from either the pacemaker or the upstream processes. For example, at 8:00 am, one cyclic-C patient (the home-downstream demand) is scheduled, and two other patients (one is a new patient or a cyclic-A patient and the other is a cyclic-B patient) are simultaneously expected to arrive at the downstream processes. In Figure 3.14, the number of patients in each scheduling interval in the downstream is constant throughout the week due to the policy explained in Chapter 3.2.3, but the types of patients in each scheduling interval change according to the daily schedule of oncologists.

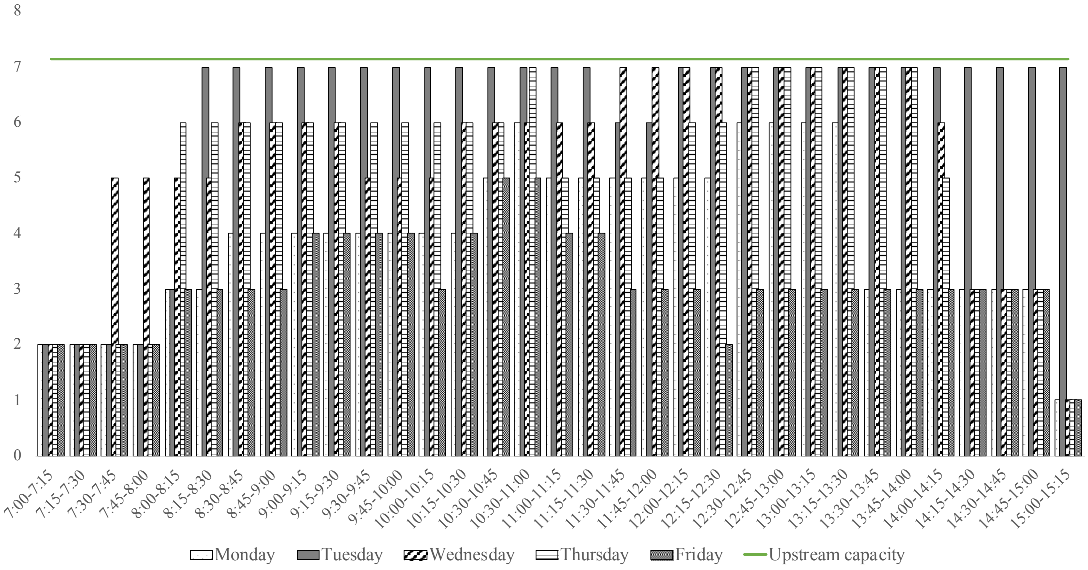


Figure 3.12 Weekly Upstream Patient Schedule after Levelling is Performed

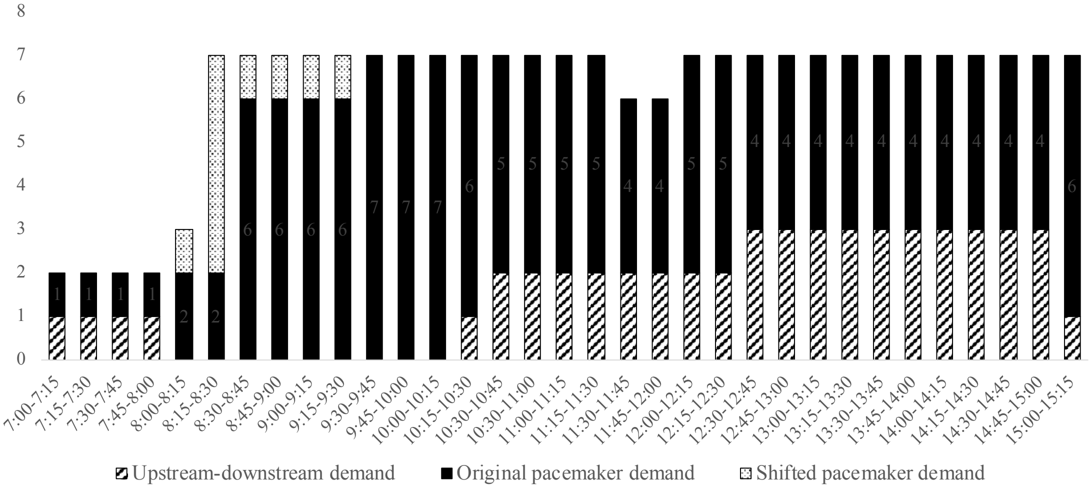


Figure 3.13 Detailed Upstream Schedule for Tuesday after Levelling is Performed

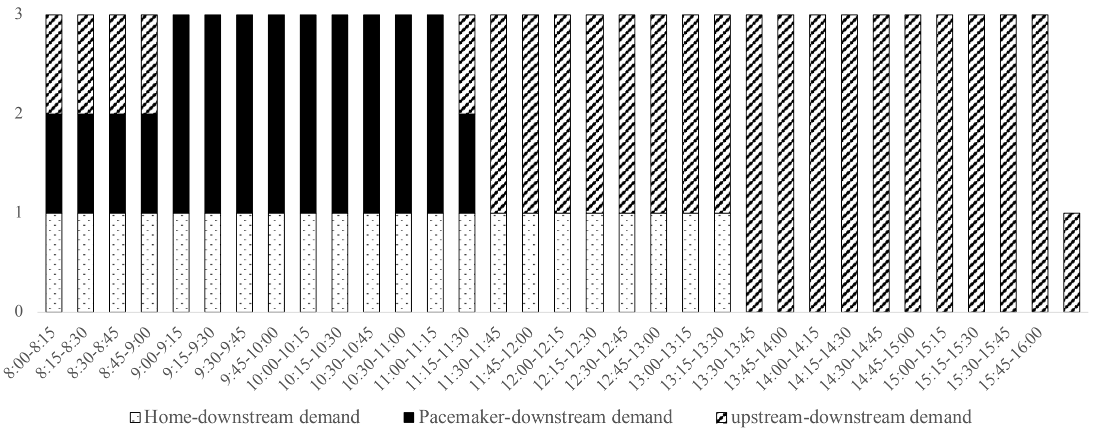


Figure 3.14 Detailed Downstream Patient Schedule for Tuesday after Levelling is Performed

Currently, the oncology department uses a two-day treatment regime, mainly because the pharmacy cell does not support the one-day treatment regime, as mentioned in Chapter 3.3.1; however, after coordinating the pharmacy resources and improving the patient schedule, applying the one-day treatment regime is possible if the schedule for the chemotherapy cell is also considered in the resulting patient schedule. Therefore, to redesign the system into a one-day structure, we schedule chemotherapy patients with/without consultations into a configuration of the longest-treatment-first (LTF) when assigning the appointment times to patients. The rule of LTF serves two purposes: first, it maximizes the chances that those patients will be treated before the shift ends, and second, longer treatments will occupy the chairs for a longer period at the beginning of the shift and allow the chemotherapy cell to fill all the chairs more quickly. Figure 3.15 illustrates the distribution of treatments by chair. Treatment 13 (T13) is the longest treatment time and treatment 1 (T1) is the shortest. Different bars represent the treatment times of 97 patients. Since the chemotherapy cell is the most downstream process, in a one-day system, this cell requires a warm-up period, the period until all chairs are occupied. Figure 3.15 does not show the warm-up period, but is a close approximation of the actual utilization of chairs. Further, the schedule for the chemotherapy cell is shifted ahead by an hour since the pharmacy cell needs approximately one hour to prepare a medication.

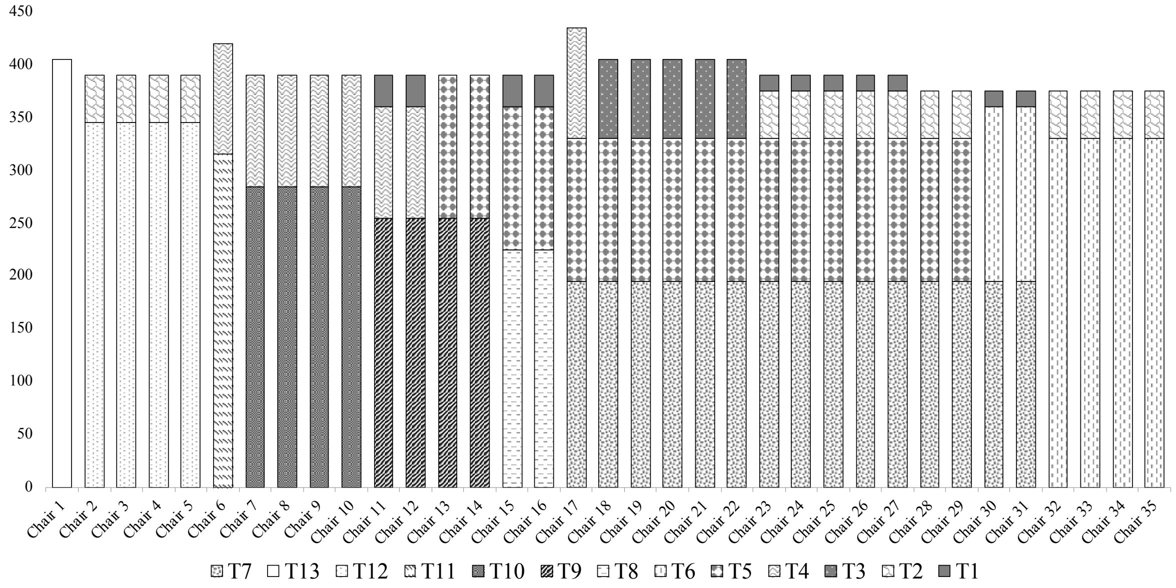


Figure 3.15 Chemotherapy Treatment Distribution

3.3.2.4 Controlling waits

Five FIFOs are used in this oncology department, each of which presents the waiting area of its succeeding cell. On average, the FIFOs of the registration, the bloodwork, consultation, pharmacy, and chemotherapy cells are 8.6 minutes, 0 minutes, 15 minutes, 7.1 minutes, and 5.7 minutes, respectively, according to Equation (3.17). Since we applied an offset time of one hour, the FIFO size of the consultation cell extends from 15.0 to 29.3 minutes. According to the processing time of the upstream processes and the sizes of the FIFOs preceding the pacemaker, patients need an average of 45.7 minutes to go through the upstream processes.

3.3.2.5 Future state of the oncology department

We demonstrate the proposed future state map in Figure 3.16, which combines the recommended improvements discussed in the previous chapters. As we can observe from Figure 3.16, the oncology department has one integrated schedule which comprises three sub-schedules, the upstream patient schedule, the pacemaker schedule, and the downstream schedule, all of which are designed in coordination. The upstream patient schedule includes a list of appointment times for upstream patients entering the registration cell based on Figure 3.12. The pacemaker schedule is a list of patient names and their scheduled consultation time with oncologists. The difference between the appointment time, i.e. patient arrival time, and the consultation time is offset by one hour, as mentioned in Chapter 3.3.2.4. The downstream patient schedule specifies the appointment times of patients coming directly to the pharmacy cell and the expected arrival time of patients that come from the bloodwork and consultation cells, in accordance with Figure 3.16. The scheduling is controlled by the patient scheduling system, while the information is communicated to the department through an integrated database. The cells that are not receiving direct scheduling signals, namely, the bloodwork and chemotherapy cells, may also have a list of patients whose attendance must be verified with no time attached to the names. The chemotherapy cell starts one hour later when the first medication comes out of the pharmacy cell. A patient then arrives every 4.9 minutes on average, which is the TT of the pharmacy cell, and fills one of the chemotherapy chairs. The chemotherapy chairs cannot be fully occupied until the warm-up period ends.

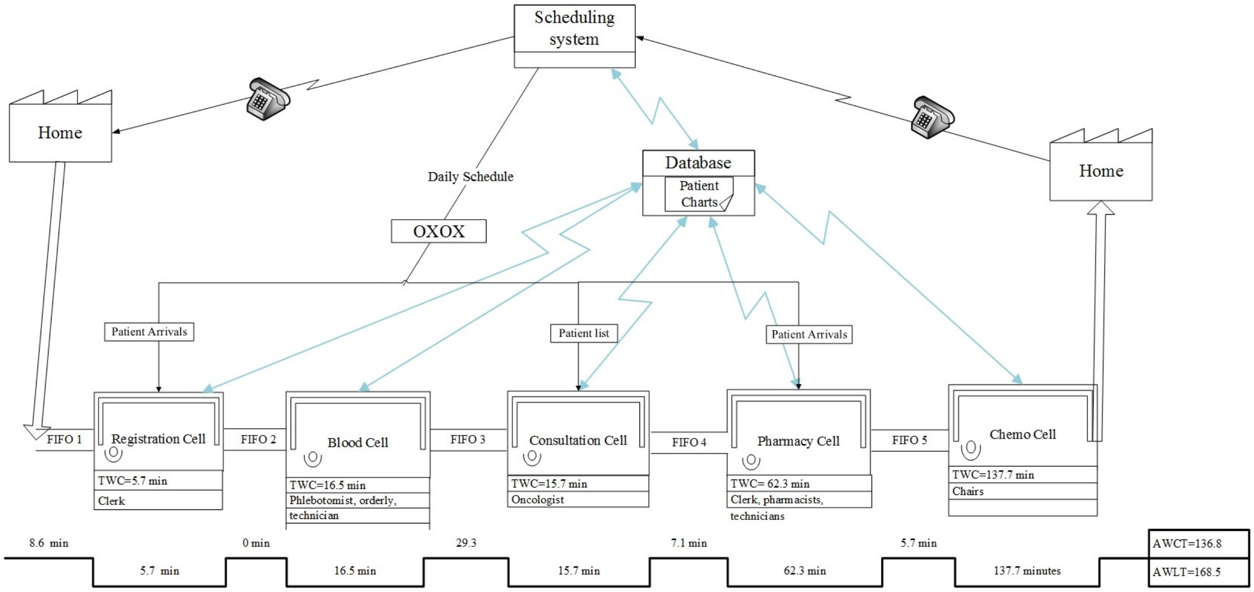


Figure 3.16 Proposed VSM for the Oncology Department

3.3.3 Analysis of results and recommendations

The proposed future state map (Figure 3.16) demonstrates that following Lean interventions, the AWCT is 136.9 minutes and the AWLT is 168.5 minutes, which yields a ρ of 81.2%. Compared to the current system, in which the AWLT is 264.4 minutes, the Lean framework reduces the AWLT by 36.3%. Additionally, the number of chemotherapy treatments is increased from 70 to 97 patients per day, an increase of 38.6%. Meanwhile, the prioritisation policies applied in the proposed patient schedule make the one-day treatment regime feasible since the consultation and pharmacy cells now prioritise patients according to the chemotherapy schedule.

Based on the analysis presented in Chapter 3.3.2.2, four additional technicians are needed in the oncology department to support the one-day treatment regime. Furthermore, the upstream resources, which include registration receptionists, phlebotomists, orderlies, and technicians, have low C/Ts compared to the TT, particularly on Monday and Friday, thereby indicating a poor utilization of their time on these days. This is one drawback of using a constant staffing level, which leads to a higher number of staff members than is needed on some days of the week. One suggestion entails toggling the resources based on the daily demand. For example, on Friday, the required number of registration receptionists is only two instead of three. Another suggestion is to manage the working hours of oncologists so their daily availability is consistent; thus the daily demand is maintained at a constant rate. However, specialists typically do not belong to the oncology department or the hospital; thus, controlling their schedules may not be feasible at this point. Future collaborations between specialists and the hospital regarding managing the schedules of the former may resolve this problem.

We also identified that in a one-day system, with a TT of 4.9 minutes at the pharmacy cell and 35 chairs, the warm-up period at the chemotherapy cell can be over 172 minutes, during which time half of the chairs are unoccupied. One possible way to remedy this circumstance entails gradually increasing the staffing level in the chemotherapy cell to coincide with the number of occupied chairs during the warm-up period. Additionally, although the number of chemotherapy treatments increases, the chairs are only utilized at 63.7% of their capacity. This indicates that more patients can be accepted for treatments. However, the patient demand is determined by the capacity of the oncologists. Given that it is not feasible to increase the number of oncologists at this point, the oncology department cannot accept more patients and therefore cannot improve its utilization of the chairs due to the lack of oncologists.

3.4 Discussion and conclusions

Treatment-oriented OPDs are facing problems of long patient wait times and underutilized resources. While Lean has been applied in healthcare contexts to resolve similar issues, the literature suggests that there is no structured methodology for Lean transformation in healthcare (D’Andreamatteo et al., 2015), particularly in treatment-oriented OPDs. Therefore, in this research, we proposed a systematic framework to guide lean transformation and resolve the problems confronted by treatment-oriented OPDs by identifying demand, coordinating resources, levelling schedules, and controlling wait times.

The imbalanced demand (i.e. patient mix) in treatment-oriented OPDs leads to long patient waiting lists and fewer numbers of scheduled patients. In a traditional Lean system, the demand stems from the customer and is traced back through the value stream for a continuous flow (Radnor et al., 2012). However, the OPDs have an ‘artificial’ demand, which means that the demand is determined by two critical processes involving specialists and treatments compared with by customers in traditional Lean systems. We perceive this as a critical feature that differentiates Lean manufacturing and Lean in specialist-centred departments, which has not yet been mentioned in the literature. Our framework manages this artificial demand by coordinating the two critical processes for the best patient compositions to reduce wait times and exploit specialists’ time to maximize the number of patients. The results of the case study demonstrate that the framework increases the total number of chemotherapy patients and eliminates indeterminate waits in the oncology department studied herein.

Uncoordinated non-specialists could create isolated processes in treatment-oriented OPDs and result in the underutilization of operators and longer patient wait times. Although resource coordination (planning) plays a central role in Lean implementation (Naiker et al., 2018; Powell et al., 2013), the Lean tools used in resource planning, such as balance charts and TT, are not well established in healthcare (Hallam & Contreras, 2018). Without the proper tools, the number of studies regarding resource planning is limited. Among the existing papers on Lean’s application to healthcare, only one by Wang et al. (2015) uses TT to calculate the number of healthcare operators needed. In this research, our framework introduces the three Lean techniques of TT, C/T, and PCT, which are adapted for application to healthcare and support the creation of patient flow by adjusting the daily number of operators based on the demand. The oncology study proves the effectiveness of these techniques in meeting patient demand by supplying the appropriate number of non-specialists and reducing daily patient wait times.

The variability in patient arrivals due to specialists’ daily schedules imposes an imbalanced workload on non-specialists and thus results in patient waits and underutilized resources. Based on Litvak et al. (2005), this hidden variability is both artificial and avoidable if patient appointments are carefully managed. One means of handling this variability using Lean methods is volume balancing, which has been highlighted as a major feature of Lean transformation (Duggan, 2018; R. Shah & Ward, 2007). Like resource planning, volume levelling is also underdeveloped in healthcare due to a lack of proper tools. Thus, our Lean framework, which considers the unique features of treatment-oriented OPDs, proposes a better patient schedule strategy. This patient schedule maintains a constant number of patients throughout the treatment process, prioritises consultation patients in a manner that respects specialists’ schedules, and fills the remaining patient appointment slots based on the remaining capacity of the non-specialists. This schedule offers a balanced flow of patients throughout the day, decreases patient wait time, and improves resource utilization in the oncology department.

The outcomes discussed above demonstrate the ways our framework addresses the major problems confronted by treatment-oriented OPDs with the proper use of Lean tools. Wait time is then monitored through FIFOs and a future VSM is established to demonstrate the resulting improvements. The framework illustrates the steps toward a Lean system following Lean principles defined by Womack and Jones (1997). Identifying the treatment-oriented OPD demand allows the subsequent determination of the customer service rate, i.e. the ideal speed of operation to meet the demand. The resources are then coordinated to run at this speed to create a patient flow that is supported by a levelled patient schedule. As the flow is introduced, scheduled patients can receive the required services when needed with a minimum wait time. The fifth principle of Lean, that of seeking perfection, is not explicitly described in this framework; however, we anticipate that by using this replicable framework, treatment-oriented OPDs can repeat the improvement process until a state of perfection is achieved.

4. Lean Transformation Framework for Shared Processes in Outpatient Departments

4.1 Introduction

There are two defining characteristics of shared processes in the OPDs under discussion. First, such processes usually serve the two competing and different priority groups of elective (scheduled) and urgent (unscheduled) patients without hospital admission. Second, the capacity of such processes is usually limited by a critical resource involving expensive machinery or scarce human resources. For instance, such a critical process can be the CT examination process.

A typical example of shared processes in OPDs is in radiology departments. The radiology department is a critical healthcare service owing to its imperative role in hospital patient care delivery; healthcare operators in many other departments are dependent on the results of diagnostic testing to work on patient treatment plans. The ever-increasing reliance on diagnostic services creates a rapid growth in demand for such services and longer patient wait times (Hitti et al., 2017; Idigo et al., 2019). Delays in receiving diagnostic services contribute to a longer LOS (Kanzaria et al., 2014), delayed cancer diagnosis (Byrne et al., 2015), etc. Meeting demand with supply in a timely fashion while improving resource allocation and patient satisfaction becomes one of the biggest challenges in the daily operations of such departments (De Mast et al., 2011; Holbrook et al., 2016).

Scheduling competing patients for a critical process has a crucial impact on wait times, patient safety, and the efficient use of resources in shared processes in OPDs. Many researchers apply designated timeslots in scheduling patients for routine examinations and match the number of patients/examinations to the capacity of the machines. These researchers either divide the working-hour capacity into two types of timeslots for elective patients and urgent patients (Bhattacharjee & Ray, 2016; Geng et al., 2017; Huang & Marcak, 2013) or pool all patients together to share the capacity (Côté & Smith, 2018; Idigo et al., 2019). The results show improved patient access (usually measured by the daily number of visits), reduced wait times, and increased resource utilisation. However, Ahmadi-Javid et al. (2017) underlined that these two competing patient classes should be separated because urgent cases, mostly coming from the ED, have a higher priority in receiving services than do elective patients. From this perspective, given the ED demand variability (Kadri et al., 2014) on a daily basis, a consequent dilemma is that a large capacity for urgent patients leads to decreased elective patient access and a high probability of resources being wasted; in contrast, a small reserved capacity for urgent patients leads to longer wait times and compromised urgent care. Therefore, a trade-off is required.

It was well recognised in the previous studies that improved scheduling strategies lead to improved resource utilisation (Huang & Marcak, 2013; Patrick & Puterman, 2007); however, little is known regarding how the improved scheduling impacts the use of resources other than the critical resource. For example, in the study conducted by Huang and Marcak (2013), an 83% increase in the utilisation of technicians was reported since the number of exams increased as a result of rescheduling at the examination process. In addition to the efficient use of critical resources, it is of immense importance to identify proper non-critical resource levels along the entire value stream, such as the number of registration clerks needed, to support the work of critical resources, further reduce wait times, and meet the demand in a timely fashion.

Given the ability to improve performance and create flow, Lean has been applied in the context of shared processes in OPDs. For example, MacDonald et al. (2013) used Lean to reduce the waste hidden in the value stream of a radiology department, resulting in decreased wait times for elective patients and better services for emergency and hospitalised patients. Another study by Kruskal and others (2012) describes the use of several Lean techniques such as VSM and A3 in a radiology department. The reported results are a better understanding of the patient flow and reduced non-value-added activities. Similarly, Shah et al. (2013) conducted a process improvement project using a Lean technique of VSM. The VSM demonstrates major steps in the radiology process and was later used for the waste identification and workplace organisation. An approximately 70% decrease in wait times was achieved with improved operator utilisation. In a recent study, Lean was declared to be effective in not only wait time reduction but dramatic cost savings as well as employee satisfaction in radiology departments (Baccei et al., 2020). Many other Lean studies with regard to shared processes in OPDs have been conducted by Karstoft and Tarp (2011), Cheung et al. (2016), Verbano and Crema (2019), etc.

Not surprisingly, Lean implementations in shared processes in the OPDs are tool-oriented and lack structured methodologies. Aiming to overcome the challenge of meeting a growing demand with supply in a timely fashion while improving resource allocation and patient satisfaction, this chapter proposes a systematic Lean framework for shared processes in the OPDs. The goal of this framework is to better manage the shared resources to facilitate the work of other departments and reduce overall patient wait times and thus contribute to hospital-wide patient flows. This framework, which includes a series of activities and adapted Lean tools following the Lean principles, enhances performance and creates patient flow by improving patient scheduling and staffing strategies. The application of the proposed framework is demonstrated in a radiology department at a local community hospital in Montreal, Canada.

4.2 Lean framework for shared processes in OPDs

The main purpose of this study is to develop a Lean framework adapted to shared processes in OPDs. A shared process in an OPD could provide multiple services; each service operates independently under its own schedule. The critical resource in a value stream of one service functions as the pacemaker for such a Lean system. A framework for the Lean transformation of shared processes in OPDs per service is established and illustrated in Figure 4.1.

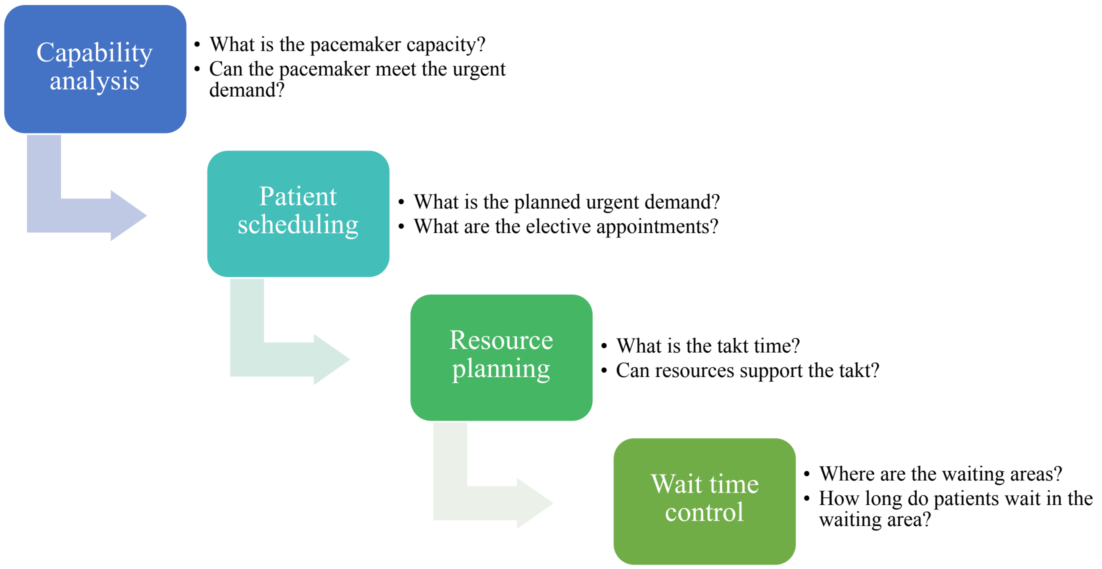


Figure 4.1 Lean Transformation Framework for Shared Processes in OPDs

As shown in Figure 4.1, this framework is composed of a capability analysis, patient and resource scheduling, and wait time control. In the framework, it is first determined if the pacemaker has a sufficient capacity to meet the expected urgent demand, which is followed by scheduling elective patients in coordination with the arrivals of urgent patients under a desired service level. Then, based on the total patient demand, staffing levels are adjusted to run at a proper speed to meet the demand and create patient flow. Wherever wait times cannot be eliminated, they are controlled throughout the value stream. Details are provided below. Note that according to Lyon (2014), the daily urgent demand is assumed to be normally distributed in this study.

4.2.1 Capability analysis

A priority of shared processes in OPDs is to offer timely services to urgent patients. Thus, we investigate the probability of the critical resources meeting an urgent demand without delay by comparing the urgent demand with the pacemaker capacity. Given a varying urgent demand, we match it more closely by dividing it into scheduling intervals (SIs). The mean demand is considered constant in one interval but varies between intervals. A common practice for selecting the scheduling interval is to use a given target. For example, 2 hours is Quebec's target of the time from a request to the completion of an emergency CT scan (Ministère de la santé et des services sociaux, 2006).

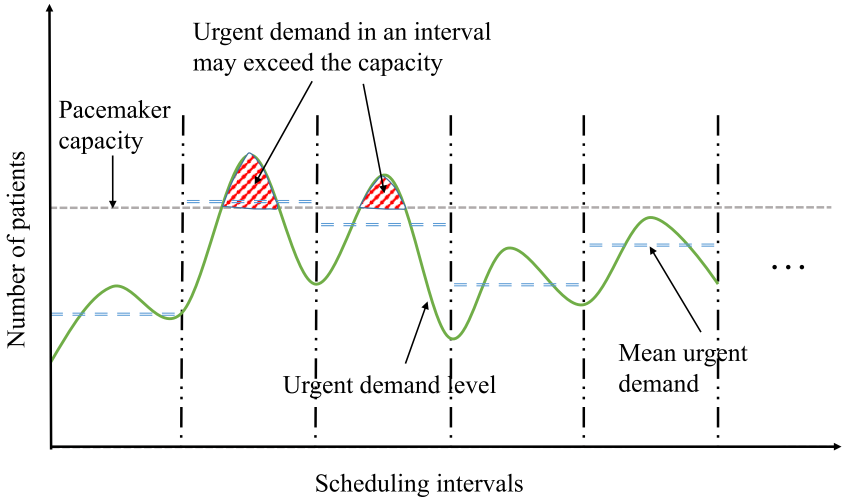


Figure 4.2 Demonstration of Pacemaker Capacity and Urgent Demand

Figure 4.2 illustrates a possible situation of urgent demand in comparison with the pacemaker capacity. The urgent demand in each interval is assumed to be normally distributed with a mean of μ_i and a standard deviation (SD) of σ_i. According to Equation (3.1), the probability of the urgent demand not being greater than the capacity in an interval is:

$$P(x_i \le c_i) = P\left(x_i \le \frac{\sum_N T_{n_i}}{t_u}\right) = P(z_i \le \frac{\sum_N T_n/t_u - \mu_i}{\sigma_i}), \qquad (4.1)$$

where x_i is the urgent demand in interval i, c_i is the interval capacity, N is the number of critical resources, t_u is the examination time of urgent patients, z_i is the standard score for interval i, and T_{n_i} is the net available time of the n^{th} critical resource in interval i. Given a z score, the probability of the pacemaker meeting the urgent demand without delay in that interval is calculated. A probability of 50% will ensure that at least the half of the urgent patients would receive timely services. Otherwise, increasing the pacemaker capability is required to match the patient demand.

4.2.2 Patient scheduling

Having realised the urgent demand, the remaining pacemaker capacity is assigned to elective patients. These elective patients are assigned into appropriate time slots in coordination with the expected urgent arrivals to minimise wait times as demonstrated below.

Intuitively, the elective demand is calculated by subtracting the urgent demand from the pacemaker capacity; however, the future urgent demand is a random variable. We therefore use a forecast to predict the future urgent demand. The number of urgent patients who can be treated without elective patients in wait is calculated as follows:

$$d^P = \mu + k\sigma, \tag{4.2}$$

where d^P is the daily planned urgent demand, μ is the daily mean forecast for urgent demand, σ is the SD, and k is a desired service level. The desired service level for urgent patients is a management-determined parameter that also impacts the number of scheduled elective patients and their wait times.

The next step is to obtain the appointment times of elective patients. Urgent arrivals, despite being unscheduled, follow a pattern within the day, based on which elective patients are scheduled. The number of elective patients to be scheduled, d_i^E, in scheduling interval i is given by

$$d_i^E = \frac{\sum_N T_{n_i} - t_u \theta_i d^P}{t_e}, \tag{4.3}$$

where t_e is the examination time of elective patients, and θ_i is the percentage of urgent patients in scheduling interval i, obtained from a demand profile (e.g. Figure 4.7). If the resulting numbers of elective patients in intervals are not integers, we carefully round the numbers up or down to create the patient schedule. Individual appointments are assigned to elective patients by evenly distributing them in a scheduling interval to reduce the risks of being delayed by urgent patients.

4.2.3 Resource planning

Creating patient flow relies on proper resource allocation. As the number of critical resources is constant, the focal point is to identify the appropriate number of non-critical resources to support the demand. This can be performed with BCs as we mentioned in Chapter 3.2.2. The BCs for shared resources in OPDs are observed at the resource level, and the number of non-critical resources m is computed by Equation (3.5). If there are m operators, the new C/T becomes $1/m$ of the original C/T, which helps bring it below TT to meet the demand.

Additionally, BCs of shared processes in OPDs involve two TTs: one is created based on the planned demand, which is set to enable the department to meet certain demand variations in addition to the average demand, and the other is based on the average demand to estimate the system performance, such as average resource utilisation.

4.2.4 Wait time control

Patients themselves flow in the value stream to access healthcare services. Waiting occurs if a process treats patients slower than in its preceding process as mentioned in Chapter 3.2.4 and if urgent patients surpass the planned urgent demand. We refer both waits to time buffers. A single

FIFO lane is used prior to processes serving either urgent patients or elective patients and is calculated based on Equation (3.17). As all resources respect the same TT, if a process has a preceding process whose C/T is much lower than the TT, patients wait in a FIFO between these two processes.

Multiple FIFO lanes, known as sequenced FIFO, are used before processes serving both types of patients. In our case, the sequenced FIFO is composed of an urgent lane of urgent patients and an elective lane of elective patients. An urgent FIFO lane is akin to a single FIFO lane, whereas an elective lane backs up elective patients when there is an urgent demand surge. The size of the elective FIFO lane is estimated by identifying the number of patients that is larger than the planned demand and multiplying the number by the associated probabilities:

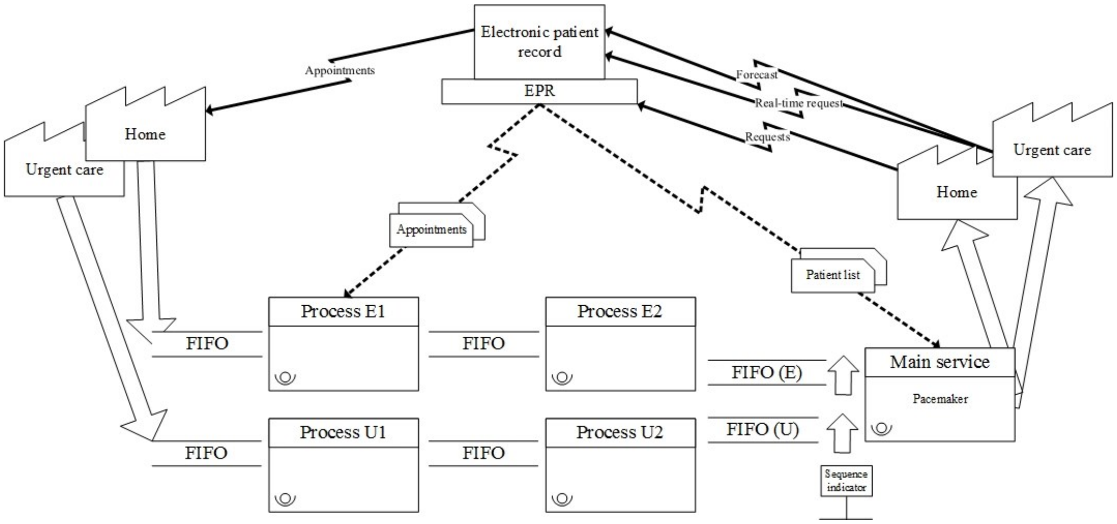
$$\text{FIFO size} = \text{TT}\ \sigma\ \left[\sum_{j=\lceil k\rceil+1}^{j=3}(j-k)(P_j - P_{j-1}) + (\lceil k\rceil - k)(P_{\lceil k\rceil} - P_k)\right], \tag{4.4}$$

where $j \in Z$, and $j \leq 3$; P_j is the probability corresponding to standard score j, and $\lceil k\rceil$ denotes the smallest integer greater than the desired service level k. In this research, the considered upper range of demand variations is the mean plus three SDs, which covers approximately 99.9% of the situations.

Both types of FIFOs keep patients in the correct sequence. An additional function of the sequenced FIFO is to prioritise urgent patients with a sequence indicator, as shown in Figure 4.3. The indicator does not move to the elective FIFO until all patients in the urgent FIFO are treated.

4.2.5 Future state of Lean shared processes in OPD

A resulting future VSM is shown in Figure 4.3. A shared process in OPD receives patient requests from its customers of urgent care units and referrals. The real-time (unscheduled) requests are sent directly to the pacemaker. If requests of urgent patients in that scheduling interval have not exceeded the planned urgent demand, patients are treated with no delay, with elective patients in wait otherwise. The elective requests are stored once they are made by referrals. The system schedules these requests according to Chapter 4.2.2 and notifies patients of their appointment times. Typically, as upstream processes are quite short, the difference between the appointment time and the time at which patients receive services at the pacemaker can be ignored. The working mechanism of the sequenced FIFO, which is composed of FIFO (U) and FIFO (E), in real clinical environments is whenever there are urgent patients, elective patients, albeit scheduled, wait. The performance is evaluated by an efficiency factor, as mentioned in Chapter 3.2.



U-urgent; E-elective

Figure 4.3 Future VMS for Shared Processes in OPDs

4.3 Case study

A case study was conducted for a CT scan process in a radiology department at a local community hospital located in Montreal, Canada. The hospital has approximately 93,000 radiology patients per year, among which approximately 50% are emergency patients. Recently, this department has been under pressure to meet an increasing demand while reducing wait times. The situation is especially severe for the modality of CT scans, so the department decided to apply the proposed Lean framework to its CT service. Night shifts and weekends are excluded in this study because very few patients are treated during the night, and weekends have different work patterns. The study was approved by the Local Research Ethics Committee.

4.3.1 Current system analysis

The CT service involves three types of patients: urgent patients from the ED and hospital wards and elective patients sent by referrals. The current process is described as follows. For urgent patients, a service requesting department (ED or hospital ward) faxes patient requisitions directly to the CT workstation (an isolated room designed for CT exams with a scanner and technicians), awaits technicians to assign a service time for patients by phone calls, transports patients at the given time, and waits for patients to be transported back. Once the called patient arrives at the CT workstation, the technician registers the patient, performs the exam, and calls the porter to send the patient back to the origin (service requesting department). The flow of elective patients is quite straightforward. They arrive according to their appointments, register at the front desk with clerks (sometimes with technicians at the CT workstation), change into hospital gowns, have the exam, and leave the department. The radiographs of all patients are available to the radiologists once their exams are done. The focus of this study is on the patient flow before the radiographs are dictated.

The current process flow is drawn in a VSM, as shown in Figure 4.4. The flow of hospitalised patients prior to their arrival at the CT workstation was not accessible at the time of this study and therefore is not included in the VSM. Information in the VSM was collected through direct observations, interviews with staff members and managers, and data extracted from the hospital software.

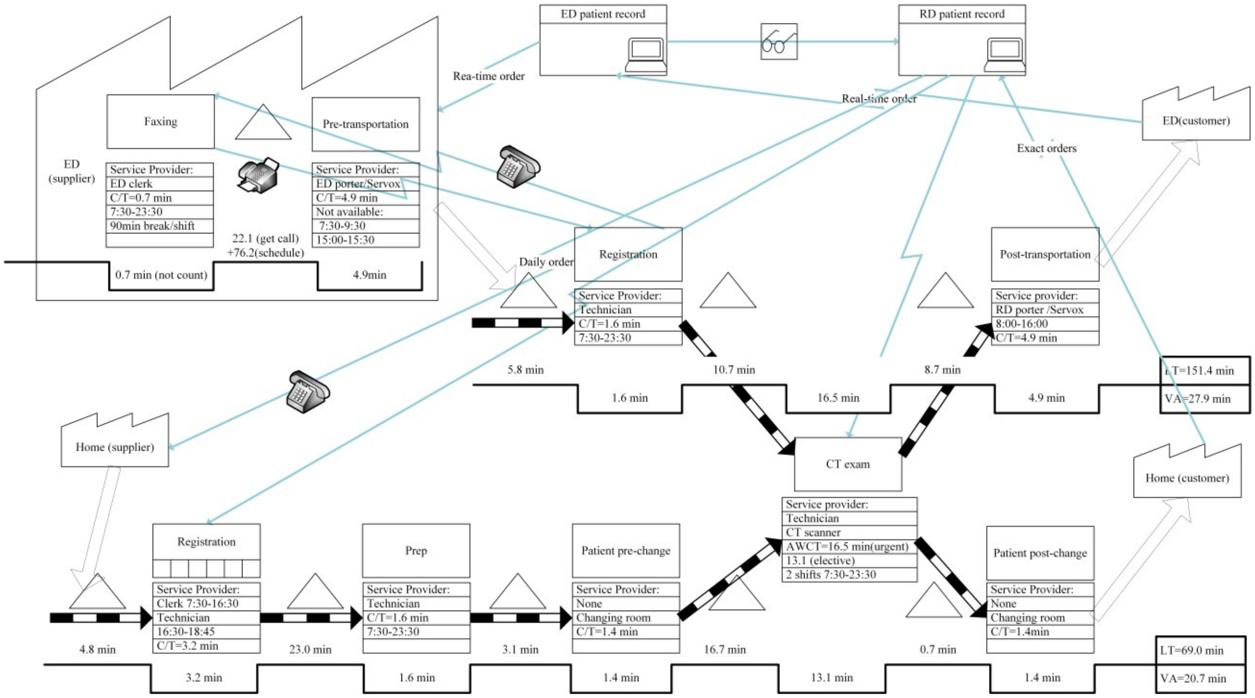


Figure 4.4 Current VSM of CT Scan Process

As shown in Figure 4.4, this CT service communicates with other departments by fax machine and phone calls, both of which require a lot of attention and create distractions and interruptions in operators' daily work. There is a major delay of 98.3 minutes before patient transportations in the emergency patient flow. This consists of time to schedule and time to transport. The time to schedule is the time that elapses between a CT request and a timeslot assigned to the patient, which could be hours due to an imbalance of demand and supply. The time to transport is the time from a timeslot being assigned to the patient physically leaving the origin, the longest of which is approximately 1.5 hours due to the uncoordinated staffing plan. Additionally, technicians in evening shifts waste an average of five minutes per patient waiting for the patient to be transported. Other delays include the information delay, which is defined as the time from the request having been faxed until the technician checks the fax machine, the transportation delay attributed to the centralised transportation service, and the delays in the processing of elective patients because of sudden urgent arrivals.

In summary, as shown in Figure 4.4, for emergency patients, with a total C/Ts of 27.9 minutes and LT of 151.4 minutes, the efficiency factor ρ equals 18%, implying that the value-added time accounts for 18% of the total emergency patient visit time on average. Similarly, for elective patients, the total C/Ts is 20.7 minutes and the LT is 69.0 minutes, leading to an efficiency factor ρ of 29%. The wait times of urgent patients are longer than those of elective patients, indicating that not all urgent patients are given priority in the current process.

4.3.2 Lean designs

To address these issues, our Lean framework was applied to investigate the pacemaker capability, improve the elective schedule, obtain the proper staffing plans, and control wait times. Before using the framework, we first forecasted the daily number of urgent patients.

Daily patients from the ED and hospital wards were analysed independently by using the historical data for the hospital fiscal years of 2009 to 2018. Boxplots were created for each type of patient as shown in Figure 4.5 and Figure 4.6. As we can see from the figures, over the years, the daily numbers of emergency patients have been increasing slightly, while the daily numbers of hospitalised patients are quite stable. Therefore, we decided to use the double exponential smoothing (DES) method to forecast the number of emergency patients while using the single exponential smoothing (SES) method to predict the number of hospitalised patients.

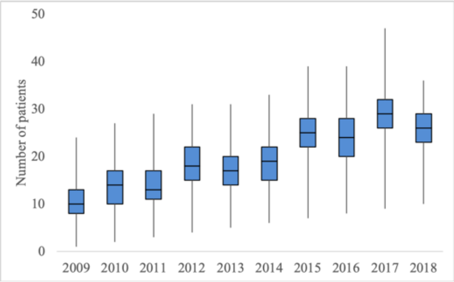


Figure 4.5 Daily Emergency Patients by Year

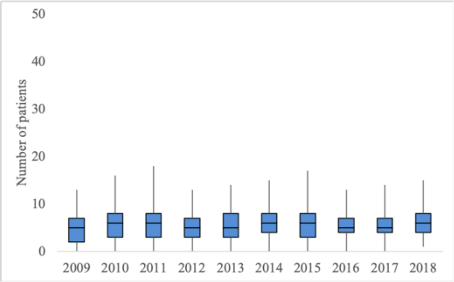


Figure 4.6 Daily Hospitalised Patients by Year

We applied a policy to use every 4-week (28 days) as one observation (period) and divided this period of patients by 28 to obtain the average daily number of patients in that period. As the increase in the number of emergency patients is quite slow and the number of hospitalised patients is stable, the effect of this policy on the forecast is negligible. The DES model can be defined as:

$$\hat{y}_t = 0.16y_t + (1-0.16)(\hat{y}_{t-1} + b_{t-1}) \tag{4.5}$$

$$b_t = 0.02(\hat{y}_t - \hat{y}_{t-1}) + (1-0.02)b_{t-1}, \tag{4.6}$$

where $\hat{y}_t$ is the fitted value, and y_t is the actual observation any time period t. The SES model is

$$\hat{y}'_t = 0.067y'_t + (1-0.067)\hat{y}'_{t-1}, \tag{4.7}$$

where $\hat{y}'_t$ is the fitted value, and y'_t is the actual observation at time t. Using Equations (4.5), (4.6), and (4.7), we obtain a total of 26.1 urgent patients per day. Next, we follow the framework to improve this CT scan process.

4.3.2.1 Capability analysis

The scheduling interval is two hours according to the Guide de Gestion de L'urgence (Ministère de la Santé et des Services sociaux, 2009). The arrival profile of urgent patients is created and shown in Figure 4.7.

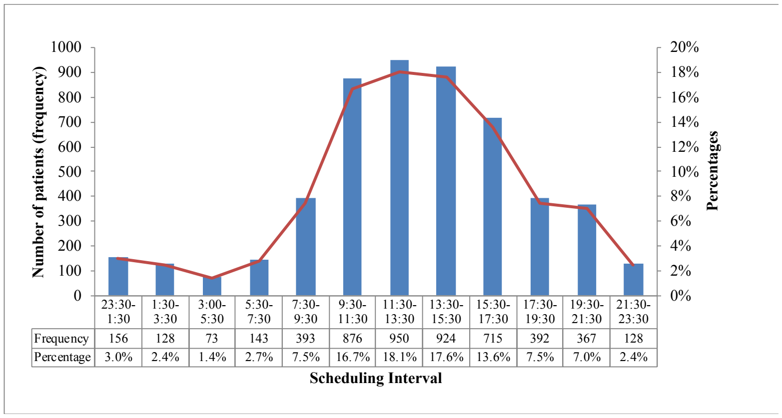


Figure 4.7 Patient Arrival Profile

Then, we compare the pacemaker capacity with the urgent demand. With one scanner and an average examination time of 17 minutes, the bi-hourly pacemaker capacity is 7.06 urgent patients, per Equation (3.7). The average demand in every 2 hours interval is given by the mean forecast of 26.1 daily urgent patients and the arrival profile in Figure 4-7. The bi-hourly SD of the urgent demand is obtained from historical data. According to Equation (4.1), in which the interval mean, SD, and pacemaker capacity are known, we calculate the standard scores and probabilities for the intervals, which are shown in Table 4.1. For example, from 11:30 to 13:30, the probability of the pacemaker (with capacity of 7.06) meeting the urgent demand (which is normally distributed with a mean and SD of 5.12 and 2.69, respectively) without delay is 76.4%. The probabilities for all the scheduling intervals are tabulated in Table 4.1. Elective patient scheduling is implemented next.

Table 4.1 Capability Analysis Over Scheduling Intervals

	Scheduling intervals							
	7:30–9:30	9:30–11:30	11:30–13:30	13:30–15:30	15:30–17:30	17:30–19:30	19:30–21:30	21:30–23:30
Mean	2.12	4.73	5.12	4.99	3.86	2.12	1.98	0.69
SD	1.73	2.59	2.69	2.66	2.34	1.73	1.67	0.99
Score	2.85	0.90	0.72	0.78	1.37	2.86	3.04	6.43
Probability	99.8%	81.6%	76.4%	78.2%	91.5%	99.8%	99.9%	>99.9%

4.3.2.2 Elective scheduling

Before scheduling elective patients, a desired service level is chosen. Three possible levels are investigated in this research: $k = 0$, $k = 1$, and $k = 2$, corresponding to the probabilities of 50%, 84.1%, and 97.7% of treating elective patients without delay. The resulting daily planned urgent demand is 26 patients, 34 patients, and 42 patients per Equation (4.2), respectively. Based on Equation (4.3), the schedule of elective patients is illustrated in Table 4.2. Please note that the service levels are rounded to the nearest integers.

Table 4.2 Elective Schedule Under Different Service Levels

Service levels	**Scheduling intervals**								**Total**
	7:30–9:30	9:30–11:30	11:30–13:30	13:30–15:30	15:30–17:30	17:30–19:30	19:30–21:30	21:30–23:30	
$k = 0$	9	3	3	3	5	8	7	8	46
$k = 1$	7	2	2	2	2	6	6	7	34
$k = 2$	6	6	6	6	6	6	6	6	23

In this study we choose a service level of $k = 1$ for urgent patients to identify the planned urgent demand. The probability that the demand for urgent patients will remain below the planned urgent demand, and hence the scheduled elective patients will be served without waiting, is 84.1%. Elective patients in every interval are evenly distributed into that interval as mentioned in Chapter 4.2.2 to share the risks of being delayed by urgent arrivals.

4.3.2.3 Resource planning

Given the demand for urgent and elective patients, this section creates two BCs to identify the proper non-critical resource levels to support this demand. An intermediate BC shows the current resource levels, and a proposed BC shows the proposed resource levels. The proposed resource levels are identified by closing the gap between the TT and the C/Ts of resources, as mentioned in Chapter 3.2.2. Analyses of the CT scanner, technician, assistant, and clerk are provided; however, resources such as ED porters who do not belong to this radiology department and have less or no effect on Lean transformation are not discussed herein.

Before analysing non-critical resources, we first present the BCs of the CT scanner in Figure 4.8. The TT of the original BC is obtained based on current demand, and that of the proposed BC is based on the average demand and the planned demand (average plus one SD). The C/Ts in the original BC are far below the TT except during 11:30 to 13:30. The C/T is 72.4% of the TT on average, indicating this department can accept more patients; however, from 11:30 to 13:30, fewer patients should be scheduled. This observation demonstrates the need for this department to reschedule the demand in the appropriate timeslots as has been shown in the previous chapters. The C/T in the proposed BC is 87.3% of the TT calculated based on the average demand, showing an 87.3% machine utilisation and a 12.7% gap left to absorb minor interruptions and system variations. The C/T is on average 99.2% of the TT calculated based on the planned demand. If the daily number of urgent patients increases to the planned urgent demand, the scanner will still be able to perform all the exams at nearly full capacity without sacrificing the elective patients' wait times. The analysis of the technicians follows.

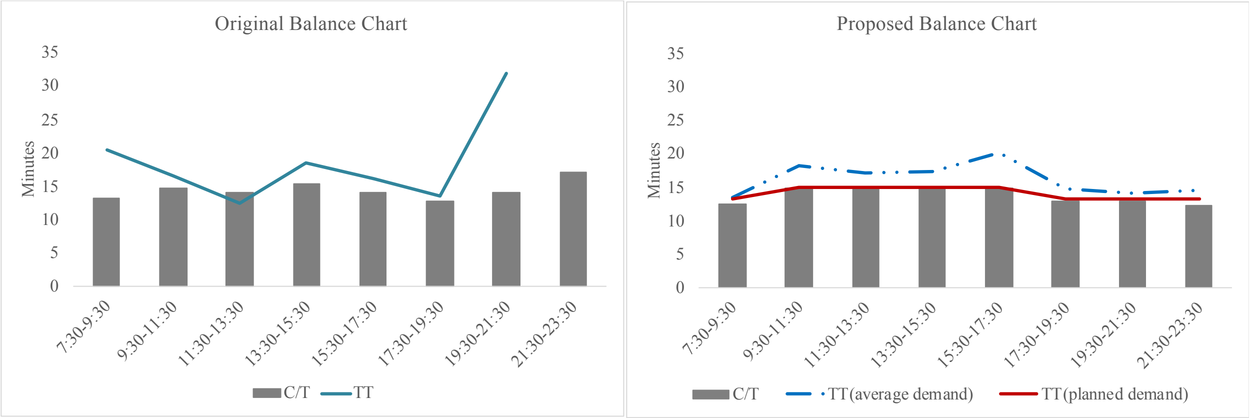


Figure 4.8 CT Scanner Machine Balance Charts

The BCs of the technicians are shown in Figure 4.9. The TT is calculated with Equation (3.4), where the number of technicians (m) is as shown in the availability graph (Figure 4.10).

In the intermediate BC, the C/T comprises exam time (time from the start to the end of an exam), administrative time (time for clerical tasks, such as registering patients), and waiting. The intermediate BC shows that in the first half of scheduling intervals, the C/T is much lower than the

TT, and in the last half of scheduling intervals, the C/T is higher than the TT. Targeting where the C/T is higher than the TT, we bring the C/T down to the TT either by eliminating the waste or managing the staffing plan to increase the number of operators m, thus reducing the C/T to less than the TT. Waiting is an absolute waste and needs to be minimised. Minimising waste is accomplished by assigning an assistant at the CT workstation, which allows the CT workstation to have more than one patient at a time, so that the technician can continuously work on patients without waiting for the next patient to be transported. After removing the waste, the C/Ts at the sixth and seventh scheduling intervals are still higher than the TT, as shown in Figure 4.9. Taking a closer look at Figure 4.10, we find a 1.5-h service downtime owing to the unavailability of the technician. To eliminate this downtime and raise the TT in this region, a new staffing plan is proposed that uses the same number of technicians but shifts the working hours to cover the breaks. The resulting availability graph is also depicted in Figure 4.10. As a result of this technician rescheduling, the gap between the C/Ts and TTs in the first four intervals is closed.

However, after these interventions, there are still scheduling intervals having C/Ts higher than the TT. Hence, we propose to reassign 1.6 minutes of clerical tasks to other operators (assistants), because clerical tasks do not require special training. The resulting BC is illustrated in Figure 4.9, where, except for the C/T of 09:30 to 11:30, other C/Ts are all under the TT. From 09:30 to 11:30, the C/T is still higher than the TT by 1.7%. This gap is so small and brief that adding one extra technician would not be an economical choice; therefore, we decided to retain the proposed staffing. If the demand increases to more than the planned demand in this interval, elective patients will need to wait to be treated in the next interval. On average, the C/T is 86.6% of the TT calculated based on the planned demand and 76.4% of the TT calculated based on the average demand.

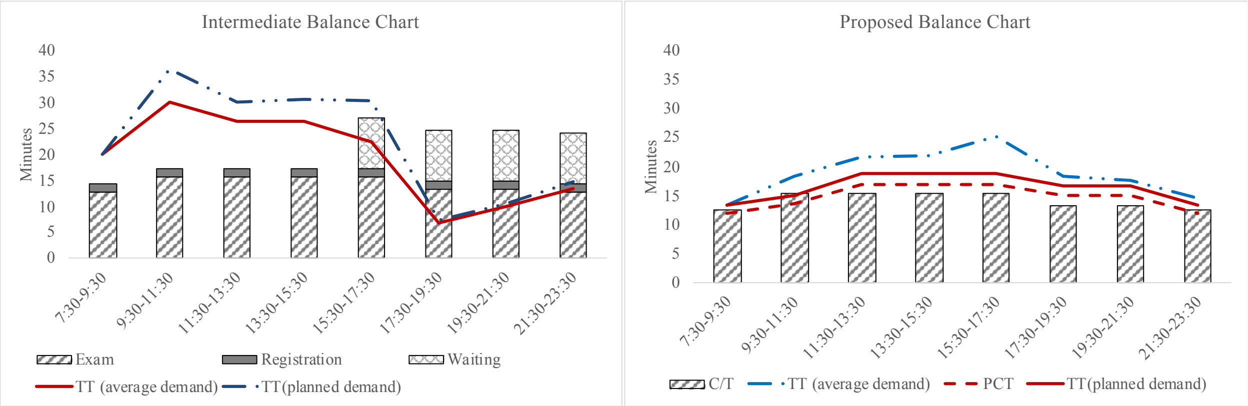
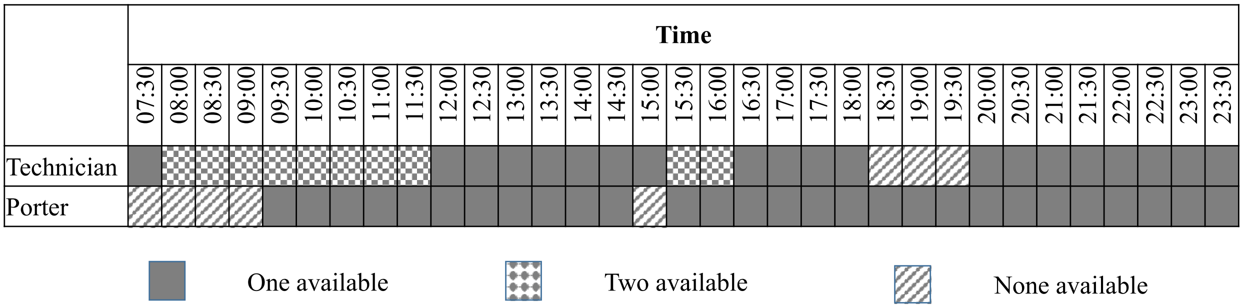


Figure 4.9 Balance Charts of Technicians



	Time																																
	07:30	08:00	08:30	09:00	09:30	10:00	10:30	11:00	11:30	12:00	12:30	13:00	13:30	14:00	14:30	15:00	15:30	16:00	16:30	17:00	17:30	18:00	18:30	19:00	19:30	20:00	20:30	21:00	21:30	22:00	22:30	23:00	23:30
Technician	■	▦	▦	▦	▦	▦	▦	▦	▦	■	■	■	■	■	■	■	▦	▦	■	■	■	■	▨	▨	▨	■	■	■	■	■	■	■	■
Porter	▨	▨	▨	▨	■	■	■	■	■	■	■	■	■	■	■	▨	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

■ One available ▦ Two available ▨ None available

Figure 4.10 Technician Availability Graph

The assistant performs clerical tasks to reduce the C/T of technicians and is a liaison between the radiology department and its supplying departments to accelerate information transmission and patient transportation. The assistant picks up a patient at the corresponding department, transports the patient to the CT workstation in a timely manner, and sends the patient back to the patient's original department upon the completion of the exams. Figure 4.11 shows the BC of the assistant, in which the demand doubles as each patient needs two transportations. Clearly, despite all the benefits of hiring the assistants, they would be underutilised given an average gap of 49.7% between the C/T and the TT. Further actions are needed to bring the C/T to 85–90% of the TT.

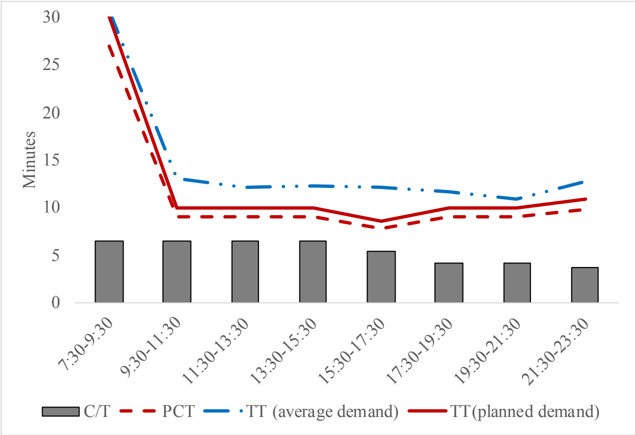


Figure 4.11 Proposed Assistant Balance Chart

Clerks are shared resources serving multiple services in the radiology department. The C/Ts of the clerk are identical for different services because clerical tasks are mostly the same. The TT is estimated by Equation (3.4), where the effective working time is the time that the clerks spent on CT scan patients. CT scan patients account for 19% of the total radiology patients, and thus, on average, per interval, 22.8 minutes of the clerk's time are assumed to be spent on CT scan patients. The BCs of clerks are illustrated in Figure 4.12. In the intermediate BC, the C/T of the clerk is much lower than the TT, indicating that the registration can run with fewer clerks. A recalculation using Equation (3.6) indicates that one clerk is adequate. In the proposed BC, in which there is only one clerk, the first C/T is slightly higher than the PCT but lower than the TT. Further actions should be taken to bring the C/T down to the PCT and reduce the operator's stress at the beginning of the shift.

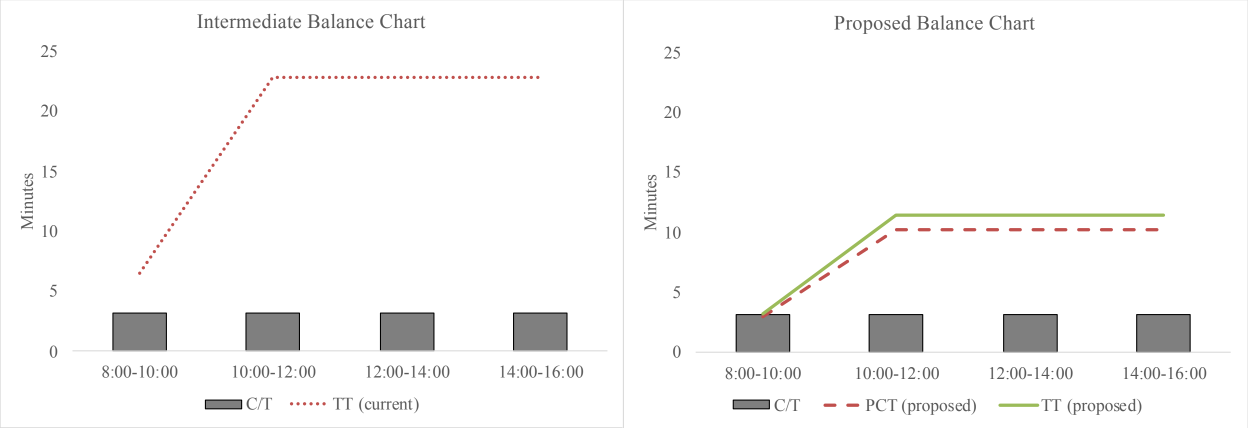


Figure 4.12 Balance Charts of Clerks

4.3.2.4 Wait time control

Upon the completion of the patient and resource scheduling, wait times in FIFOs were estimated. There are six FIFO lanes in the value stream as illustrated in Figure 4.13: FIFO 1 represents the waiting areas before registration, and the rest are the waiting areas outside the CT workstation. Patients do not wait in FIFO 2, FIFO 4, and FIFO 6 because these FIFOs' succeeding processes work quicker than their preceding processes (lanes are still separated due to physical distances). FIFO 5 and FIFO 3 constitute a sequenced FIFO, whose size is given by Equation (4.4), where $k = 1$, and $\sigma = 7.1$ (obtained from historical data). The sizes of the rest of the FIFOs are computed using Equation (4.4). The resulting wait times can be found in the future VSM (Figure 4.13).

4.3.2.5 Future state of the system

Combining the previous designs, a future state of this system is shown in a VSM (Figure 4.13). This CT service has one patient schedule that consists of appointment times and bi-hourly planned urgent demands. The appointment times are created in coordination with the urgent arrivals according to Table 4.2 (k = 1) after receiving phone calls from the referrals. These appointment times are provided to the clerk for checking-in patients and technicians for patient information verification. The bi-hourly planned demand is sent to technicians at the CT station for assessing the patient volume. If technicians sense an overflow of urgent patients, the management investigates the process and takes action to avoid a significant increase in wait times. Upon arrival, elective patients go through the process but may be stopped at any time by emergency requests. In Figure 4.13, urgent patients in FIFO 5 of the sequenced FIFO always have priority in receiving exams. Urgent patients pull the CT service whenever they need it, and the assistant transports and registers patients as soon as possible before the technician performs the exams, thus reducing the wait time. Although the assistant works as a liaison to reduce the negative impact (e.g. information delay) of using disconnected systems among departments, an integration of systems is still highly recommended for further improvements.

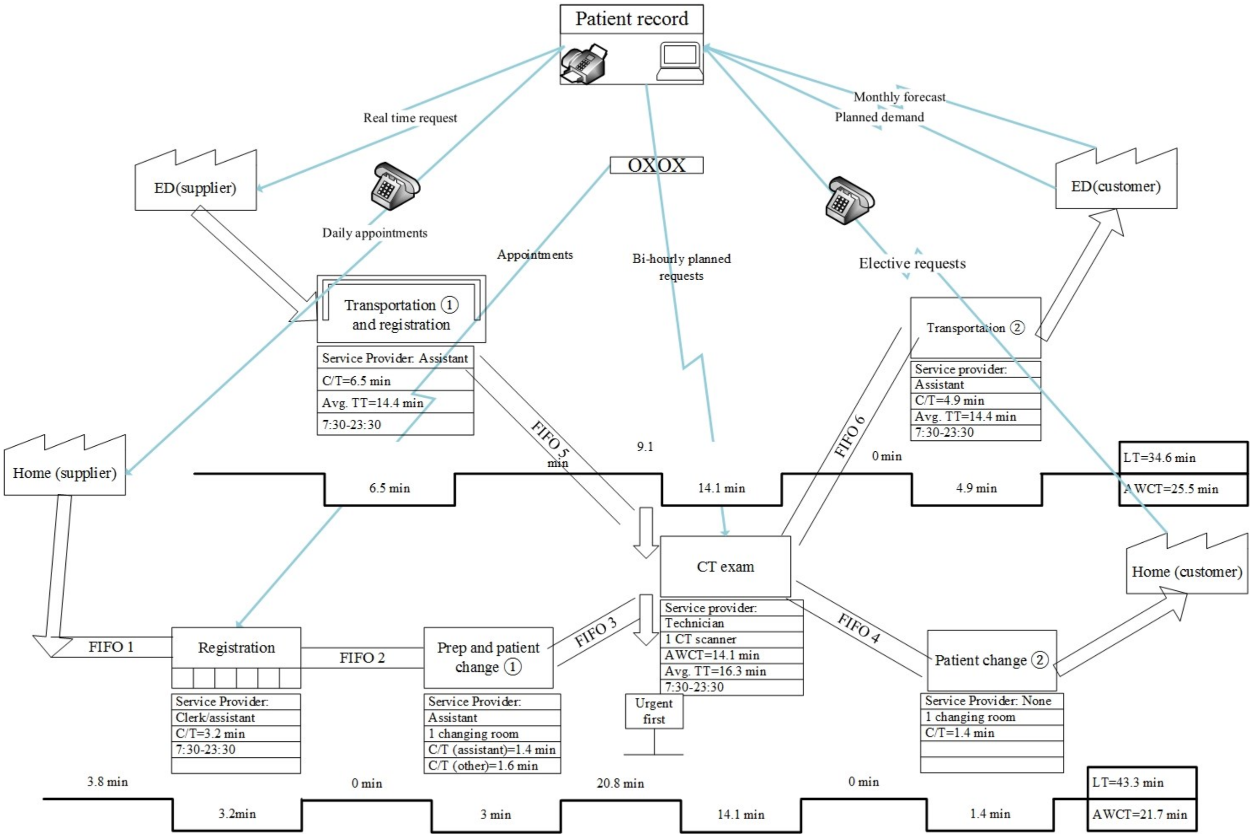


Figure 4.13 Future VSM for CT Scan Process

As shown in Figure 4.13, with the total C/T of 25.5 minutes and the LT of 34.6 minutes, the value-added time of emergency patients accounts for 73.7% of their visit time on average, which produces an efficiency factor ρ of 73.7%. The total C/T of elective patients is 21.7 minutes and the LT of elective patients is 43.3 minutes, which leads to an efficiency factor ρ of 50.1%.

4.3.3 Result analysis

In this section, the outcomes of the Lean transformation for the CT scan process are presented.

4.3.3.1 Lead times

We first discuss the patient LTs. As illustrated in the current and proposed VSMs (Figure 4.4 and Figure 4.13), the Lean framework reduces the LT of emergency patients by 77.1%, from 151.4 minutes to 34.6 minutes, and that of elective patients by 37.2%, from 69.0 minutes to 43.3 minutes. While using Lean, the same type of patient is assumed to be homogeneous in terms of mean service time, for example; however, in reality, departments operate in a stochastic environment in which service times vary and there are uncertainties in unscheduled arrivals. Considering the complexities brought by the stochastic environment, this section uses discrete-event simulation (DES) to evaluate LTs in a realistic environment.

A simulation of the case department was created in Rockwell Arena software based on the system described in Chapter 4.3.1. The simulation imitates the behaviour of the studied CT scan process and provides metrics of interest, such as LTs, by the end of each run. The simulation was verified by following the entities (representing patients) in the model and then validated by comparing a new set of real data collected in the next month to the simulated values. Once verified and validated, we ran a total of 60 replications of before and after Lean scenarios, 30 for each scenario. It is known that 30 samples are usually enough for a statistically sound analysis. Figure 4.14 is the resulting boxplot demonstrating patient LTs before and after the Lean transformation under stochastic conditions.

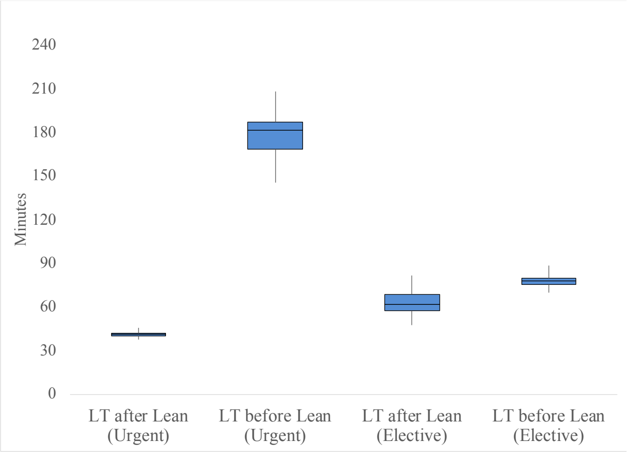


Figure 4.14 Boxplots of Patient Lead Times Before and After Lean

As can be seen in Figure 4.14, before and after Lean, under stochastic conditions, the average LTs of urgent patients are 178.6 minutes and 41.3 minutes, respectively, in comparison with the average LTs of urgent patients of 151.4 minutes and 34.6 minutes under a deterministic environment. The average LTs of elective patients before and after Lean are 77.8 minutes and 62.6 minutes under a stochastic environment, respectively, in comparison with the average LTs of elective patients of 69.0 minutes and 43.3 minutes under a deterministic environment. Clearly, under a stochastic environment, the LTs before and after Lean for patients worsened as expected; nevertheless, these simulation runs helped us validate the expected outcomes of Lean recommendations. In addition, Figure 4.14 indicates that, after urgent patients are prioritised in the proposed system, the variations in LTs of urgent patients are smaller than those in the current system, hence improving urgent patient care.

4.3.3.2 Other results

The proposed patient schedule increases elective patient access by 21.4% from 28 patients to 34 patients per day. These elective patients are treated without a wait for 83.9% of the situations, and the possibility of urgent patients in an interval being delayed to the next interval is less than 10% based on Table 4.1. Additionally, the average utilisation of the CT scanner, which is considered the critical resource in this study, is improved to 87.3%, representing a 19.9% increase compared to the current utilisation of 72.8%.

The proposed system runs with one additional assistant, one less clerk, and the same number of technicians based on the analysis in Chapter 4.3.2.3. The role of the assistant is critical because it allows the technician to work at a speed under the TT to meet the patient demand, eliminates downtime and transportation delays, oversees urgent patients, and improves technicians' physical health by helping them with moving patients according to the interviews with technicians. On the other hand, the assistant BC demonstrates the low resource utilisation. One suggestion is that, in the future, this assistant can be shared by other services such as MRI, whose workstation location is close with the same working hours, resulting in an improved utilisation of the assistant. Furthermore, at the first scheduling interval in which the C/T of clerks is higher than the TT, instead of compromising the patient wait time, the assistant can be temporarily assigned to the registration desk to help the clerks. It is recommended that the extra clerk be trained in the registration process to serve as an assistant with no added cost, and further research is needed to validate the use of the assistant's time elsewhere.

We also found that, while ED arrivals show an increasing trend (Figure 4.5), the critical resource, the CT scanner in this case, has almost reached its full capacity based on Figure 4.8. The machine utilisation is at 99.2% when the urgent demand reaches the planned demand. Thus, the CT scanner becomes the bottleneck of this service. If the capacity of the CT scanner cannot be expanded, such as acquiring a new machine, according to Equations (4.2) and (4.4), patient wait times will increase provided there are the same number of elective patients; the department could otherwise reduce the scheduled elective patients; however, this increases the wait time of elective patients to initiate their exams. Thus, acquiring additional machines would be the best solution to meet the increasing CT requests of this department.

4.4 Discussion

The shared processes in OPDs significantly contribute to the creation of patient flow in a hospital. Attempting to meet demand while maintaining a high resource utilisation and quality patient care, shared processes in OPDs have been exploring improvement approaches. The Lean methodology seems to be an effective solution for similar issues; however, the literature shows a lack of well-established Lean methodologies. This research, therefore, proposes a structured framework for shared processes in OPDs to guide their Lean transformation process and improve performance. This framework involves four steps: capability analysis, patient scheduling, resource planning, and wait time control.

As shown in this research, a capability analysis and patient scheduling compose the foundation of a Lean transformation by identifying and scheduling the patient demand. Considering society as a whole, patient demand can be enormous. Identifying demand becomes a problem of finding the maximum number of patients that can be accepted at the pacemaker process involving critical resources. When urgent and elective patients coexist in the studied shared processes in OPD, for example, a priority is to determine whether the urgent demand can be satisfied by the pacemaker, which is done in the framework through a capability analysis. The subsequent question is the time and quantity of elective patients to be scheduled to create a smooth patient flow, which can be referred to as a Lean concept of production levelling (Duggan, 2018). While levelling did not appear to be a conscious part of Lean research in shared processes in OPDs, our framework moves Lean implementation in healthcare towards maturity through a better patient schedule, in which elective patients are distributed into timeslots in coordination with projected urgent arrivals.

When we level the schedule, a key question is coordinating the demand of the two competing patient groups. Common practices in the previous studies set the urgent demand to a constant level and rescheduled or reoriented elective patients when the real demand exceeded the defined level (Geng et al., 2017; Huang & Marcak, 2013). What sets our research apart is that we offer several scheduling options by introducing a managerial indicator, called the desired service level. According to Chapter 4.2, a larger service level provides a higher planned urgent demand, fewer scheduled elective patients, and lower wait times for elective patients but also includes a risk of idle resources. On the other hand, a lower service level leads to more scheduled elective patients but higher chances of overtime, rescheduling, or reorientation. Although in the case study, only one service level was selected by expert opinion, an alternative is to estimate the performance under all service levels to finally determine which situation best balances the number of patients accepted, wait times, and resource utilisation.

Resource planning plays a central role in the daily management of shared processes in OPDs and in the Lean literature (Amaratunga & Dobranowski, 2016; Waring & Bishop, 2010). Closing the gap between the CT of a resource and its corresponding TT is a strategy that has been widely used in Lean manufacturing to obtain an appropriate number of resources to support customer demand (Duggan, 2018). In contrast, Lean techniques that are used for resource planning are not well established in healthcare and are missing proper definitions and instructions. Furthermore, again, due to the uncertainties in urgent arrivals, scheduled resources in such departments need to accommodate a certain level of demand variations, which has not yet been investigated in the reviewed literature. In response to these, our framework clearly defines Lean techniques for resource planning, adapts them to shared processes in OPDs to ensure that a certain degree of patient overflow is acceptable, and embeds resource planning into the proper stage of a Lean transformation process.

A Lean buffer in manufacturing is defined as the minimum level of materials necessary and sufficient to buffer the variations in downstream demand and the capability of the upstream process (Shook & Marchwinski, 2014). As patients themselves are transformed in the core healthcare process, such buffers do not appear in healthcare. However, there exists another form of Lean buffer, known as a time buffer. This is the wait time resulting from the variability in urgent arrivals and uncoordinated resources. Decreasing the sizes of time buffers depends on improved patient and resource schedules to smooth the patient flow. Our framework also proposes to control these time buffers through a Lean technique of FIFOs. Whenever the wait time in the FIFOs are larger than calculated, managers would look into the process, identify the causes, and take countermeasures to prevent longer wait times.

Based on this study, we conclude that using a structured Lean framework in alignment with Lean principles assists practitioners to improve performance and reduce wait times. This framework identifies the demand of urgent and elective patients, allocates an appropriate number of resources to meet the demand, and creates a patient flow. Wherever patients cannot flow, there is a time buffer indicating the expected wait times of patients. The framework also serves as a guideline for hospital mangers to repeatedly conduct Lean activities in shared processes in OPD to maintain or keep pursuing better patient service and better resource allocation.

5. Lean Transformation Framework for Emergency Departments

5.1 Introduction

The emergency department is an integral part of a hospital in which emergency services are offered for unscheduled patients. The most important issue in the ED is overcrowding, which is defined as a situation in which the demand for emergency services exceeds the timely supply (Hoot & Aronsky, 2008; Velt et al., 2018). ED overcrowding is an ongoing and critical issue associated with negative outcomes including increased deaths and hospital admissions (McCusker et al., 2014), increased operational costs (Bayley et al., 2005), and decreased patient satisfaction (Wang et al., 2017).

ED overcrowding involves a complex network of interwoven processes ranging from the ED itself to the entire hospital. To develop potential solutions, researchers have been using a conceptual ED overcrowding model proposed by Asplin et al. (2003), which partitions the causes of ED overcrowding into three interdependent components of input, throughput, and output. The input component includes the factors that contribute to the demand for ED services. While nonurgent visits, frequent-flyer patients, and the flu season are the commonly studied factors that result in ED overcrowding (Hoot & Aronsky, 2008), the impacts either are brief (Schull et al., 2004) or require interactions among healthcare organisations and community (Moe et al., 2017). On the hospital level, the input explicitly refers to the quantity and timing of ED service. While ED arrivals are characterised by large daily, weekly, and seasonal variations, along with a degree of inherent variability, it is nevertheless predictable (Asheim et al., 2019). A study by Carvalho-Silva et al. (2018) showed that an accurate forecasting of ED arrivals optimises resource allocation to the real demand, thus resulting in better patient care.

The throughput component looks internally at the ED care process for the factors that have the largest effect on longer ED LOS and low resource utilisation. The ED care process lies in and outside the ED. Within the ED, a typical care process consists of triage, assessments, interventions, and discharge, operated by nurses and physicians. Research has demonstrated that improved nurse/physician scheduling (Wong et al., 2014), as well as better stretcher allocation (Luscombe & Kozan, 2016), enhances the efficiency in EDs. Other activities, for example, the redesign of the ED layout, show positive influences on the LOS reduction (Wang et al., 2015). Outside the ED, external providers are frequently asked to offer various tests (CT scans, for example), consulting services, etc. during the patient stay. Reducing the time elapsed between the request and the result can significantly decrease the ED LOS (Storrow et al., 2008). However, a barrier to reducing this time is the uncertainty in demand. Vasanawala and Desser (2005) suggest support departments to project and evaluate the quantity and timing of ED requests to reduce wait times for both the ED and its scheduled patients and to efficiently use resources.

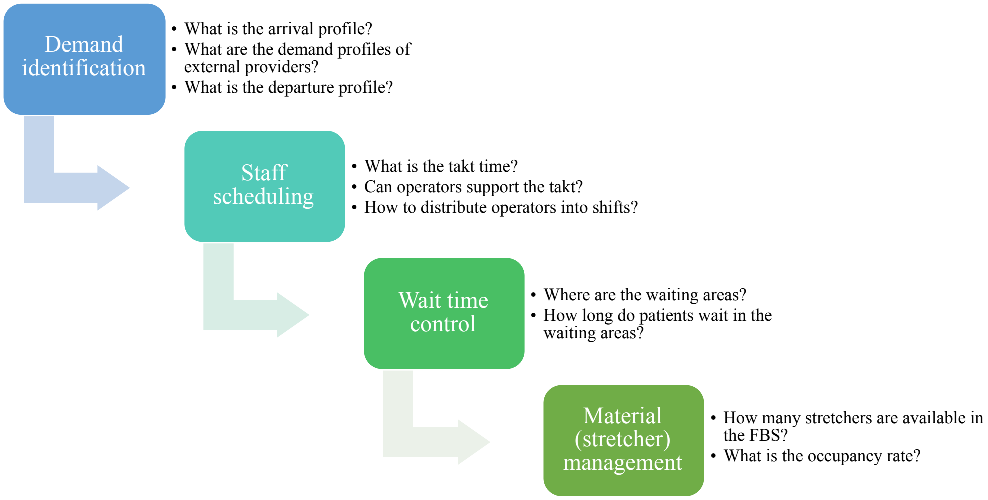
The inability to move admitted patients to hospital units is one of the most frequently cited reasons for ED overcrowding from the output perspective (Shi et al., 2015; White et al., 2013). This forces the ED to board these patients, who may require ongoing care and thus consume ED resources, until beds are made available in hospital units. The time from an inpatient bed request to the patient's physical departure is defined as ED boarding. Barak-Corren et al. (2017) pointed out that an early identification of ED demand for hospital beds helps the units manage their daily operations and significantly reduce boarding time.

Lean has been positively received by managers in the ED for ED overcrowding alleviation (Bucci et al., 2016). A popular Lean activity illustrated in the literature is waste elimination. Waste refers to the non-value-added steps that require extra resources, added delay, and, ultimately, increased costs (Burgess & Radnor, 2013). In an Ontario hospital, the ED LOS was reduced by an average of 0.8 hours through identifying and eliminating the seven sources of waste in the ED care process (Ng et al., 2010). Murrell et al. (2011) rearranged the value-added step in their triage process and developed a rapid triage and treatment system to reduce ED LOS and decrease the rate of left without being seen (LWBS) in a community hospital. Similarly, Lowthian et al. (2015) decreased the median of the ED LOS by approximately 1.5 hours without compromising safety and quality by eliminating the non-value-added steps in the process. While these Lean applications emphasise the waste hidden behind each step of the ED care process, another stream of literature pays more attention to the overall patient flow by using VSM techniques. Bal et al. (2017) used the VSM technique to draw a diagram of every step involved in the ED care process and associated information such as the number of operators. Improta et al. (2018) presented their ED care process using the VSM technique, by which they captured the overall sense of the system and identified areas for improvement. Another Lean study in ED comprehensively mapped the ED patient flow from arrival to departure using the VSM technique and demonstrated the C/T and value-added and non-value-added components along the value stream (Wang et al., 2015).

While Lean implementations in EDs are not exceptionally tool-dominated studies and lack structured methodologies, there arises a research question of how to effectively use Lean in the ED with an objective of mitigating ED overcrowding. Therefore, this chapter establishes such a Lean framework, in which Lean techniques are adapted to the context of EDs to overcome the issues of ED overcrowding at a system level considering the dependent input, throughput, and output components.

5.2 Lean framework for EDs

This section introduces a framework adapted to the ED Lean transformation, as shown in Figure 5.1. This framework consists of demand identification, staff scheduling, wait time control, and material management, and aims to streamline patient flows, reduce wait times, and efficiently use resources, thus alleviating ED overcrowding. Starting with the arrival profile, this framework traces patient demand along the value stream and synchronises the pace of departure with the pace of arrival while considering the impacts of external providers. The number of operators needed is obtained using the Lean technique of TT to support demand, followed by the distribution of operators into appropriate shifts. Whenever the demand outpaces the supply, patients are in a controlled wait. Ultimately, to support the flow of patients, the amount of required materials, such as stretchers, are calculated.



- What is the arrival profile?
- What are the demand profiles of external providers?
- What is the departure profile?

- What is the takt time?
- Can operators support the takt?
- How to distribute operators into shifts?

- Where are the waiting areas?
- How long do patients wait in the waiting areas?

- How many stretchers are available in the FBS?
- What is the occupancy rate?

Figure 5.1 Lean Transformation Framework for EDs

5.2.1 Demand identification

The management of the patient flow is significantly affected by the level of demand for ED care. According to Baker et al. (2009), managing demand in an ED is to free up resources in a timely manner by discharging/admitting patients at the rate of arrival. However, the external provider is a decisive factor for the timing of patient discharge as we mentioned in Chapter 5.1. Therefore, identifying the demand means finding a departure profile that closely reflects the arrival profile and is subject to the constraints of external providers, such as working hours. Through this process, we can also capture the demand profiles of external providers. The profiles illustrate the number of patients/requests in discharge intervals (DIs). The DI is the frequency with which ready-to-discharge patients are sent home and/or admitted to hospital units.

The external providers, in general, include the laboratory, radiology (which includes the CT scan process discussed in Chapter 4), and specialists. Based on Ministère de la Santé et des Services Sociaux (2006), it is recommended that laboratory and radiology test results are made available before the consults of specialists, and the results of all services should be available prior to a physician's final assessment. The stepwise procedure to obtain the departure profile is described as follows.

Step 1) Capture the arrival profile illustrating the number of arrivals per DI based on historical data.

Step 2) Obtain the demand profile of the laboratory, namely, the number of patients who need blood work per DI. The demand in each DI in the arrival profile is shifted to a DI that is t minutes later, where $t = t_{la}$, and t_{la} is the time to start the blood work. Depending on the hospital policy, this time can be time to triage or time to physician initial assessment (PIA).

Step 3) Set the preliminary departure profile to be the arrival profile, which is an ideal state in which the pace of arrival and the pace of departure are perfectly synchronised. Patients in the profile are differentiated on the basis of services required: imaging, consults, both imaging and consults, or neither imaging nor consults.

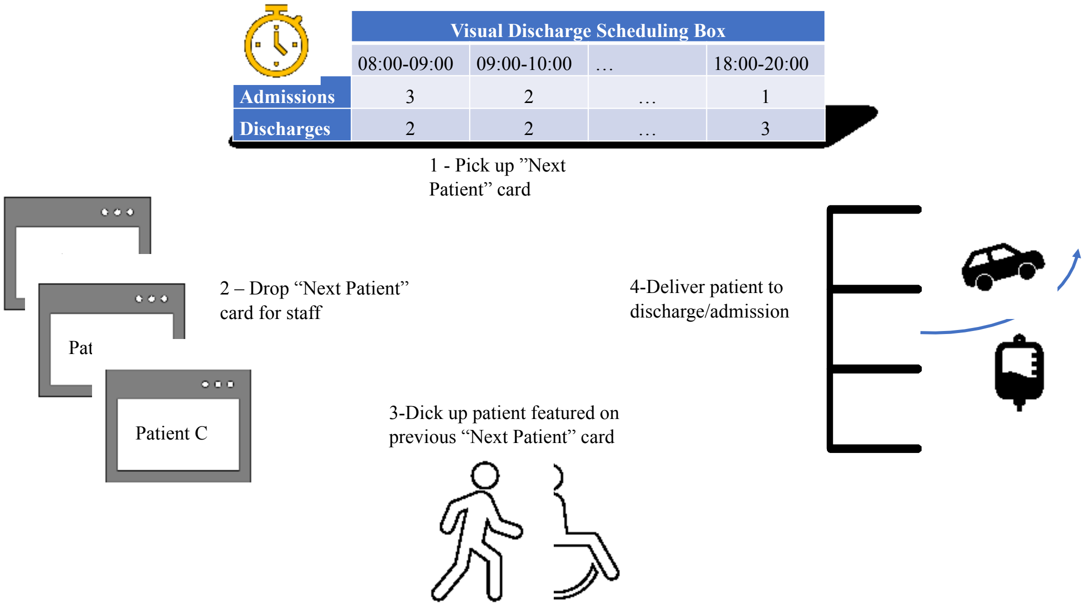
Step 4) Obtain the preliminary demand profile of consulting. Patients requiring consults in a DI in the preliminary departure profile are shifted to the DI that is t minutes later, where $t = t_{con}$, and t_{con} is the processing time (PT) of consulting. PT is defined as the time from when a patient/sample goes to the external provider until the result is available.

Step 5) Obtain the preliminary demand profile of imaging. Patients requiring imaging in a DI in the preliminary departure profile are shifted to the DI that is t_1 minutes later, and patients requiring both imaging and consulting in a DI in the preliminary departure profile are shifted to the DI that is t_2 minutes later. $t_1 = t_{im}$, $t_2 = t_{con}$, and t_{im} is the PT of imaging.

Step 6) Capture the final demand profiles of consulting and imaging. The demand in DIs in which external providers are not available is redistributed to appropriate time slots, meaning the start time of the corresponding services are delayed, hence delaying the patient departure.

Step 7) Reflect the changes made in Step 6 on the preliminary departure profiles to obtain the final departure profile.

The resulting demand profiles are sent to the corresponding external providers for scheduling purposes, and the departure profile, which shows the estimated number of departures, including both admitted and discharged patients, is used to guide the ED discharge process. The working mechanism is shown in Figure 5.2. The departure cycle repeats every DI and includes four major steps. First, the discharge staff goes to the visual discharge scheduling box and picks up one patient card. This person then drops the card to a staff member in the ED (e.g. the coordinator) and takes the patient who is ready to be discharged or admitted. The next patient card tells the emergency team that they need to prepare a patient for discharge by the next cycle. If the number of patients in a DI is not an integer, we carefully round the numbers up or down to create the scheduling box.



Visual Discharge Scheduling Box

	08:00-09:00	09:00-10:00	…	18:00-20:00
Admissions	3	2	…	1
Discharges	2	2	…	3

Figure 5.2 “Milk Round” for ED Discharge (adapted from Baker et al., 2009)

5.2.2 Staff scheduling

After the demand is identified, preparing operators to meet the demand helps create a smooth ED patient flow to reduce ED overcrowding. The challenges are the efficient allocation of operators to meet the daily demand and minimisation of wait times due to demand variability during the day. To overcome these challenges, two stages are involved in staff scheduling. The first stage identifies the number of required operators on a daily basis with a Lean technique of BCs, as discussed in Chapter 3.2.2. The demand variability within a shift causes mismatches between the capacity and demand given a constant staffing level in that shift. The second stage distributes operators across the proper shifts to minimise the discrepancies between demand and capacity (i.e. supply). We use a mathematical model to identify the optimal staffing levels throughout the day for this purpose. To match capacity to demand, we deploy a four-shift policy, i.e. one additional mid-shift to the traditional morning, evening, and night shifts. The fourth shift (mid-shift) can start as early as 9:00 am and overlaps with the morning and evening shifts. For practical reasons, it must end before the night shift starts. This results in seven possible options for the start time of the mid-shift. A mathematical model is developed to determine the optimal number of operators in each shift, including the mid-shift, that minimises the hourly discrepancies between demand and capacity throughout the day. The model is as follows.

Indices: i time index; I denotes the total number of hours per day (24);
j shift index; J denotes the number of shifts per day (4);
k option index; K denotes total number of mid-shift options (7);

Variables: x_{kj} = number of operators assigned to shift j under the k^{th} option;
y_{ki} = number of operators during the i^{th} hour of the day under the k^{th} option;
z_k = square of the sum of discrepancy between the demand and capacity throughout the day under the k^{th} option;

Parameters: m = total number of operators per day;
t_{twc} = total work content;
t_i = net available time in the i^{th} hour of the day
d_i = demand in i^{th} hour of the day;
α =human factor;
β_{ki} = 1 if mid-shift is effective in i^{th} hour under the k^{th} option; otherwise 0.

Minimise:

$$\sum_{k=1}^{K} z_k$$

Subject to:

$$y_{ki} = x_{k1} + \beta_{ki} x_{k4} \qquad i = 9, \dots, 16 \quad k = 1, \dots, K \tag{5.1}$$

$$y_{ki} = x_{k2} + \beta_{ki} x_{k4} \qquad i = 17, \dots, 24 \quad k = 1, \dots, K \tag{5.2}$$

$$y_{ki} = x_{k3} \qquad i = 1, \dots, 8 \quad k = 1, \dots, K \tag{5.3}$$

$$\sum_{j=1}^{J} x_{jk} = m \qquad k = 1, \dots, K \tag{5.4}$$

$$y_{ki} > 0 \qquad i = 1, \dots, 24 \quad k = 1, \dots, K \tag{5.5}$$

$$z_k = \sum_{i=1}^{I} \left(\frac{\alpha t_i y_{kj}}{t_{twc}} - d_i \right)^2 \qquad k = 1, \dots, K \tag{5.6}$$

The objective function minimises discrepancies between the demand and the capacity of all mid-shift options. It should be noted that we are only interested in finding the mid-shift option k that has the smallest z_k value. Constraints (5.1), (5.2) and (5.3) set the number of operators in the day, evening, and night shifts, respectively. Constraint (5.4) ensures that the total number of operators equals what we have obtained in OBCs. Constraint (5.5) makes sure that there is at least one operator per shift for practical reasons. Constraint (5.6) calculates the squared difference between the effective capacity according to Equation(3.7) and demand under each option. The resulting staffing plans of all options are compared, and the one associated with the smallest discrepancy is the optimal staffing plan. This optimal staffing plan reflects the scheduled capacity of the ED, which is the maximum number of patients that can be treated by each operator, for example, nurses. The selected optimal staffing plan minimises the situation of imbalanced demand and supply but definitely cannot eliminate it. This implies that within the day, patient demand may exceed the scheduled capacity at times; however, the scheduled capacity meets the average patient demand on a daily basis.

5.2.3 Wait time control

When demand outpaces the scheduled capacity during the day, there are inevitable patient waits, called time buffers. The goal of wait time control is to monitor the time buffers and estimate the length of the wait using FIFO lanes, a Lean technique. The FIFO lanes maintain patients in the right sequence after being prioritised lowest-acuity-first (LAF) (Bullard et al., 2008) at the triage (see Figure 5.3). Physically, FIFO lanes are ED waiting rooms.

To compute wait times in FIFOs, we divide the day into several cycles. Within each cycle, patients wait due to temporary lack of capacity; however, all patients are treated by the end of each cycle. Explicitly, assuming the q^{th} cycle is from the ${l_q}^{th}$ hour to the ${l_{q+1}}^{th}$ hour (inclusive) of the day, the untreated patients in this cycle are subject to

$$\Delta_i = max\{\Delta_{i-1} + d_i - c_i, 0\} > 0 \qquad i = l_q, \dots, l_{q+1} - 1 \tag{5.7}$$

$$\Delta_{l_{q+1}} = max\left\{\Delta_{l_{q+1}-1} + d_{l_{q+1}} - c_{l_{q+1}}, 0\right\} = 0, \tag{5.8}$$

where c_i is the capacity in the i^{th} hour of the day, d_i is the demand in the i^{th} hour of the day, and Δ_i is the number of untreated patients in the i^{th} hour of the day. The size of the FIFO, which represents the average patient waiting time, is computed as follows:

$$\text{FIFO size} = \frac{1}{\mu}\sum_{q=1}^{Q}\sum_{i=I_q}^{I_{q+1}}\Delta_{qi} \ , \tag{5.9}$$

where μ is the mean forecast of the daily demand, and Q is the number of cycles.

5.2.4 Material management

The patient flow in the ED care process is also supported by materials such as stretchers. Insufficient stretchers would cause important human resources to be wasted and prolong patient wait times. We address this issue by using a finished stretcher supermarket (FSS). An FSS holds a predetermined number of stretchers and supplies these stretchers to acute patients in a timely fashion. These acute patients are on stretchers during the entire ED stay. An FSS comprises the cycle stock, buffer stock, and safety stock as shown in Table 5.1, where μ is the average demand of acute patients, σ is the standard deviation, t_B is the boarding time, and β is the admission rate.

Table 5.1 Finished Stretcher Supermarket

FSS component	**Patient demand**	**Stretcher time per patient**
Cycle stock	μ	LOS-t_B
Buffer stock	2σ	LOS-t_B
Safety stock	$\beta\mu$	t_B

As shown in Table 5.1, the cycle stock is the average number of stretchers to meet the demand of the average number of acute patients (μ). The number of stretchers in the buffer stock is set to meet a demand surge (2σ). Two standard deviations are the recommended buffer stock, as there is a 97.7% probability that the demand surge will not surpass this value. The number of stretchers in the safety stock should meet the demand of boarded patients. Distinguishing boarded patients from others helps reveal admission issues and stimulates the motivation for further reducing the boarding time by improving the collaboration between the ED and hospital units. The stretcher time is the amount of time a patient spends on a stretcher. The times for which patients utilise the stretchers of different types of stock are presented in Table 5.1.

Given Table 5.1, according to Equations (3.4) and (3.6), the total number of stretchers needed, i.e. the size of an FSS, is calculated as follows:

$$\text{FSS size} = \frac{(\mu+2\sigma)\times(LOS-t_B)+\beta\mu t_B}{Effective\ working\ time}, \tag{5.10}$$

where the effective working time of stretchers is the available time (1440 minutes) minus the time needed for cleaning.

When the system is running, the cycle stock begins to be utilised by patients; however, it automatically refills according to the discharge plan. The buffer stock is used when the ED experiences a demand surge greater than the average. When this happens, a Code Red is sent, calling for interventions to free up more stretchers. The safety stock buffers the mismatches between the TT of the ED and the TT of hospital units and is normally used by only boarded patients. It can be occasionally used for new arrivals if patients in process cannot be discharged due to some unexpected reasons.

Given the total number of stretchers needed, the occupancy rate of the ED is estimated by

$$\text{Occupancy rate} = \frac{\text{Number of stretchers occupied per day}}{\text{Total stretchers in the ED}}. \tag{5.11}$$

A higher occupancy rate implies a busier and more crowded ED; on the other hand, a lower occupancy rate may be an indicator of excess stretchers or waste.

5.2.5 Future state of Lean EDs

A future state of the ED is presented in a VSM, as shown in Figure 5.3. The ED forecasts patient arrivals and creates a schedule that includes demand profiles of external providers and a departure profile based on arrivals. The departure profile is sent to the discharge process for patient departure preparation as well as hospital units for bed preparation. The demand profiles are given to the corresponding departments so that they can offer timely services to ED patients. When patients arrive, they are triaged, assigned stretchers in the FSS, and assessed by a group of healthcare operators. External services are requested if necessary, and results are normally returned electronically. When all results of the requested services are available, patients are assessed for discharge or admission. Discharged patients are sent home directly while admitted patients may wait for available beds. Clearly, if operators, and the scheduled capacity, are sufficient, patients would flow in the value stream without waiting in the FIFO lanes. The FSS in Figure 5.3 is where the stretchers are located. Upon discharge, an available stretcher flows back to the FSS. The performance of such a Lean system is estimated by an efficiency factor, as mentioned in Chapter 3.2.

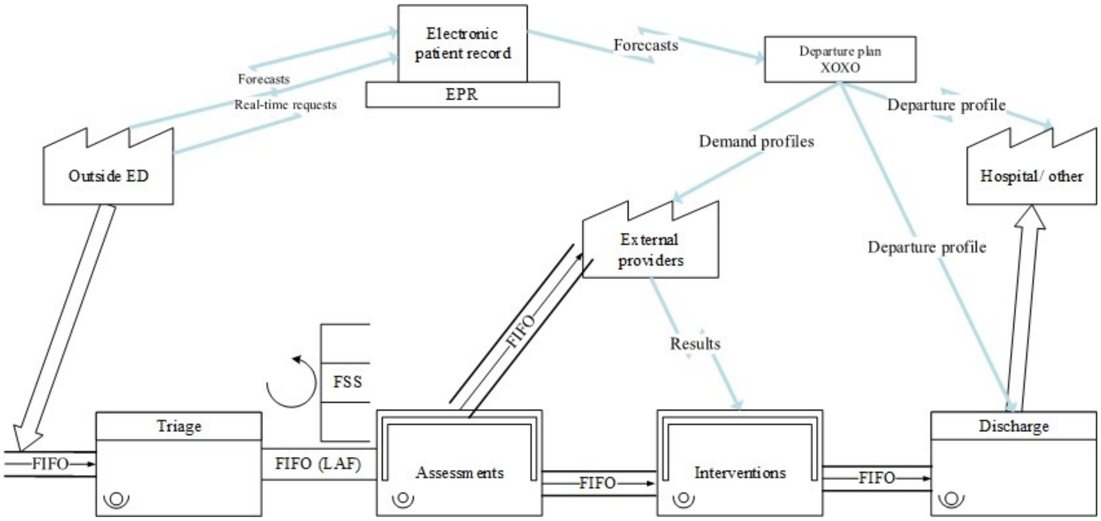


Figure 5.3 Future VSM for EDs

5.3 Case study

A case study was conducted in an ED at a local community hospital in Montreal, Canada, which serves approximately 38,000 ED patients annually. For years, this ED has been persistently trying different methods to improve its system to alleviate ED overcrowding, reduce patient LOS, and provide better patient care. However, it is still undergoing issues of, for example, prolonged wait times and resource shortages (e.g. short of physicians). The study was approved by the Local Research Ethics Committee.

5.3.1 Current system analysis

This research is conducted at the acute side for stretcher patients, and trauma patients are not included. The patient flow is described as follows. Upon arrival, a patient is seen by a triage nurse, assigned a stretcher, and registered by clerks. What follows are nurse observations, a physician's initial assessment, and nurse interventions, including tests and treatments. To create a proper treatment plan before discharge, the physician awaits the results of external services that he/she may have requested. Subsequently, a nurse disconnects equipment, prepares the patient's leave, and finishes charting. Overall, there are three processes: before initial assessment activities (including triage and registration), initial assessment activities (including observation, assessment, and intervention), and reassessment activities (including reassessment and discharge).

The process flow is depicted as a VSM in Figure 5.4. The average data shown in the map were collected through direct observations, interviews with staff and management, and hospital patient records. As illustrated in Figure 5.4, on average, there is a 60.2-minute wait before the PIA, one-hour wait for nursing, 77.4-minute wait for imaging to start, 367.2-minute wait for consults to begin, and 507.8-minute boarding. In addition to the reported shortages of nurses and physicians, available ED operators sometimes cannot start treating patients due to a lack of stretchers. Prolonged wait times to start services and boarding are the result of inefficient coordination between the ED and its support departments. Other issues that are hidden in the ED care process include patient safety concerns. For example, nurses are responsible for the primary activities indicated in Figure 5.4 and hourly patient observations; however, an inadequate nursing staff forces nurses to often overlook hourly observations and make time for the main activities. Furthermore, although deployed with 28 regular stretchers, the ED has to use temporary stretchers placed in corridors to meet the patient demand. Placing patients on temporary stretchers exacerbates ED overcrowding and reduces patient safety. Overall, the efficiency factor is 33.8%, calculated by a 355.2-minute AWCT divided by a 1051.9-minute AWLT. This means that only 33.8% of the time during a patient visit is value-added. The current occupancy rate is 110% per Equation (5.11), in which temporary stretchers are included only in the numerator.

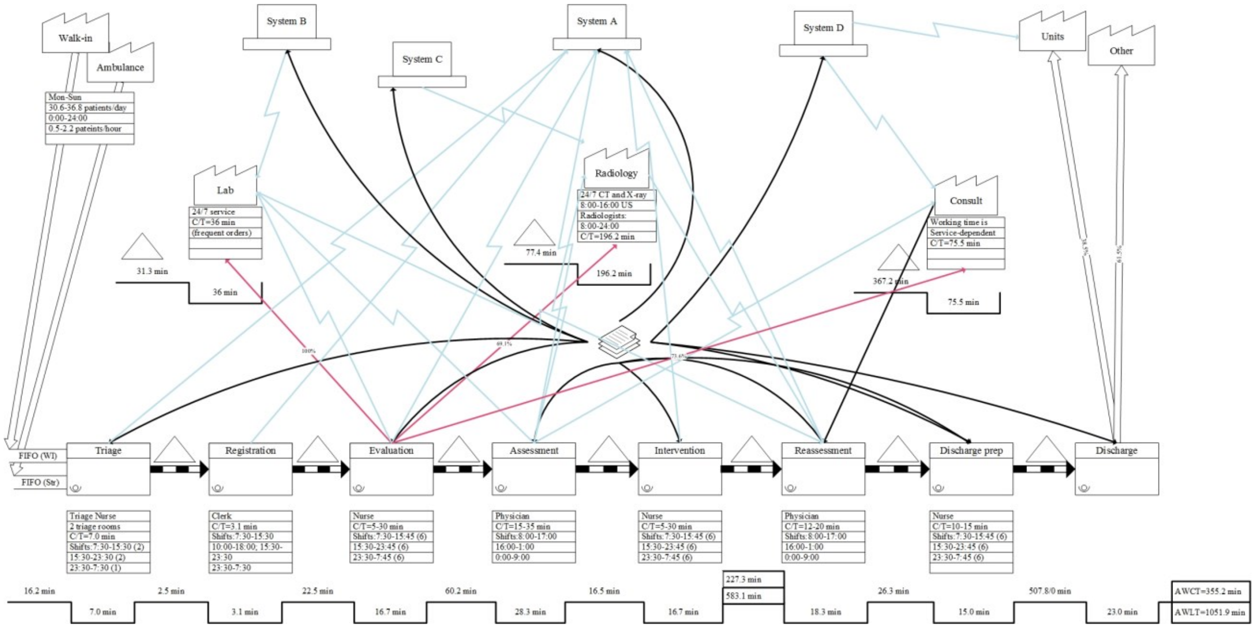


Figure 5.4 Current VSM of the ED

5.3.2 Lean designs

In this chapter, we demonstrate the application of our Lean framework for improving the ED system through an efficient discharge plan supported by proper demand profiles of external providers, sufficient staffing and stretchers, and controlled wait times.

The most recent three years of data on daily ED arrivals were analysed. The resulting boxplots are shown in Figure 5.5, illustrating that the number of daily visits has scarcely increased over the years, has no monthly pattern, and is stable on weekdays but less so on weekends. Therefore, we decided to use the most recent year of data and forecast the ED demand using SES. The observation unit is one week, and daily forecasts were obtained by proportionally distributing the weekly forecast according to Figure 5.5.

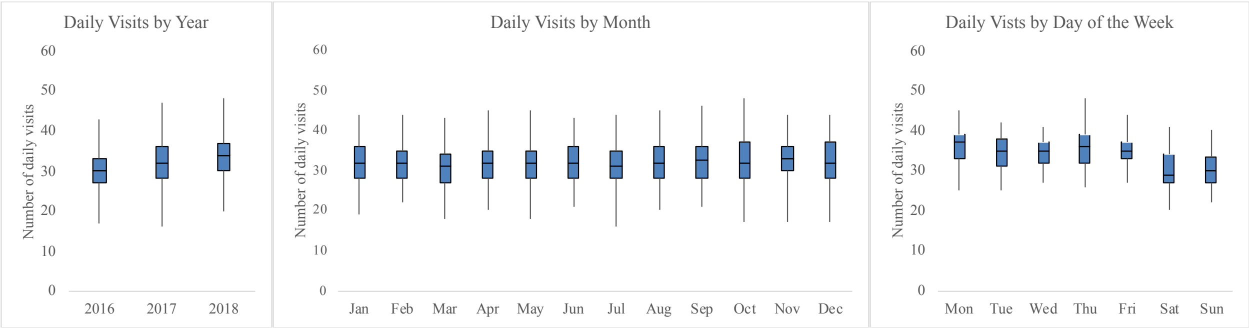


Figure 5.5 Daily Visits by Year, Month and Day of the Week

The SES model for this ED is

$$\hat{y}_t = 0.273 y_t + (1 - 0.273)\hat{y}_{t-1}, \tag{5.12}$$

where $\hat{y}_t$ is the fitted value, and y_t is the actual observation any time period t. A mean average percentage error (MAPE) was used to validate the model. The resulting MAPE is 3.8%, implying the difference between the fitted values and the real observations is only 3.8% on average. Applying this model, we obtain a weekly forecast of 238.5 patients, 35.5 patients on each weekday and 30.5 patients on Saturday and Sunday. Note that the following analysis is for weekdays; however, the same analysis applies to weekends as well.

5.3.2.1 Demand identification

The hourly arrival profile based on historical data and the mean forecast of daily demand is illustrated in Figure 5.6. The chosen DI is one hour based on expert opinion and site observations. Next, we find the demand profiles of imaging, consults, and blood tests.

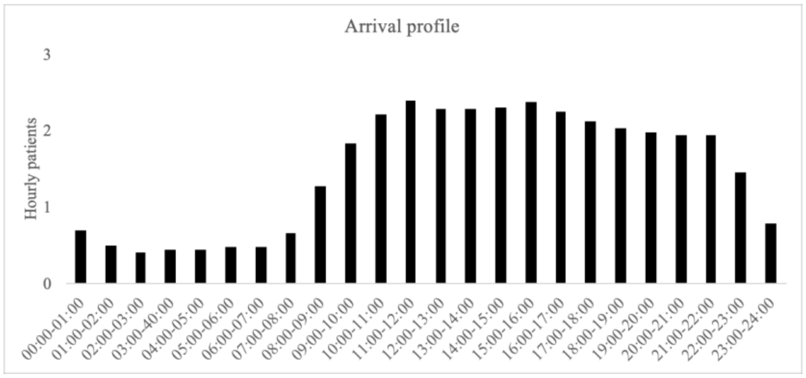


Figure 5.6 Hourly Arrival Profile

In this ED, more than 99% of stretcher patients need bloodwork upon arrival, 69.1% of stretcher patients need imaging, 73.6% of stretcher patients need consults, and only 8% of stretcher patients need neither imaging nor a consult. Consults are given by specialists from 8:00 am to

10:00 pm, while the laboratory works 24 hours a day. Although imaging is also available 24 hours a day, 30% of the tests cannot be done from midnight to 6:00 am. The time from arrival to bloodwork is estimated to be one hour. The PTs of bloodwork, imaging, and consults are 36 minutes, 196.2 minutes, and 75.5 minutes, respectively, as shown in Figure 5.4. Because this research is limited to the ED, we assumed the C/Ts of external services would not change before and after the Lean implementation. The resulting hourly demand profile is illustrated in Figure 5.7 as per Chapter 5.2.1. In determining these profiles, we applied two scheduling strategies. First, we assumed that the 30% of the imaging requests that cannot be completed during the night will be handled from 6:00 am to 8:00 am, which is the current practice for one of the imaging modalities. Second, the consults are evenly distributed into each hour from 8:00 am to 10:00 pm to even out the workload of specialists.

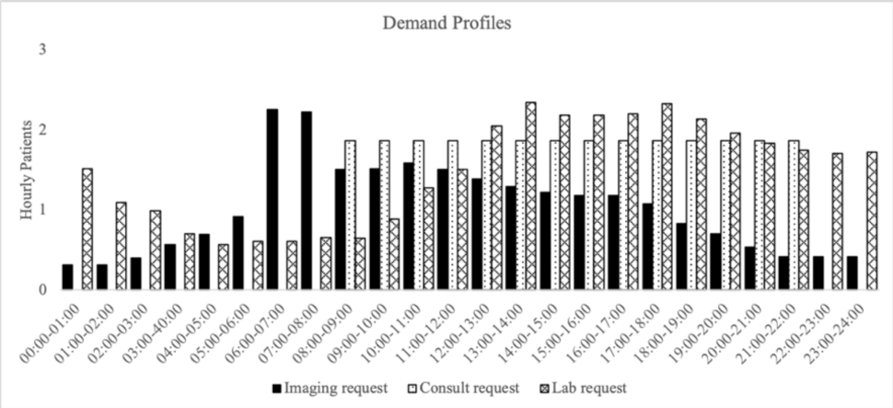


Figure 5.7 Hourly Demand Profiles

The resulting hourly departure profile is illustrated in Figure 5.8, in which the admission rate is 38.5% on average and hospital units work from 8:00 am to 10:00 pm. To facilitate the work of hospital units, we evenly distributed the daily admissions into the working hours of the hospital units, which results in one admission per hour from 8:00 am to 10:00 pm.

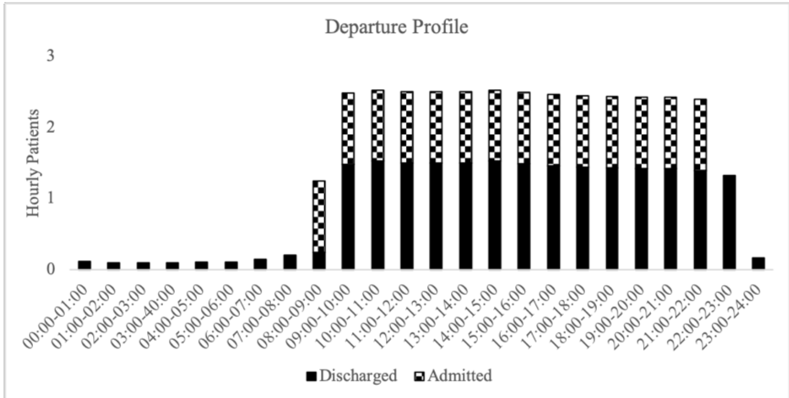


Figure 5.8 Hourly Departure Profile

This departure profile (Figure 5.8) will be used, along with Figure 5.2, to guide the discharge process, while demand profiles (Figure 5.7) will be sent to support departments to ensure that they have enough capacity available to meet the proposed ED requests.

5.3.2.2 Staff scheduling

This section identifies the daily staffing levels and their distribution into appropriate shifts according to Chapter 5.2.2. In this case study, we considered only physicians and nurses, whose working hours are from 8:00 am to 4:00 pm for the day shift, from 4:00 pm to 12:00 am for the evening shift, and from 12:00 am to 8:00 am for the night shift. There are three physicians, one in each shift, and 17 nurses, six in the day shift, six in the evening shift, and five in the night shift. The TT and C/T are given by Equation (3.6) in which the average TWC is 44.6 minutes for physicians and 195.2 minutes for nurses. The 195.2-minute nursing time is composed of time for primary activities and hourly observations, the first of which accounts for 24.8% of the TWC, which is 48.4 minutes as shown in Figure 5.4.

The resulting OBC of physicians and nurses is shown in Figure 5.9, which combines the current and proposed future staffing levels. As we can see from this figure, the current C/Ts of physicians exceed the TT, indicating that more physicians are needed to meet the current demand, whereas the nurse staffing is adequate. To bring the C/Ts of physicians below the TT, we calculated the required number of physicians to be four. With four physicians, the proposed C/Ts of both physicians and nurses are under the TT. The ratio between the C/T and TTs of physicians is 83%, and the ratio between the C/T and TTs of nurses is 85%. These gaps are intentionally left to absorb minor interruptions, as mentioned in Chapter 3.2.2.

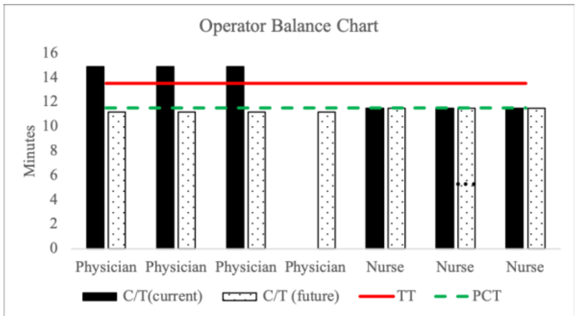


Figure 5.9 Combined Operator Balance Chart

Next, we applied the model described in Chapter 5.2.2 for both physicians and nurses. The demand is obtained from Figure 5.6 and Figure 5.8. The possible start times for the mid-shift that are considered in the model are from 9:00 until 15:00 so that the mid-shift ends before midnight. Table 5.2 lists the resulting discrepancies between demand and capacity, z_k, for all possible mid-shift options. As we can see in this table, the best mid-shift option (identified by the smallest z_k value) for nurses is the one that starts at 1:00 pm and for physicians at 2:00 pm. The optimal staffing levels corresponding to these mid-shifts are shown in Table 5.3.

Table 5.2 Staffing Results

		Start of the mid-shift						
		09:00	10:00	11:00	12:00	13:00	14:00	15:00
z_k	Nurse	4.86	3.62	3.86	4.26	2.82	4.21	5.49
	Physician	375.78	288.76	222.58	176.85	150.58	143.01	153.35

Table 5.3 Proposed Staffing Levels

		Day shift	**Evening shift**	**Night shift**	**Mid-shift**
Nurse	Working hours	8:00–16:00	16:00–24:00	0:00–8:00	13:00–21:00
	Number of operators	7	5	2	3
Physician	Working hours	8:00–16:00	16:00–24:00	0:00–8:00	14:00–22:00
	Number of operators	1	1	1	1

Figure 5.10 compares the cumulative demand and the operator capacity throughout the day and demonstrates that the designed staffing levels are able to handle the daily patients despite the fact that patients between 9:00 to 21:00 may be in wait.

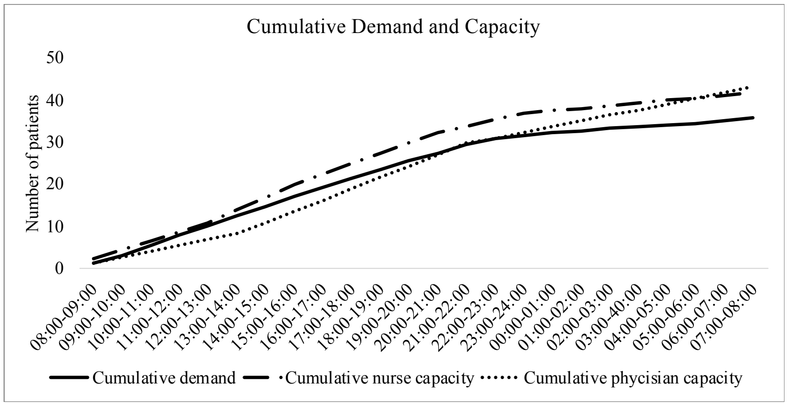


Figure 5.10 Comparison Between Cumulative Demand and Capacity Over Time

5.3.2.3 Wait time control

After the staff scheduling, patients could still be in wait when the demand outpaces capacity within the day as shown in Figure 5.10. The wait times, or the sizes of FIFOs, are presented in the future VSM (Figure 5.11). For example, the FIFO before imaging has 17 cycles: one 8-hour cycle from 0:00 to 8:00, and 16 one-hour cycles from 8:00 to 24:00. In the cycle of 0:00 to 8:00, the total wait times is 14.9 hours, and patients do not wait from 8:00 to 24:00 as the hourly capacity is greater than the hourly demand. Given an average of 24.5 imaging patients per day, the size of the FIFO, which represents the average patient wait time, is 0.6 hours (36.5 minutes) per Equation (5.9).

5.3.2.4 Finished bed supermarket

After the previous designs, the ED LOS is 501.98 minutes on average, including a 38.3-minute boarding time. The size of the FSS is 17 stretchers, calculated by Equation (5.10) in which the effective working time is 1440 minutes, the stretcher cleaning time is 30 minutes, and the standard deviation of the daily demand is 3.4. This FSS is composed of the cycle stock of 13 stretchers, the buffer stock of three stretchers, and the safety stock of one stretcher. In this case study, considering variations in the admission rate (with a mean of 38.5% and standard deviation of 10.6% from historical data), we increased the safety stock to two stretchers, which, under approximately 97.7% of situations, provide sufficient stretchers for boarded patients. The size of the FSS is therefore 18 stretchers with a 64.3% occupancy rate.

5.3.2.5 Future state of the ED

Based on the above discussion, we present the future state of the ED in a VSM as shown in Figure 5.11. A single ED schedule based on a forecast is sent to both the discharge process and the external providers. This schedule includes all the demand profiles that are identified in Chapter 5.3.2.1 to guide and coordinate the ED discharge as well as the daily operations of other departments to accelerate the ED care process. When patients ask for ED care, after triage, they occupy stretchers from the FSS and are immediately assessed and prescribed the necessary tests by ED care providers. Patients would smoothly flow in the value stream without waits, provided that the scheduled capacity is greater than the demand and external departments perform ED requests in a timely fashion with the help of demand profiles. Patients are discharged/admitted by the discharge staff within a DI, and the emptied stretchers are sent back to the FSS. This staff monitors the level of stretchers in the FSS and reports any abnormality to the management as soon as it happens. An additional recommendation is to integrate hospital systems, so the information about patients is shared on one platform and facilities the Lean transformation by quickly informing the support departments of changes in the ED schedule or demand surge. Overall, the

AWLT is 517.1 minutes and the AWCT is 359.1 minutes, which leads to an efficiency factor of 69.4%.

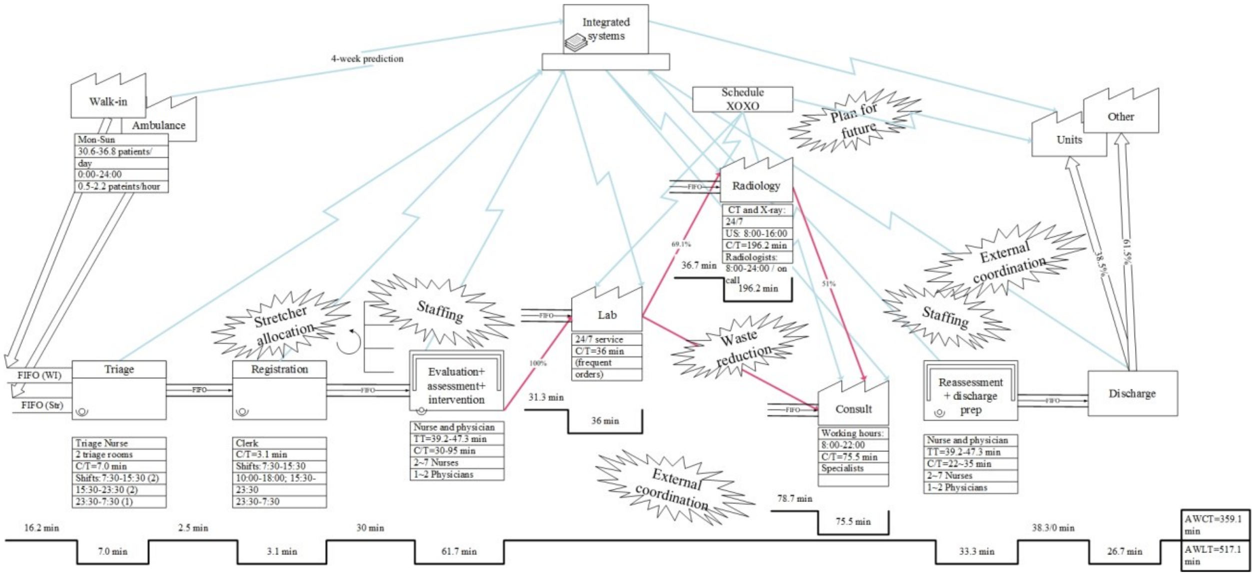


Figure 5.11 Future VSM of the ED

5.3.3 Results and analysis

This section summarises the outcomes of the Lean transformation. Note that due to the limited time span and scope of the case study, the framework has not yet been completely implemented in the real clinical environment, and the results are obtained by calculation and simulation.

A summary of some ED indicators before and after Lean is illustrated in Table 5.4. As it shows, the solutions proposed by this Lean framework reduce the ED LOS (AWLT) by 50.8%, which doubles the efficiency factor (i.e. percent of value-added time). The time to PIA is reduced by 54.1% through the addition of one physician and staff rescheduling. The wait times to start imaging and consults are dramatically reduced by 52.6% and 80% provided that the external providers fulfil the ED requests presented in demand profiles (Figure 5.7). The boarding time is remarkably decreased by 92.4% due to better coordination between the ED and hospital units based on the proposed departure plan (Figure 5.8). Additionally, the occupancy rate of stretchers is reduced by 41.5%. This is discussed in more detail below.

Table 5.4 ED Indicators Before and After Lean

	Indicator (min)					**Indicator (%)**	
	LOS	PIA	Imaging	Consult	Boarding	Efficiency	Occupancy
Before	1050.9	128.2	77.4	367.2	507.8	33.8	110

After	517.1	58.8	36.7	73.8	38.3	69.4	64.3

5.3.3.1 Lead times

Table 5.4 demonstrates the LOS using Lean calculations; however, to evaluate this Lean transformation under stochastic situations that are close to the real clinical environment in which arrivals and C/Ts of processes vary, we developed a simulation model using Rockwell Arena software and compared the simulated LOS before and after the Lean scenarios. In the simulation, patient flow is based on our observations as demonstrated in the VSMs (Figure 5.4 and Figure 5.11), and the processes are governed by time distributions obtained from hospital administrative databases. This model was verified by consulting the ED staff members and validated by comparing a new set of observed data and simulated results.

The results of 60 replications of before and after Lean scenarios were collected (30 for each are usually enough for a statistically sound analysis as we mentioned in Chapter 4.3.3). The resulting boxplot of simulated results before and after Lean is shown in Figure 5.12. As it is shown, the average LOS before Lean is 1105.4 minutes, compared to 1051.9 minutes in the current VSM (Figure 5.4), and with a 90% confidence interval, the LOS is between 1085.5 and 1125.2 minutes. The average LOS after Lean is 555.7 minutes, compared to 517.1 minutes obtained in the future VSM (Figure 5.11), and with a 90% confidence interval, the LOS is between 549.2 and 562.2 minutes. Another improvement we can see from Figure 5.12 is a decreased variability due to streamlined processes and reduced waste. Although the results under the stochastic environment have worsened compared to the deterministic analysis, the solutions offered by our framework demonstrated its abilities to significantly decrease the ED LOS.

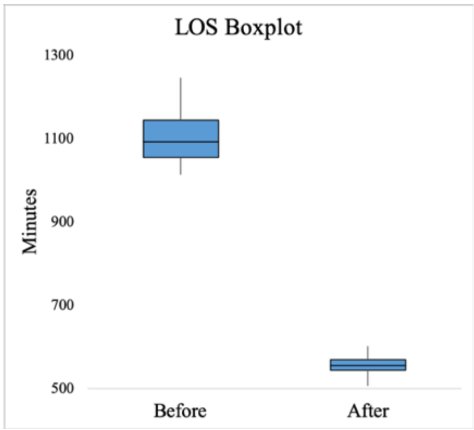


Figure 5.12 Boxplot of the LOS Before and After Lean

5.3.3.2 Other results

The proposed ED system needs one more physician and the same number of nurses. The OBC (Figure 5.9) shows a gap of 83% between the C/T and the TT, indicating an average physician utilisation of 83%, which is slightly lower than the recommended range of 85–90%. However, without this additional physician, the utilisation would be 110%. As for nurses, shift adjustments leave them enough time for both primary activities as well as hourly observations, which possibly increases ED patient safety. Although the OBC (Figure 5.9) indicates that the current nurse staffing is sufficient, future research is needed to finalise this because the situation of two nurses per night shift violates the Nurse Staffing Ratio Law (Tellez & Seago, 2013); on the other hand, the reduced LOS will result in fewer hourly observations, hence a lower nursing workload.

The future state is supported by the demand profiles of external providers (Figure 5.7) and the departure profile (Figure 5.8). The recommendation is that, by using the demand profiles, external departments can schedule their daily operations in advance and offer immediate service to ED patients when requested. Chapter 4 serves as a good example of how to use the demand profile of imaging in a radiology department, in the CT scan process in particular, to devise better patient schedules and reduce patient wait times. The departure profile is used as described in Chapter 5.2.1. Each hour, the ED coordinator takes one card indicating one admission and notifies the related personnel to send a patient to the hospital units. The cards of discharged home patients are used as reminders for the ED team to assess patients who are possibly ready to be discharged.

After applying the framework, the occupancy rate is reduced to 64.3%. A significant result would be the reduction in the incidence of mortality and ED returns according to McCusker et al. (2014) who conducted their research in the same hospital ED. We also identified that, despite no upward trend in Figure 5.5, the current stretcher occupancy rate of 110% and field observations of patients being misplaced to the non-acute side are indications of a possible increase in the number of ED visits, which is hidden due to the shortage of resources. The reduced occupancy would reveal an upward trend of acute patients. If this happens, it is recommended to reapply this framework, update the demand profiles, and recalculate the optimal resource allocation to maintain the improvements.

5.4 Discussion

In this study, we set out to explore a structured framework for the Lean transformation of EDs to deal with ED overcrowding problems and improve ED performance. This framework proposes to identify the demand for ED care and the corresponding ED demand for external services such as imaging. An appropriate staffing plan is designed to meet the hourly demand, optimise the utilisation of operators, and minimise wait times of patients. The wait times in the ED are monitored through controlled time buffers, and the numbers of materials needed to support patient flow are computed. Finally, a future state of the ED with highlighted improvement directions is illustrated in a future VSM. The proposed framework involves these gradual and interactive interventions that lead to Lean EDs where patient flow is created, resources are efficiently allocated, and wait times are reduced, resulting in less crowding.

The case study shows that the ability to identify demand at the beginning of the ED Lean transformation is crucial for designing strategies aimed at ED overcrowding alleviation. While research on the modelling and forecasting of ED arrivals has accounted for a majority of ED overcrowding papers (Carvalho-Silva et al., 2018), this topic is scarcely included in the ED Lean implementations, and it is unclear in the literature how the forecasted arrivals serve the ED in the improvement designs. It is well known that the fundamental philosophy of Lean calls for the production of what the customer wants, when they want it, and in the quantities requested (Womack & Jones, 1997). Therefore, our framework translates this philosophy into synchronising the pace of patient departures with the pace of the arrivals, hence ensuring that incoming patients will successfully pull the ED service when needed. Furthermore, the pace of offering external services is also synchronised through the design of demand profiles to support the desired pace of patient departures. In that sense, from the perspective of ED overcrowding, we connected the crowding factors of input, throughput, and output and, from the perspective of Lean, the framework moves the understanding of the healthcare version of just-in-time towards maturity.

A key concept of Lean is managing buffers (Womack & Jones, 1997), through which the ED LOS can be minimised. According to Hopp and Spearman (2004), there are three types of buffers: inventory buffers, capacity buffers, and time buffers. Our framework identifies all three types of buffers in EDs and proposes solutions to manage them. Capacity buffers are the scheduled resources such as ED operators or machinery that are needed in the core ED care process to treat patients in a timely manner. Inventory buffers refer to the essentials supporting the core ED care process, such as the stretchers. A sufficient capacity and inventory buffers can provide protection against fluctuations in demand and/or in the patient care process. In addition, sufficient inventory buffers decrease the probability of the scheduled capacity being wasted. Time buffers in healthcare refer to the waits of patients who are both the customers and transformed resources in the process. A time buffer is typically controlled with respect to the capacity and inventory buffers. While our framework applies OBCs to obtain the capacity buffer that meets the average ED demand with small fluctuations (interruptions) in the ED care process, the demand surges are not protected by the calculated capacity buffer. However, actions such as temporarily adding operators to treat patients more quickly than normal are suggested when demand surges (i.e. observed wait times are greater than the calculated time buffer in FIFOs). Still, future research on the design of capacity buffers covering demand variability would be an interesting area to further reduce ED overcrowding.

Mitigating the ED overcrowding is not a simple nor short-term effort, and neither is a Lean transformation. This framework follows Lean principles and targets the three interacting ED overcrowding components as much as possible to help the ED achieve a state of reduced wait times and better resource utilisation, resulting in less crowding and better patient care. Despite the fact that the ED case study shows only short-term outcomes, we envision that the long-term effects of applying this framework is to foster a culture of continuous improvement owing to its reproducibility and ability to monitor wait times. However, this perception needs to be validated through future research of the Lean implementation in a longer-term empirical study.

This research has several limitations. It is possible that the outcomes of Lean are not as positive as what we demonstrated in Chapter 5.3.3 because the real clinic environment can be more complicated and affected by operators other than just nurses and physicians, such as social workers. Additionally, an external service may include various modalities done by different operators and/or machines. For example, plain radiography, CT scan, and ultrasound are the three most frequent modalities of imaging conducted by different machinery in the radiology department. This study did not distinguish the demand in the profiles on the basis of the modalities, and thus the obtained improvements could be overestimated owing to the fact that some modalities are less accessible than others in the ED. However, these demand profiles definitely shed some light on system improvement at a hospital-level and could direct further improvement areas. Another limitation is that the framework assumes that patients are triaged by their acuity level at the triage process and that the subsequent treatments are given in a first-in-first-serve manner, which holds true in the case study department but may be different in other EDs.

6. Summary, Conclusions and Future Studies

Full applications of Lean have been less common in healthcare systems than in production systems, for example, in part because Lean practices in the healthcare system are mostly patchy and localised and cannot be easily replicated and in part because of the lack of a well-documented methodology that specifically links Lean concepts with the healthcare system. In this thesis, we targeted three typical types of hospital departments, which are treatment-oriented OPDs, shared processes in OPDs, and EDs, and created systematic frameworks for Lean transformation in these departments.

Treatment-oriented OPDs face issues of unbalanced patient mix, demand variations along the value stream, and daily demand variations attributed to the availability of specialists. In Chapter 3, given the known demand of such departments, we created a Lean framework for treatment-oriented OPDs. The Lean transformation process proposed by this framework starts with identifying the pacemaker capacity and correct patient composition at the pacemaker to obtain a balanced system, followed by creating appropriate staffing levels to support the system. A patient schedule in which patient type and appointment time are specified is then proposed to balance the demand variations in the value stream, improve resource utilisation, and reduce wait times. Wait times in the system are monitored through FIFO lanes in which patients are sequenced in a first-in-first-serve manner. We developed a structured and replicable Lean framework for treatment-oriented OPDs to identify the patient mix, coordinate resources, level schedules, and control waits. The framework was applied to an oncology department that suffers from long patient wait times and low resource utilisation. The results demonstrated that applying the framework helps the oncology department transform from a two-day treatment regime to a one-day treatment regime, thus reducing patient visit time by 36.3% and increasing the number of daily treatments by 38.6%.

Shared processes in OPDs serve both scheduled and unscheduled patients, thus the challenge in their daily operations is to timely meet the urgent demand while offering as many outpatient services as possible with minimum wait times and maintaining a high resource utilisation. The demand is only partially known in such systems. In Chapter 4, we developed a Lean framework for shared processes in OPDs, in which the pacemaker was tested for its ability to meet the urgent demand, and a proper number of electives were scheduled at a desired service level indicating the probability of electives being delayed. Because urgent patients could appear during any time within a day, a carefully designed staffing plan is required to offer a continuous and sufficient service supply. Similar to the treatment-oriented departments, wait times of scheduled patients are monitored through FIFO lanes; however, scheduled patients may wait longer according to the selected service level. The implementation of the framework was illustrated in a radiology department in Montreal, Canada. The results show a 21.4 % increase in the daily number of elective patients, a 77.1% decrease in the visit time of urgent patients, and a 37.2% decrease in the visit time of elective patients through improved patient and staff schedules.

Crowding has been an issue for worldwide EDs and is associated with many small but interrelated factors starting from the upstream process of triage until the very end process of discharge, as well as in other departments whose services are requested by emergency patients. Given the unknown demand in EDs, Chapter 5 introduces a Lean framework for EDs that proposes to work on the pacemaker (discharge) process and identify an optimal departure profile, according to which resources for new arrivals are regularly freed up, thus avoiding crowding. On the other hand, discharge decisions are made by physicians upon the availability of test results; therefore, to support the departure profile, demand profiles including the number of ED requests for external services within the day are proposed based on arrivals and the availability of service-providing departments. Resources are calculated to support the patient flow, and patient wait times are estimated as final steps of Lean transformations in EDs. A case study was undertaken in an ED at a local community hospital in Montreal, Canada. After applying this framework, the patient LOS was reduced by 50.8% from 1051.9 min to 517.1 min, and the occupancy rate of stretchers was reduced by 41.5% from 110% to 64.3%.

The priorities of Lean transformations in treatment-oriented and any appointment-based department are to identify the number of patients that can be accepted by the system and to balance the system so that the variations can be minimised, thus ensuring a smooth patient flow. It is of great importance in shared processes in OPDs to identify or forecast urgent demand before scheduling electives in coordination with the arrival of urgent patients. As for EDs, where arrivals are not controllable, we manage the patient discharge process to meet the incoming demand, with the help of external providers. A similarity in all three situations is that the staffing levels need to run the processes at appropriate speeds to support the known/forecasted demand.

Although the frameworks are utilised in different types of departments for solving different challenges, a unified Lean approach behind these frameworks should be highlighted here. First, this approach starts with a better understanding of value in healthcare in the eyes of customers. It identifies the value from a patient's perspective (e.g. less waiting and better care), the value from a healthcare provider's perspective (e.g. reduced work stress), and the value from other stakeholders' perspectives such as government (e.g. efficient use of resources). Understanding the value allows a better interpretation of the impediments/challenges to increase the value. Subsequently, this approach focuses on delivering perfect service to the customer through a perfect value creation process that has (ideally) zero waste. In achieving this, two questions are asked: what is patient demand, and how to assess major value streams to ensure each step is equipped with the proper resources and treats patients at the right time and that all the steps are connected by flow, pull, and levelling (scheduling). This approach answers these questions by adapting and organising a series of traditional Lean tools for improving the processes of the studied contexts.

In summary, this thesis explored the challenges, those that impede value creation, in three typical types of departments, used a unified Lean approach to establish transformation frameworks, and demonstrated the applications of these frameworks in real clinical settings. The contributions of this thesis are twofold. Theoretically, the framework moves our understanding of Lean in healthcare by emphasising that a Lean transformation is a series of interrelated and gradual rather than fragmented activities associated with the proper use of adapted Lean tools in alignment with Lean principles. Practically, this research is a comprehensive guideline for practitioners to implement Lean in their organisations to reduce wait time and make better decisions in allocating scarce resources to optimise patient care. There are two possible extensions of this research on the hospital level and on the societal level. First, on the hospital level, it raises a question of how the established Lean frameworks can be applied to departments of a single hospital to create a hospital-level patient flow, thus further enhancing hospital performance and reducing operation cost. For example, instead of conducting an individual urgent demand forecast in Chapter 4, how can the demand profile of imaging identified in the ED in Chapter 5 be used to improve the patient schedule for a CT scan process in a radiology department. Second, on the societal level, healthcare organisations that are not within hospitals, such as community clinics, that are similar to any of the studied hospital departments are expected to be able to practice the corresponding Lean framework. Future research is needed to validate this claim.

References

10 IHI Innovations to improve health and health care. (2017). Institute for Healthcare Improvement. http://www.ihi.org/resources/Pages/Publications/10-IHI-Innovations-to-Improve-Health-and-Health-Care.aspx

Aakre, K. T., Valley, T. B., & O'Connor, M. K. (2010). Quality initiatives: Improving patient flow for a bone densitometry practice: Results from a Mayo Clinic Radiology Quality Initiative. *RadioGraphics*, *30*(2), 309–315. https://doi.org/10.1148/rg.302095735

Abouzahra, M., & Tan, J. (2014). Remaking Rosa Medical Center: A 5-step approach to transitioning with Lean. In *Lean Thinking for Healthcare* (pp. 239–264). Springer.

Ahmadi-Javid, A., Jalali, Z., & Klassen, K. J. (2017). Outpatient appointment systems in healthcare: A review of optimization studies. *European Journal of Operational Research*, *258*(1), 3–34. https://doi.org/10.1016/j.ejor.2016.06.064

Al- Araidah, O., Momani, A., Khasawneh, M., & Momani, M. (2010). Lead-time reduction utilizing Lean tools applied to healthcare: The inpatient pharmacy at a local hospital. *Journal for Healthcare Quality*, *32*(1), 59–66. https://doi.org/10.1111/j.1945-1474.2009.00065.x

Amaratunga, T., & Dobranowski, J. (2016). Systematic review of the application of Lean and Six Sigma quality improvement methodologies in Radiology. *Journal of the American College of Radiology*, *13*(9), 1088-1095.e7. https://doi.org/10.1016/j.jacr.2016.02.033

Armony, M., Israelit, S., Mandelbaum, A., Marmor, Y. N., Tseytlin, Y., & Yom-Tov, G. B. (2015). On patient flow in hospitals: A data-based queueing-science perspective. *Stochastic Systems*, *5*(1), 146–194. https://doi.org/10.1287/14-SSY153

Asheim, A., Bjørnsen, L. P. B. W., Næss-Pleym, L. E., Uleberg, O., Dale, J., & Nilsen, S. M. (2019). Real-time forecasting of emergency department arrivals using prehospital data. *BMC Emergency Medicine*, *19*(1), 42. https://doi.org/10.1186/s12873-019-0256-z

Asplin, B. R., Magid, D. J., Rhodes, K. V., Solberg, L. I., Lurie, N., & Camargo, C. A. (2003). A conceptual model of emergency department crowding. *Annals of Emergency Medicine*, *42*(2), 173–180. https://doi.org/10.1067/mem.2003.302

Baccei, S. J., Henderson, S. R., Lo, H. S., & Reynolds, K. (2020). Using quality improvement methodology to reduce costs while improving efficiency and provider satisfaction in a busy, academic Musculoskeletal Radiology Division. *Journal of Medical Systems*, *44*(6), 104.

Baker, M., Taylor, I., & Mitchell, A. (2009). *Making Hospitals Work: How to improve patient care while saving everyone's time and hospitals' resources*. Lean enterprise academy Ltd.

Bal, A., Ceylan, C., & Taçoğlu, C. (2017). Using value stream mapping and discrete event simulation to improve efficiency of emergency departments. *International Journal of Healthcare Management*, *10*(3), 196–206. https://doi.org/10.1080/20479700.2017.1304323

Barak-Corren, Y., Fine, A. M., & Reis, B. Y. (2017). Early prediction model of patient hospitalization from the pediatric emergency department. *Pediatrics*, *139*(5), e20162785.

Baril, C., Gascon, V., Miller, J., & Côté, N. (2016). Use of a discrete-event simulation in a Kaizen event: A case study in healthcare. *European Journal of Operational Research*, *249*(1), 327–339. https://doi.org/10.1016/j.ejor.2015.08.036

Bath, B., Goodridge, D. D., Harrison, D. L., Dobson, D. R., Sari, D. N., Jeffery, C., Lawal, M., Kutz, M., Bourassa, D. C., & Gill, D. (2016). *A first phase evaluation of Saskatchewan's Lean health care transformation: Final report* (p. 55).

Bayley, M. D., Schwartz, J. S., Shofer, F. S., Weiner, M., Sites, F. D., Traber, K. B., & Hollander, J. E. (2005). The financial burden of emergency department congestion and hospital crowding for chest pain patients awaiting admission. *Annals of Emergency Medicine*, *45*(2), 110–117. https://doi.org/10.1016/j.annemergmed.2004.09.010

Beck, M. J., Okerblom, D., Kumar, A., Bandyopadhyay, S., & Scalzi, L. V. (2016). Lean intervention improves patient discharge times, improves emergency department throughput and reduces congestion. *Hospital Practice*, *44*(5), 252–259. https://doi.org/10.1080/21548331.2016.1254559

Belter, D., Halsey, J., Severtson, H., Fix, A., Michelfelder, L., Michalak, K., & De Ianni, A. (2012). Evaluation of outpatient oncology services using lean methodology. *Oncology Nursing Forum*, *39*(2), 136.

Bhat, S., Gijo, E. V., & Jnanesh, N. A. (2014). Application of Lean Six Sigma methodology in the registration process of a hospital. *International Journal of Productivity and Performance Management*, *63*(5), 613–643. https://doi.org/10.1108/IJPPM-11-2013-0191

Bhat, S., Gijo, E. V., & Jnanesh, N. A. (2016). Productivity and performance improvement in the medical records department of a hospital: An application of Lean Six Sigma. *International Journal of Productivity and Performance Management*, *65*(1), 98–125. https://doi.org/10.1108/IJPPM-04-2014-0063

Bhattacharjee, P., & Ray, P. K. (2016). Hospital appointment scheduling in presence of walk-ins and emergency arrivals. In *Smart Technologies for Smart Nations* (pp. 175–194). Springer.

Brett Benfield, C., Brummond, P., Lucarotti, A., Villarreal, M., Goodwin, A., Wonnacott, R., Talley, C., & Heung, M. (2015). Applying lean principles to continuous renal replacement therapy processes. *American Journal of Health-System Pharmacy*, *72*(3), 218–223. https://doi.org/10.2146/ajhp140257

Bucci, S., De Belvis, A. G., Marventano, S., De Leva, A. C., Tanzariello, M., Specchia, M. L., Ricciardi, W., & Franceschi, F. (2016). Emergency Department crowding and hospital bed shortage: Is Lean a smart answer? A systematic review. *European Review for Medical and Pharmacological Sciences*, *20*(20), 4209–19.

Burgess, N., & Radnor, Z. (2013). Evaluating Lean in healthcare. *International Journal of Health Care Quality Assurance*, *26*(3), 220–235. https://doi.org/10.1108/09526861311311418

Byrne, S. C., Barrett, B., & Bhatia, R. (2015). The impact of diagnostic imaging wait times on the prognosis of lung cancer. *Canadian Association of Radiologists Journal*, *66*(1), 53–57. https://doi.org/10.1016/j.carj.2014.01.003

Carter, P. M., Desmond, J. S., Akanbobnaab, C., Oteng, R. A., Rominski, S. D., Barsan, W. G., & Cunningham, R. M. (2012). Optimizing clinical operations as part of a global emergency medicine initiative in Kumasi, Ghana: Application of lean manufacturing principals to low-resource health systems. *Academic Emergency Medicine*, *19*(3), 338–347. https://doi.org/10.1111/j.1553-2712.2012.01311.x

Carvalho-Silva, M., Monteiro, M. T. T., Sá-Soares, F. de, & Dória-Nóbrega, S. (2018). Assessment of forecasting models for patients arrival at Emergency Department. *Operations Research for Health Care*, *18*, 112–118. https://doi.org/10.1016/j.orhc.2017.05.001

Chadha, R., Singh, A., & Kalra, J. (2012). Lean and queuing integration for the transformation of health care processes: A lean health care model. *Clinical Governance: An International Journal*, *17*(3), 191–199. https://doi.org/10.1108/14777271211251309

Chan, H., Lo, S., Lee, L., Lo, W., Yu, W., Wu, Y., Ho, S., Yeung, R., & Chan, J. (2014). Lean techniques for the improvement of patients' flow in emergency department. *World Journal of Emergency Medicine*, *5*(1), 24–28. https://doi.org/10.5847/wjem.j.issn.1920-8642.2014.01.004

Cheung, Y. Y., Goodman, E. M., & Osunkoya, T. O. (2016). No more waits and delays: Streamlining workflow to decrease patient time of stay for Image-guided Musculoskeletal Procedures. *RadioGraphics*, *36*(3), 856–871. https://doi.org/10.1148/rg.2016150174

Collar, R. M., Shuman, A. G., Feiner, S., McGonegal, A. K., Heidel, N., Duck, M., McLean, S. A., Billi, J. E., Healy, D. W., & Bradford, C. R. (2012). Lean management in academic surgery. *Journal of the American College of Surgeons*, *214*(6), 928–936. https://doi.org/10.1016/j.jamcollsurg.2012.03.002

Cookson, D., Read, C., Mukherjee, P., & Cooke, M. (2011). Improving the quality of Emergency Department care by removing waste using lean value stream mapping. *International Journal of Clinical Leadership*, *17*(1).

Costa, L. B. M., & Filho, M. G. (2016). Lean healthcare: Review, classification and analysis of literature. *Production Planning & Control*, *27*(10), 823–836. https://doi.org/10.1080/09537287.2016.1143131

Côté, M. J., & Smith, M. A. (2018). Forecasting the demand for radiology services. *Health Systems*, *7*(2), 79–88. https://doi.org/10.1080/20476965.2017.1390056

Damle, A., Andrew, N., Kaur, S., Orquiola, A., Alavi, K., Steele, S. R., & Maykel, J. (2016). Elimination of waste: Creation of a successful Lean colonoscopy program at an academic medical center. *Surgical Endoscopy*, *30*(7), 3071–3076. https://doi.org/10.1007/s00464-015-4599-6

D'Andreamatteo, A., Ianni, L., Lega, F., & Sargiacomo, M. (2015). Lean in healthcare: A comprehensive review. *Health Policy*, *119*(9), 1197–1209. https://doi.org/10.1016/j.healthpol.2015.02.002

De Brandao, S. L. (2009). Trends and approaches in lean healthcare. *Leadership in Health Services*, *22*(2), 121–139. https://doi.org/10.1108/17511870910953788

De Mast, J., Kemper, B., Does, R. J., Mandjes, M., & van der Bijl, Y. (2011). Process improvement in healthcare: Overall resource efficiency. *Quality and Reliability Engineering International*, *27*(8), 1095–1106. https://doi.org/10.1002/qre.1198

Doğan, N. Ö., & Unutulmaz, O. (2016). Lean production in healthcare: A simulation-based value stream mapping in the physical therapy and rehabilitation department of a public hospital. *Total Quality Management & Business Excellence*, *27*(1–2), 64–80.

Duggan, K. J. (2018). *Creating mixed model value streams: Practical lean techniques for building to demand*. CRC Press.

Duska, L. R., Mueller, J., Lothamer, H., Pelkofski, E. B., & Novicoff, W. M. (2015). Lean methodology improves efficiency in outpatient academic Gynecologic Oncology clinics. *Gynecologic Oncology*, *138*(3), 707–711. https://doi.org/10.1016/j.ygyno.2015.07.001

Fine, B., Golden, B., Hannam, R., & Morra, D. (2009). Leading Lean: A Canadian healthcare leader's guide. *Healthcare Quarterly*, *12*(3), 32–41. https://doi.org/10.12927/hcq.2013.20877

Ford, A. L., Williams, J. A., Spencer, M., McCammon, C., Khoury, N., Sampson, T. R., Panagos, P., & Lee, J.-M. (2012). Reducing door-to-needle times using Toyota's Lean manufacturing principles and value stream analysis. *Stroke*, *43*(12), 3395–3398. https://doi.org/10.1161/STROKEAHA.112.670687

Fournier, P.-L., & Jobin, M.-H. (2018). Understanding before implementing: The context of Lean in public healthcare organizations. *Public Money & Management*, *38*(1), 37–44. https://doi.org/10.1080/09540962.2018.1389505

Geng, N., Chen, L., Liu, R., & Zhu, Y. (2017). Optimal patient assignment for W queueing network in a diagnostic facility setting. *International Journal of Production Research*, *55*(19), 5609–5631. https://doi.org/10.1080/00207543.2017.1324650

Gijo, E. V., & Antony, J. (2014). Reducing patient waiting time in outpatient department using Lean Six Sigma methodology. *Quality and Reliability Engineering International*, *30*(8), 1481–1491. https://doi.org/10.1002/qre.1552

Gill, P. S. (2012). Application of Value Stream Mapping to eliminate waste in an emergency room. *Global Journal of Medical Research*, *12*(6), 51–56.

Gleeson, H., Calderon, A., Swami, V., Deighton, J., Wolpert, M., & Edbrooke-Childs, J. (2016). Systematic review of approaches to using patient experience data for quality improvement in healthcare settings. *BMJ Open*, *6*(8), e011907. https://doi.org/10.1136/bmjopen-2016-011907

Glover, W. J., Farris, J. A., & Aken, E. M. V. (2014). Kaizen events: Assessing the existing literature and convergence of practices. *Engineering Management Journal*, *26*(1), 39–61. https://doi.org/10.1080/10429247.2014.11432003

Gonzalez, C. M., Lau, K. H., & Wickramasinghe, N. (2014). Using value stream mapping to improve processes in a urology department. In *Lean Thinking For Healthcare* (pp. 479–494). Springer.

Gupta, D., & Denton, B. (2008). Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions*, *40*(9), 800–819. https://doi.org/10.1080/07408170802165880

Hallam, C. R. A., & Contreras, C. (2018). Lean healthcare: Scale, scope and sustainability. *International Journal of Health Care Quality Assurance*, *31*(7), 684–696. https://doi.org/10.1108/IJHCQA-02-2017-0023

Hitti, E. A., El-Eid, G. R., Tamim, H., Saleh, R., Saliba, M., & Naffaa, L. (2017). Improving Emergency Department radiology transportation time: A successful implementation of lean methodology. *BMC Health Services Research*, *17*(1), 625. https://doi.org/10.1186/s12913-017-2488-5

Holbrook, A., Glenn, H., Mahmood, R., Cai, Q., Kang, J., & Duszak, R. (2016). Shorter perceived outpatient MRI wait times associated with higher patient satisfaction. *Journal of the American College of Radiology*, *13*(5), 505–509. https://doi.org/10.1016/j.jacr.2015.11.008

Holden, R. J. (2011). Lean thinking in Emergency Departments: A critical review. *Annals of Emergency Medicine*, *57*(3), 265–278. https://doi.org/10.1016/j.annemergmed.2010.08.001

Holden, R. J., & Hackbart, G. (2012). From group work to teamwork: A case study of “Lean” rapid process improvement in the ThedaCare Information Technology Department. *IIE Transactions on Healthcare Systems Engineering*, *2*(3), 190–201. https://doi.org/10.1080/19488300.2012.709584

Hoot, N. R., & Aronsky, D. (2008). Systematic review of emergency department crowding: Causes, effects, and solutions. *Annals of Emergency Medicine*, *52*(2), 126-136.e1. https://doi.org/10.1016/j.annemergmed.2008.03.014

Hopp, W. J., & Spearman, M. L. (2004). To pull or not to pull: What Is the question? *Manufacturing & Service Operations Management*, *6*(2), 133–148. https://doi.org/10.1287/msom.1030.0028

Hospitals see benefits of Lean and Six Sigma. (2009). American Society for Quality. /Read/16654/asq-hospitals-see-benefits-of-lean-six-sigma

Huang, Y. L., & Marcak, J. (2013). Radiology scheduling with consideration of patient characteristics to improve patient access to care and medical resource utilization. *Health Systems*, *2*(2), 93–102. https://doi.org/10.1057/hs.2013.1

Hydes, T., Hansi, N., & Trebble, T. M. (2012). Lean thinking transformation of the unsedated upper gastrointestinal endoscopy pathway improves efficiency and is associated with high levels of patient satisfaction. *BMJ Quality & Safety*, *21*(1), 63–69. https://doi.org/10.1136/bmjqs-2011-000173

Idigo, F. U., Agwu, K. K., Onwujekwe, O. E., Okeji, M. C., & Anakwue, A.-M. C. (2019). Improving patient flows: A case study of a tertiary hospital radiology department. *International Journal of Healthcare Management*, *0*(0), 1–9. https://doi.org/10.1080/20479700.2019.1620476

Improta, G., Romano, M., Di Cicco, M. V., Ferraro, A., Borrelli, A., Verdoliva, C., Triassi, M., & Cesarelli, M. (2018). Lean thinking to improve emergency department throughput at AORN Cardarelli hospital. *BMC Health Services Research*, *18*(1), 914. https://doi.org/10.1186/s12913-018-3654-0

Jayasinha, Y. (2016). Decreasing turnaround time and increasing patient satisfaction in a safety net hospital–based pediatrics clinic using Lean Six Sigma methodologies. *Quality Management in Health Care*, *25*(1), 38–43. https://doi.org/info:doi/10.1097/QMH.0000000000000083

Kadri, F., Harrou, F., Chaabane, S., & Tahon, C. (2014). Time series modelling and forecasting of Emergency Department overcrowding. *Journal of Medical Systems*, *38*(9), 107. https://doi.org/10.1007/s10916-014-0107-0

Kanzaria, H. K., Probst, M. A., Ponce, N. A., & Hsia, R. Y. (2014). The association between advanced diagnostic imaging and ED length of stay. *The American Journal of Emergency Medicine*, *32*(10), 1253–1258. https://doi.org/10.1016/j.ajem.2014.07.038

Karstoft, J., & Tarp, L. (2011). Is Lean Management implementable in a department of radiology? *Insights into Imaging*, *2*(3), 267–273. https://doi.org/10.1007/s13244-010-0044-5

Kenney, C. (2012). *Transforming health care: Virginia Mason Medical Center's pursuit of the perfect patient experience*. CRC Press.

Kinsman, L., Rotter, T., Stevenson, K., Bath, B., Goodridge, D., Harrison, L., Dobson, R., Sari, N., Jeffery, C., & Bourassa, C. (2014). "The largest Lean transformation in the world": The Implementation and evaluation of Lean in Saskatchewan healthcare. *Healthcare Quarterly*, *17*(2), 29–32. https://doi.org/10.12927/hcq.2014.23880

Kruskal, J. B., Reedy, A., Pascal, L., Rosen, M. P., & Boiselle, P. M. (2012). Quality initiatives: Lean approach to improving performance and efficiency in a Radiology Department. *RadioGraphics*, *32*(2), 573–587. https://doi.org/10.1148/rg.322115128

LaGanga, L. R. (2011). Lean service operations: Reflections and new directions for capacity expansion in outpatient clinics. *Journal of Operations Management*, *29*(5), 422–433. https://doi.org/10.1016/j.jom.2010.12.005

Lamm, M. H., Eckel, S., Daniels, R., & Amerine, L. B. (2015). Using lean principles to improve outpatient adult infusion clinic chemotherapy preparation turnaround times. *American Journal of Health-System Pharmacy*, *72*(13), 1138–1146. https://doi.org/10.2146/ajhp140453

Leggat, S. G., Bartram, T., Stanton, P., Bamber, G. J., & Sohal, A. S. (2015). Have process redesign methods, such as Lean, been successful in changing care delivery in hospitals? A systematic review. *Public Money & Management*, *35*(2), 161–168.

Liang, B., Turkcan, A., Ceyhan, M. E., & Stuart, K. (2015). Improvement of chemotherapy patient flow and scheduling in an outpatient oncology clinic. *International Journal of Production Research*, *53*(24), 7177–7190. https://doi.org/10.1080/00207543.2014.988891

Litvak, E., Buerhaus, P. I., Davidoff, F., Long, M. C., McManus, M. L., & Berwick, D. M. (2005). Managing unnecessary variability in patient demand to reduce nursing stress and improve patient safety. *The Joint Commission Journal on Quality and Patient Safety*, *31*(6), 330–338. https://doi.org/10.1016/S1553-7250(05)31044-0

Lowthian, J., Curtis, A., Straney, L., McKimm, A., Keogh, M., & Stripp, A. (2015). Redesigning emergency patient flow with timely quality care at the Alfred. *Emergency Medicine Australasia*, *27*(1), 35–41. https://doi.org/10.1111/1742-6723.12338

Luscombe, R., & Kozan, E. (2016). Dynamic resource allocation to improve emergency department efficiency in real time. *European Journal of Operational Research*, *255*(2), 593–603. https://doi.org/10.1016/j.ejor.2016.05.039

Lyon, A. (2014). Why are normal distributions normal? *The British Journal for the Philosophy of Science*, *65*(3), 621–649. https://doi.org/10.1093/bjps/axs046

MacDonald, S. L., Cowan, I. A., Floyd, R., Mackintosh, S., Graham, R., Jenkins, E., & Hamilton, R. (2013). Measuring and managing radiologist workload: Application of lean and constraint theories and production planning principles to planning radiology services in a major tertiary hospital. *Journal of Medical Imaging and Radiation Oncology*, *57*(5), 544–550. https://doi.org/10.1111/1754-9485.12090

MacKenzie, J., & Hall, W. (2014). Current state of Lean in Canadian health care. *Ottawa, ON: The Conference Board of Canada*.

Marchildon, G. (2013). Implementing Lean health reforms in Saskatchewan. *Health Reform Observer – Observatoire Des Réformes de Santé*, *1*(1). https://doi.org/10.13162/hro-ors.01.01.01

Martin, A. J., Hogg, P., & Mackay, S. (2013). A mixed model study evaluating lean in the transformation of an Orthopaedic Radiology service. *Radiography*, *19*(1), 2–6. https://doi.org/10.1016/j.radi.2012.09.005

Matthias, O., & Brown, S. (2016). Implementing operations strategy through Lean processes within health care: The example of NHS in the UK. *International Journal of Operations & Production Management*, *36*(11), 1435–1457. https://doi.org/10.1108/IJOPM-04-2015-0194

Mazzocato, P., Holden, R. J., Brommels, M., Aronsson, H., Bäckman, U., Elg, M., & Thor, J. (2012). How does lean work in emergency care? A case study of a lean-inspired intervention at the Astrid Lindgren Children's hospital, Stockholm, Sweden. *BMC Health Services Research*, *12*(1), 28. https://doi.org/10.1186/1472-6963-12-28

Mazzocato, P., Savage, C., Brommels, M., Aronsson, H., & Thor, J. (2010). Lean thinking in healthcare: A realist review of the literature. *BMJ Quality & Safety*, *19*(5), 376–382. https://doi.org/10.1136/qshc.2009.037986

Mazzocato, P., Stenfors-Hayes, T., von Thiele Schwarz, U., Hasson, H., & Nyström, M. E. (2016). Kaizen practice in healthcare: A qualitative analysis of hospital employees' suggestions for improvement. *BMJ Open*, *6*(7), e012256. https://doi.org/10.1136/bmjopen-2016-012256

McCulloch, P., Kreckler, S., New, S., Sheena, Y., Handa, A., & Catchpole, K. (2010). Effect of a "Lean" intervention to improve safety processes and outcomes on a surgical emergency unit. *BMJ*, *341*. https://doi.org/10.1136/bmj.c5469

McCusker, J., Vadeboncoeur, A., Lévesque, J.-F., Ciampi, A., & Belzile, E. (2014). Increases in Emergency Department occupancy are associated with adverse 30-day outcomes. *Academic Emergency Medicine*, *21*(10), 1092–1100. https://doi.org/10.1111/acem.12480

McDermott, A. M., Kidd, P., Gately, M., Casey, R., Burke, H., O'Donnell, P., Kirrane, F., Dinneen, S. F., & O'Brien, T. (2013). Restructuring of the Diabetes Day Centre: A pilot lean project in a tertiary referral centre in the West of Ireland. *BMJ Quality & Safety*, *22*(8), 681–688. https://doi.org/10.1136/bmjqs-2012-001676

McDermott, C. M., & Venditti, F. J. (2015). Implementing lean in knowledge work: Implications from a study of the hospital discharge planning process. *Operations Management Research*, *8*(3), 118–130. https://doi.org/10.1007/s12063-015-0103-7

Migita, R., Del Beccaro, M., Cotter, D., & Woodward, G. A. (2011). Emergency Department overcrowding: Developing Emergency Department capacity through process improvement. *Clinical Pediatric Emergency Medicine*, *12*(2), 141–150. https://doi.org/10.1016/j.cpem.2011.04.001

Ministère de la santé et des services sociaux, Quebec, Direction des communications, & Association québécoise d'établissements de santé et de services sociaux. (2006). *Guide de gestion de l'urgence.* Association québécoise d'établissements de santé et de services sociaux.

Moe, J., Kirkland, S. W., Rawe, E., Ospina, M. B., Vandermeer, B., Campbell, S., & Rowe, B. H. (2017). Effectiveness of interventions to decrease Emergency Department visits by adult frequent users: A systematic review. *Academic Emergency Medicine*, *24*(1), 40–52. https://doi.org/10.1111/acem.13060

Moore, S., & Arthur, R. (2019). A Quality Improvement Initiative: Using Lean methodology to improve efficiency of the morning cycle monitoring at an ambulatory academic fertility clinic. *Journal of Obstetrics and Gynaecology Canada*, *41*(6), 755–761. https://doi.org/10.1016/j.jogc.2018.10.027

Murrell, K. L., Offerman, S. R., & Kauffman, M. B. (2011). Applying Lean: Implementation of a rapid triage and treatment system. *Western Journal of Emergency Medicine*, *12*(2), 184–191.

Naik, T., Duroseau, Y., Zehtabchi, S., Rinnert, S., Payne, R., McKenzie, M., & Legome, E. (2012). A structured approach to transforming a large public hospital emergency department via Lean methodologies. *Journal for Healthcare Quality*, *34*(2), 86–97. https://doi.org/10.1111/j.1945-1474.2011.00181.x

Naiker, U., FitzGerald, G., Dulhunty, J. M., & Rosemann, M. (2018). Time to wait: A systematic review of strategies that affect out-patient waiting times. *Australian Health Review*, *42*(3), 286–293.

Newell, T. L., Steinmetz- Malato, L. L., & Dyke, D. L. V. (2011). Applying Toyota Production System techniques for medication delivery: Improving hospital safety and efficiency. *Journal for Healthcare Quality*, *33*(2), 15–22. https://doi.org/10.1111/j.1945-1474.2010.00104.x

Ng, D., Vail, G., Thomas, S., & Schmidt, N. (2010). Applying the Lean principles of the Toyota Production System to reduce wait times in the emergency department. *Canadian Journal of Emergency Medicine*, *12*(1), 50–57. https://doi.org/10.1017/S1481803500012021

Niemeijer, G. C., Trip, A., Ahaus, K. T. B., Does, R. J., & Wendt, K. W. (2010). Quality in trauma care: Improving the discharge procedure of patients by means of Lean Six Sigma. *Journal of Trauma and Acute Care Surgery*, *69*(3), 614–619. https://doi.org/10.1097/TA.0b013e3181e70f90

Ohno, T. (1988). *Toyota Production System: Beyond large-scale production*. CRC Press.

Parkhi, S. S. (2019). Lean management practices in healthcare sector: A literature review. *Benchmarking: An International Journal*, *26*(4), 1275–1289. https://doi.org/10.1108/BIJ-06-2018-0166

Patrick, J., & Puterman, M. L. (2007). Improving resource utilization for diagnostic services through flexible inpatient scheduling: A method for improving resource utilization. *Journal of the Operational Research Society*, *58*(2), 235–245. https://doi.org/10.1057/palgrave.jors.2602242

Piggott, Z., Weldon, E., Strome, T., & Chochinov, A. (2011). Application of Lean principles to improve early cardiac care in the emergency department. *Canadian Journal of Emergency Medicine*, *13*(5), 325–332. https://doi.org/10.2310/8000.2011.110284

Powell, D., Riezebos, J., & Strandhagen, J. O. (2013). Lean production and ERP systems in small- and medium-sized enterprises: ERP support for pull production. *International Journal of Production Research*, *51*(2), 395–409. https://doi.org/10.1080/00207543.2011.645954

Radnor, Z. J., Holweg, M., & Waring, J. (2012). Lean in healthcare: The unfilled promise? *Social Science & Medicine*, *74*(3), 364–371. https://doi.org/10.1016/j.socscimed.2011.02.011

Rico, R. A., & Jagwani, J. M. (2013). Application of lean methods to compounding services in hospital pharmacy. *European Journal of Hospital Pharmacy: Science and Practice*, *20*(3), 168–173. https://doi.org/10.1136/ejhpharm-2012-000221

Robinson, S., Radnor, Z. J., Burgess, N., & Worthington, C. (2012). SimLean: Utilising simulation in the implementation of lean in healthcare. *European Journal of Operational Research*, *219*(1), 188–197. https://doi.org/10.1016/j.ejor.2011.12.029

Roemeling, O. P., Land, M. J., Ahaus, K., Slomp, J., & van den Bijllaardt, W. (2017). Impact of lean interventions on time buffer reduction in a hospital setting. *International Journal of Production Research*, *55*(16), 4802–4815. https://doi.org/10.1080/00207543.2017.1301687

Rother, M., & Shook, J. (2003). *Learning to see: Value stream mapping to add value and eliminate muda*. Lean Enterprise Institute.

Rutledge, J., Xu, M., & Simpson, J. (2010). Application of the Toyota Production System improves core laboratory operation. *American Journal of Clinical Pathology*, *133*(1), 24–31. https://doi.org/10.1309/AJCPD1MSTIVZI0PZ

Sanz, M. D., Yagüe- Fabra, J. A., & Matilla, R. G. (2019). Use of Lean techniques in health care in Spain to improve involvement and satisfaction of workers. *The International Journal of Health Planning and Management*, *34*(1), e274–e290. https://doi.org/10.1002/hpm.2646

Schull, M. J., Mamdani, M. M., & Fang, J. (2004). Community influenza outbreaks and emergency department ambulance diversion. *Annals of Emergency Medicine*, *44*(1), 61–67. https://doi.org/10.1016/j.annemergmed.2003.12.008

Shah, C. J., Sullivan, J. R., Gonyo, M. B., Wadhwa, A., & DuBois, M. S. (2013). Practice policy and quality initiatives: Using lean principles to improve screening mammography workflow. *RadioGraphics*, *33*(5), 1505–1517. https://doi.org/10.1148/rg.335135017

Shah, R., & Ward, P. T. (2007). Defining and developing measures of lean production. *Journal of Operations Management*, *25*(4), 785–805. https://doi.org/10.1016/j.jom.2007.01.019

Shi, P., Chou, M. C., Dai, J. G., Ding, D., & Sim, J. (2015). Models and insights for hospital inpatient operations: Time-dependent ED boarding time. *Management Science*, *62*(1), 1–28. https://doi.org/10.1287/mnsc.2014.2112

Shook, J., & Marchwinski, C. (2014). *Lean Lexicon: A graphical glossary for Lean Thinkers*. Lean Enterprise Institute.

Silvester, K., Lendon, R., Bevan, H., Steyn, R., & Walley, P. (2004). Reducing waiting times in the NHS: Is lack of capacity the problem? *Clinician in Management*, *12*(3).

Simons, F. E., Aij, K. H., Widdershoven, G. A., & Visse, M. (2014). Patient safety in the operating theatre: How A3 thinking can help reduce door movement. *International Journal for Quality in Health Care*, *26*(4), 366–371.

Skeldon, S. C., Simmons, A., Hersey, K., Finelli, A., Jewett, M. A., Zlotta, A. R., & Fleshner, N. E. (2014). Lean methodology improves efficiency in Outpatient Academic Uro-oncology Clinics. *Urology*, *83*(5), 992–998. https://doi.org/10.1016/j.urology.2013.11.048

Smith, C., Wood, S., & Beauvais, B. (2011). Thinking Lean: Implementing DMAIC methods to Improve efficiency within a Cystic Fibrosis Clinic. *Journal for Healthcare Quality*, *33*(2), 37–46. https://doi.org/10.1111/j.1945-1474.2010.00130.x

Smith, G., Poteat-Godwin, A., Harrison, L. M., & Randolph, G. D. (2012). Applying Lean principles and Kaizen Rapid Improvement Events in public health practice. *Journal of Public Health Management and Practice*, *18*(1), 52–54. https://doi.org/10.1097/PHH.0b013e31823f57c0

Spear, S. J. (2005). Fixing health care from the inside, today. *Harvard Business Review*, *83*(9), 78.

Storrow, A. B., Zhou, C., Gaddis, G., Han, J. H., Miller, K., Klubert, D., Laidig, A., & Aronsky, D. (2008). Decreasing lab turnaround time improves Emergency Department throughput and decreases emergency medical services diversion: A simulation model. *Academic Emergency Medicine*, *15*(11), 1130–1135. https://doi.org/10.1111/j.1553-2712.2008.00181.x

Teichgräber, U. K., & de Bucourt, M. (2012). Applying value stream mapping techniques to eliminate non-value-added waste for the procurement of endovascular stents. *European Journal of Radiology*, *81*(1), e47–e52. https://doi.org/10.1016/j.ejrad.2010.12.045

Tellez, M., & Seago, J. A. (2013). California nurse staffing law and RN workforce changes. *Nursing Economics*, *31*(1), 18–26.

Vasanawala, S. S., & Desser, T. S. (2005). Accommodation of requests for emergency US and CT: Applications of queueing theory to scheduling of urgent studies. *Radiology*, *235*(1), 244–249. https://doi.org/10.1148/radiol.2351040289

Vashi, A. A., Sheikhi, F. H., Nashton, L. A., Ellman, J., Rajagopal, P., & Asch, S. M. (2019). Applying Lean principles to reduce wait times in a VA emergency department. *Military Medicine*, *184*(1–2), e169–e178. https://doi.org/10.1093/milmed/usy165

Velt, K. B., Cnossen, M., Rood, P. P. M., Steyerberg, E. W., Polinder, S., & Lingsma, H. F. (2018). Emergency department overcrowding: A survey among European neurotrauma centres. *Emergency Medicine Journal*, *35*(7), 447–448. https://doi.org/10.1136/emermed-2017-206796

Verbano, C., & Crema, M. (2019). Applying lean management to reduce radiology turnaround times for emergency department. *The International Journal of Health Planning and Management*, *34*(4), e1711–e1722. https://doi.org/10.1002/hpm.2884

Vermeulen, M. J., Stukel, T. A., Guttmann, A., Rowe, B. H., Zwarenstein, M., Golden, B., Nigam, A., Anderson, G., Bell, R. S., Schull, M. J., Afilalo, M., Anderson, G., Bell, R. S., Carew, D., Carter, M., Cooke, M., Golden, B., Guttmann, A., Nigam, A., … Zwarenstein, M. (2014). Evaluation of an Emergency Department Lean process improvement program to reduce length of stay. *Annals of Emergency Medicine*, *64*(5), 427–438. https://doi.org/10.1016/j.annemergmed.2014.06.007

Vogenberg, F. R., & Santilli, J. (2018). Healthcare Trends for 2018. *American Health & Drug Benefits*, *11*(1), 48–54.

Wang, H., Kline, J. A., Jackson, B. E., Robinson, R. D., Sullivan, M., Holmes, M., Watson, K. A., Cowden, C. D., Phillips, J. L., Schrader, C. D., Leuck, J., & Zenarosa, N. R. (2017). The role of patient perception of crowding in the determination of real-time patient satisfaction at Emergency Department. *International Journal for Quality in Health Care*, *29*(5), 722–727. https://doi.org/10.1093/intqhc/mzx097

Wang, T. K., Yang, T., Yang, C. Y., & Chan, F. T. (2015). Lean principles and simulation optimization for emergency department layout design. *Industrial Management & Data Systems*, *115*(4), 678–699. https://doi.org/10.1108/IMDS-10-2014-0296

Waring, J. J., & Bishop, S. (2010). Lean healthcare: Rhetoric, ritual and resistance. *Social Science & Medicine*, *71*(7), 1332–1340. https://doi.org/10.1016/j.socscimed.2010.06.028

White, B. A., Biddinger, P. D., Chang, Y., Grabowski, B., Carignan, S., & Brown, D. F. (2013). Boarding inpatients in the Emergency Department increases discharged patient length of stay. *The Journal of Emergency Medicine*, *44*(1), 230–235. https://doi.org/10.1016/j.jemermed.2012.05.007

White, B. A., Chang, Y., Grabowski, B. G., & Brown, D. F. (2014). Using Lean-based systems engineering to increase capacity in the Emergency Department. *Western Journal of Emergency Medicine*, *15*(7), 770–776. https://doi.org/10.5811/westjem.2014.8.21272

Womack, J. P., & Jones, D. T. (1997). Lean thinking—Banish waste and create wealth in your corporation. *Journal of the Operational Research Society*, *48*(11), 1148–1148.

Wong, R., Levi, A. W., Harigopal, M., Schofield, K., & Chhieng, D. C. (2012). The positive impact of simultaneous implementation of the BD FocalPoint GS Imaging System and Lean Principles on the operation of Gynecologic Cytology. *Archives of Pathology & Laboratory Medicine*, *136*(2), 183–189. https://doi.org/10.5858/arpa.2011-0139-OA

Wong, T. C., Xu, M., & Chin, K. S. (2014). A two-stage heuristic approach for nurse scheduling problem: A case study in an emergency department. *Computers & Operations Research*, *51*, 99–110. https://doi.org/10.1016/j.cor.2014.05.018

Young, T. P., & McClean, S. I. (2008). A critical look at Lean thinking in healthcare. *BMJ Quality & Safety*, *17*(5), 382–386. https://doi.org/10.1136/qshc.2006.020131

Yousri, T. A., Khan, Z., Chakrabarti, D., Fernandes, R., & Wahab, K. (2011). Lean thinking: Can it improve the outcome of fracture neck of femur patients in a district general hospital? *Injury*, *42*(11), 1234–1237. https://doi.org/10.1016/j.injury.2010.11.024