Real-Time Simulation of Patient Care Processes in Healthcare

Sepideh Bahrani

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of

Master of Science in Systems Science

uOttawa
University of Ottawa
Ottawa, Ontario, Canada

July 2013

Abstract

The increasing waiting times to access healthcare services are a major concern for patients in hospitals. Due to the unpredictability of health issues, hospitals and clinical services are provided to patients even without prescheduled medical appointments. Unexpected and random patient arrivals can result in high waiting times. Waiting occurs mostly because of insufficient resources available compared to demanding service delivery requirements at a given time. Thus, appropriate management of resource scheduling over time can help reduce patient wait times.

So far, simulation has mostly been used as a support for *strategic* decision making in healthcare environments. We are proposing a complementary approach, namely, *real-time simulation*, to support *operational* decision making rather than long-term strategic decision making. Real-time simulation is a technique used to get a timely prediction of the system status in a near future (e.g., a few hours). Hospitals can benefit from the capabilities of real-time simulations by predicting upcoming bottleneck occurrences in patient care processes and make effective decisions in the present time to avoid undesirable outcomes in the near future.

This research presents real-time simulation capabilities for short-term operational decision making of patient care processes in hospitals and the possible ways to run alternative scenarios and evaluate their results to come up with the most effective solution considering various factors. This thesis also provides tool support based on a leading simulation environment, namely Arena. The tool-supported methodology is evaluated through a realistic cardiac care process in an Ontario community hospital, with encouraging results.

Acknowledgment

First and foremost, I would like to express my sincere gratitude to my supervisor, Professor Daniel Amyot, for his continuous support, during my masters studies. I am so honored to have experienced the supervision of someone I truly admire as an expert, thinker and human being. This dissertation definitely would have not been possible without his effective guidance, patience, motivation, enthusiasm and encouragement. I will always remain grateful to him for the rest of my life.

I would like to thank the Osler project team, and especially Alain Mouttham for giving me the opportunity to work in his team and for providing me with the essentials required for conducting this research. His constant effort and incredible motivation for guiding the team in achieving our goals were truly inspiring.

I would like to express my deepest appreciation to Renaud Bougueng Tchemeube, for the brilliance of his mind and ideas, for his generosity in sharing his wide knowledge on every life-related component, and for light guiding my way through the dark (which happened a lot). I am so thankful for all his kindness.

A special thanks to Alexandre Ouellet, Director of modeling solutions at Trellisys, for his contributions in stimulating constructive suggestions and for being supportive of the whole research. His valuable and precious knowledge has definitely facilitated both the learning and experimentation process behind this research.

I owe my deepest gratitude to my beloved family, without whom none of this could have been possible. To my parents, who have been incredibly supportive and have been my main motivation: my mother, Farangis, whom I have always looked up to and who has shown me how to stay strong and make the impossible become possible; my father, Mansour, who inspired me to reach my dreams and who has always caught me when I was falling; and to my lovely sister, Setareh, who brought me infinite joy and endless care.

Last but not least, I would like to thank NSERC and the Business Intelligence Network for funding this work, and the University of Ottawa for providing the required facilities and a great environment in which the students are enabled to work productively towards raising their academic experience.

Table of Contents

List of Figures

List of Tables

List of Acronyms

Acronym	**Definition**
AHP	Analytic Hierarchy Process
BP	Business Process
BPM	Business Process Manager
BPMN	Business Process Model and Notation
CCL	Cardiac Catheterization Lab
CTAS	Canadian Triage and Acuity Scale
CV	Coefficient of Variation
CW	Cardiology Ward
DES	Discrete Event Simulation
DSS	Decision Support System
ED	Emergency Department
ER	Emergency Room
ICU	Intensive Care Unit
LWBS	Leave Without Being Seen
MOE	Margin of Error
NEDOCS	National Emergency Department Overcrowding Scale
OD	Outpatient Department
OR	Operation Room
PAN	Process Analyzer
PCI	Percutaneous Coronary Intervention
TL	Test Laboratories
TOPSIS	Technique Ordered Preference by Similarity to the Ideal Solution
VBA	Visual BASIC for Applications

Chapter 1. Introduction

This thesis presents a novel and tool-supported real-time simulation approach for healthcare processes in order to support short-term, operational decision making.

1.1. Context and Motivation

Recently, improvement in patient care has been highly valued and has drawn a lot of attention within the healthcare domain. Improvement in hospital process performance increases the efficiency and productivity of the overall hospital as well as the patient satisfaction level. One of the possible ways of measuring the quality of performance in operation and management processes (or at least its outcome) is through the patient satisfaction level [42]. However, precise analysis of process outcomes can help elucidate deficiencies at the system level. Managers will then be able to find approaches and make decisions that will help minimizing these deficiencies. Patient waiting time is one of the most important factors in the patient satisfaction level [46]. Any raise of this factor will also have a significant impact on cost. This cost does not only involve monetary loss, it also can result in life-threatening consequences [37]. Table 1 shows the average patient waiting time for each Canadian Triage and Acuity Scale (CTAS) level in hospitals across Ontario, Canada. This table highlights that the patient waiting time for admitted patients is quite higher than national targets.

Table 1 Time Spent in Emergency Room in Ontario, Canada [32]

Time spent in ER (Hours) by CTAS and by Patient Type

Provincial

October-November-December 2012												
	Admitted Patients (Hours)						Non-admitted Patients (Hours)					
CTAS Level	Target	Cases	Average Wait Time	Median Wait Time	90% completed within	Percent Completed within Target	Target	Cases	Average Wait Time	Median Wait Time	90% completed within	Percent Completed within Target
All Patients	N/A	140685	14.5	9.9	29.9	N/A	N/A	1158564	3.2	2.5	6.2	N/A
CTAS I	8	6015	11.4	6.3	26.9	59	8	4309	4.5	3.4	8.7	88
CTAS II	8	62771	15.0	10.1	30.8	39	8	175967	4.6	3.8	8.1	90
CTAS III	6	64929	14.7	10.3	29.6	25	6	521675	3.6	3.0	6.7	86
CTAS IV	4	6547	11.9	7.5	27.2	23	4	409896	2.2	1.8	4.2	89
CTAS V	4	369	9.2	5.7	20.9	37	4	46130	1.8	1.4	3.5	93

Much consideration has been devoted to Emergency Department (ED) congestion issues for finding effective approaches to reduce patient waiting times. This is particularly relevant given that ED is where all types of patients with critical health conditions need to be prioritized for immediate clinical support. However, transferring the population in a hospital from one department to another, with the intention of locally optimizing the patient flow in one given department, is not necessarily an efficient operational management strategy, because the throughput time as well as the total waiting time for a patient may still remain the same (or get worse) during the patient's journey across several units of the hospital. Global optimizations at the patient flow level rather than local optimizations at the department or unit level is what should be targeted. With regard to the latter observation, patient waiting time is a valid criterion for measuring the care process efficiency, as observed by Williams et al. [44] and Koizumi et al. [23].

In general, through a process where event occurrences are pre-organized and scheduled (e.g., in scheduled surgeries), performance outcome predictions are much closer to the real outcomes, compared to the situation where event occurrences are random (e.g., in emergencies or cardiology). Due to the random nature of patient arrivals in emergency departments of hospitals, managing not only the ED, but also the other departments connected to ED [5][42], is a challenging task. The patient population entering from ED will spread to all of the other departments to continue with the rest of their care processes.

Among different types of clinical services, cardiology departments have always been of significant importance, due to the magnitude of the cardiac diseases. Urgent care processes are required to take place in time for the cardiac patients and thus, time plays an important role in this regard. Most cardiac problems occur unexpectedly and therefore a lot of patients referred to the cardiac departments begin their journey from ED. However, for patients with more severe health issues (i.e., lower CTAS levels), the care process is not over after the ED. These patients will be referred typically to the cardiology ward (CW) and if necessary to the cardiac catheterization lab (CCL) for surgery.

For solving the issues related to patient waiting times, effective decisions should be made and actions should take place accordingly. Managerial decisions can help reduce the patient waiting time, *if* the candidate actions and their impact are clear to the decision makers. Thus, the statement "making the right decisions at the right time" can become true if the means for supporting short-term operational decisions are available.

The research conducted in this thesis emphasizes the effectiveness of *real-time* simulation capabilities for supporting short-term operational decision making in hospitals. In other words, the thesis focuses on how predicting the upcoming short-term behaviour of a hospital in a timely manner can help support the process of short-term decision making. A generic model of a cardiology department has been created. Approximate but realistic input data is fed to the simulation model and the model starts running using the current state of the hospital's process instances as initial state. The results from what-if scenarios are observed. These scenarios are actions typically used to mitigate potential bottlenecks and overcome the congestion through an entire clinical pathway. However, this whole procedure is not for a one-time (strategic) decision making. For a more effective (short-term or operational) decision-making process, frequent evaluations of the system performance are required. Hence, the system can exhibit a more timely and continuous adaptation to a range of unexpected event occurrences. Using real-time simulation techniques that support continuous short-term decision making related to patient care processes, according to the event occurrences, is the aim of this research.

In this research, a healthcare process the William Osler Health Center hospital (simply referred to as Osler) located in Brampton, Canada, has been taken as a case study. Osler is a large community hospital and was chosen to be a representative of a typical hospital and of its issues. One of their most common end-to-end clinical pathways involving cardiac patients is being used in this thesis. The patient entrance point is from the cardiac emergency department (ED) and after required investigation on physical status is done, the patient is transferred to the cardiology ward (CW) where he/she would be assigned a bed. The patient is then taken to the cardiac catheterization lab (CCL) for an angiogram and/or a Percutaneous Coronary Intervention (PCI), before returning to the cardiology ward for recovery before being discharged. This pathway is modeled using a leading commercial simulation tool, namely Arena v.14.

1.2. Research Question

The research question addressed in this thesis is as follows: *Is the real-time simulation of clinical processes feasible in the context of a patient care process, with wait time predictions represented in a way that supports short-term operational decision making related to resource allocation?*

The emphasis in our research question is on assessing the feasibility of real-time simulation capabilities. The requirements for building a real-time simulation model will be investigated and if the feasibility of the approach is confirmed, more investigation will be devoted to the benefits of using this approach in facilitating operational decision making.

1.3. Simulations

The patient waiting time problem often occurs as a result of allocating a limited number of resources to unlimited or unpredictable numbers of patients. An effective approach leading to facilitating operational decision making is essential in this context. *Simulation* is one of the most effective techniques for decision making and is increasingly being used in the healthcare domain. Simulation enables us to design and analyze the processes involved in a certain context, without necessarily implementing the actual designs. It also gives us an opportunity to *run "what-if" scenarios* and observe the consequences of potential alternative actions. This can take away the risk and the cost of actual implementation and yet still give us a holistic view of the system under observation. Comparing different scenario outcomes help understand the impact of alternatives, make decisions, and take actions accordingly. Appropriate decision making can improve the overall organizational control, in healthcare and elsewhere.

Simulation is about knowing historical information, predicting future outcomes, and giving suggestions for system performance improvement. *Real-time simulation* is about consecutive simulation runs in *shorter periods* and about the prediction of *near-future* outcomes by using the historical information related to the behaviour of the components of the system and the *current (present) state* of the system. *"What if" scenarios* can execute against the simulation model and provide impact analysis. The quicker the simulation runs are, the more frequent the decision-making process can be performed, and the more timely actions are likely to be taken for system improvement.

There are however specific challenges due to the involvement of many human factors and exceptions in healthcare processes. Thus, simulating care processes might be more complex than simulations in other industries (e.g., transportation, finances and manufacturing). Moreover, the outcome results should be reliable enough to be analyzed for possible alternatives and decision making in a context where human lives are easily at stake. Therefore, the input data should also be sufficiently accurate (in terms of quality and timeliness) in order to have a reliable outcome results.

Note that this thesis focuses on simulations in support to day-to-day *operational* decision making (e.g., by bringing in more staff or opening/closing beds), without trying to optimize process *definitions*. This is different from most of the literature on simulations in healthcare, which focuses on *strategic* decision making (e.g., through improvements to processes, scheduling surgeries or restructuring of units). Carter and Blake have studied the use of simulations for strategic decision making as well as simulation challenges in that area [9]. They have also pioneered the use of simulations for patient flows in the emergency department of the Children Hospital for Eastern Ontario in 1996 [7], but to our knowledge they have not studied the use of real-time simulations in an operational context.

1.4. Research Methodology

For conducting this research, we use a methodology based on the *Design Science* approach proposed by Hevner and illustrated in Figure 1. Note in particular the relevance and rigor cycles in this approach. From the Design Science methodology we chose the following steps to follow for conducting this research:

1. Select a problem (business needs from the healthcare environment).
2. Review the literature on the problem (foundations and methodologies).
3. Propose a solution to the problem and define its specific requirements.
4. Design and development of the proposed solution (develop/build).
5. Analyze and evaluate the results (justify/evaluate).

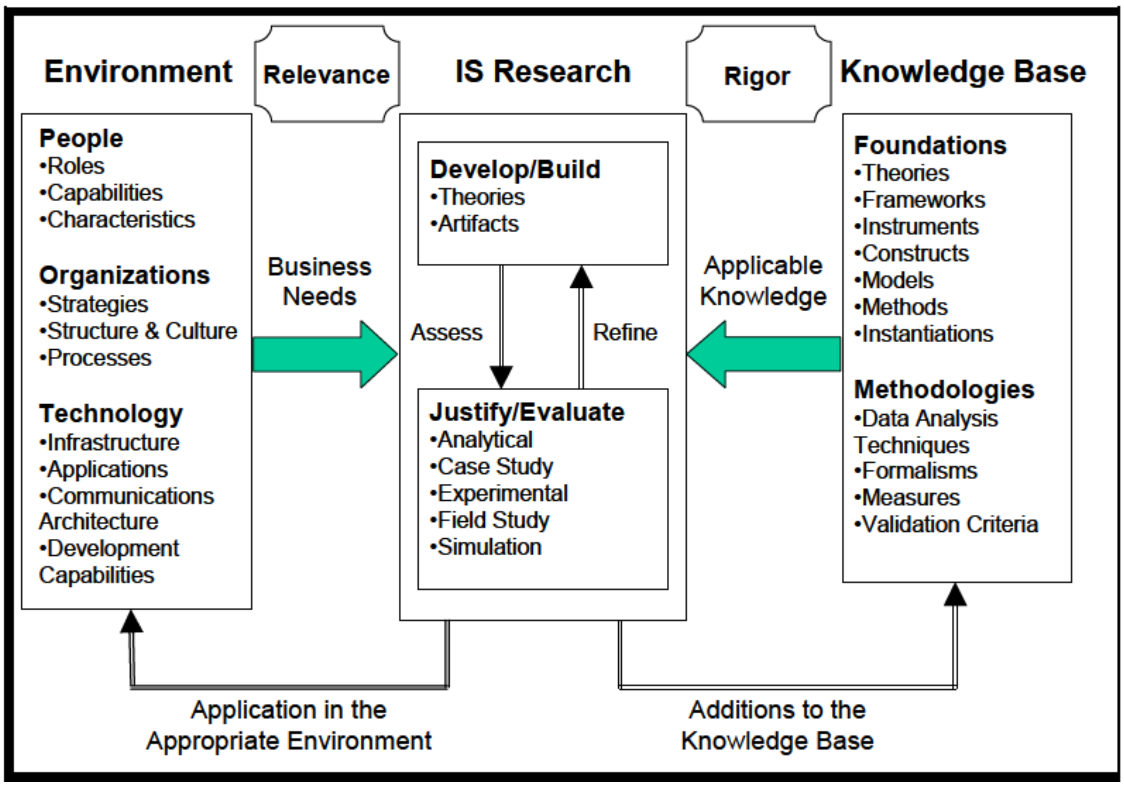

Figure 1 Design Science in information systems research (Hevner, [19])

This thesis research begins with a generic introduction to one of the most common issue recently faced by many hospitals, which is the reduction of patient waiting time. Earlier in this chapter, the motivations for selecting the waiting time problem were discussed (1) and the importance of operational decision making was highlighted for patient waiting time management.

A systematic literature review was conducted to get insight into the relevant work done so far regarding the most common healthcare issues (e.g., patient waiting time) and different approaches taken to solve the problems (2). The use of simulations in this context was also studied. Real-time simulations are suggested as an effective solution for operational decision making in order to help manage the waiting time problem, and specific requirements for designing a real-time simulation model are identified (3). The suggested model has been designed and developed for a real-life entire clinical pathway (4). The results from the application of the suggested approach are analyzed. Finally a conclusion is made based on the experiment results and the limitations on the way of conducting the research are shared along with suggested future work (5).

1.5. Thesis Contributions

The main contribution of this thesis is a demonstration that real-time simulations have the potential to support timely short-term operational decision making for patient care processes in hospitals. Scheduling of the resources is part of the operational decision making, which has a direct impact on the patient waiting time. Therefore, we are applying our real-time simulation approach for making short-term decisions that help reduce the patient waiting time by adjusting the resource scheduling.

There are also secondary contributions of this thesis, highlighted below:

- Methodology for developing real-time simulation models.
- Guidelines for mapping simplifying a clinical process expressed with the Business Process Model and Notation (BPMN) into a conceptual model, with additional mappings to a simulation model in Arena.
- Real-time simulation model of an entire clinical pathway, taken from a real-life patient flow.
- Real-time simulation capabilities based on a commercial simulation tool (Arena) that initially did not support real-time features.
- External access to the input data of the simulation model, which enables the possibility of integrating the model with a database that collects the real-time information on patient flow instances.
- Automated approach for choosing the number of replications for a simulation model based on the requirements on desirable precision of the output data.

- Introduction, set-up and evaluation of possible alternative scenarios that correspond to candidate short-term decisions that have the potential to impact the future state of the process.

The core ideas of this thesis are published in a paper presented at the 2013 Summer Computer Simulation Conference and Work in Progress (SCSC 2013 and WIP 2013), *Real-time Simulations to Support Operational Decision Making in Healthcare*, in order to share the knowledge gained through this research [2]. A poster based on this work was also presented by the thesis author at the Graduate poster competition of the Faculty of Engineering of the University of Ottawa in March 2013: *Real-time Simulations of Patient Care Processes: Support for Decision Making*. It received the second best poster award in the Computer Science category.

1.6. Thesis Outline

In Chapter 2, basic concepts in simulation and some statistical theories are discussed, while in Chapter 3 a literature review is conducted for the work related to this research. Chapter 4 focuses on the two main phases (and their related steps) of a methodology for defining a proper and accurate real-time simulation model. In Chapter 5 and Chapter 6, the two real-time simulation phases are discussed through an experiment done for a real clinical pathway. Different what-if scenarios are explored and their results are analyzed and discussed. In Chapter 7, a detailed comparison between the closely related work and our research is done, and threats to validity and limitations faced while conducting the research are shared. Chapter 8 concludes the thesis and highlights opportunities for future work.

Chapter 2. Background

In this chapter, basic simulation and statistical concepts required for this study are discussed.

2.1. Types of Simulation

There are different ways of categorizing simulation types depending on the purpose and application of a simulation. Kelton et al. [21] classify simulation types with regard to three different dimensions: time, state, and randomness.

2.1.1 Time: Static vs. Dynamic

In static models, the notion of time does not exist and the system is analyzed as a snapshot, whereas in dynamic models time plays an important role and the system's behaviour is observed over time. Usually, simulation models under observation are dynamic.

2.1.2 State: Discrete vs. Continuous

Discrete event simulation (DES) involves stating the system status in specific (discrete) time steps by showing a physical model through mathematical terms. In DES, the system is described as a sequential event-based model where the system changes to a new state after a specific time step, due to an event occurrence. Durations of steps correlate with event occurrence frequencies. It is assumed that there is no event occurrence between two consecutive time steps. DES is used mainly when precise analysis on detailed statistical values is required. This type of simulation has many applications in, for example, queuing systems and production planning.

When the dynamic behaviour of the whole system is the main focus of observation, a continuous simulation is used. In a continuous simulation, the variables change continuously over time and the continuous change of states is what needs to be observed rather than at what time specific states occur.

2.1.3 Randomness: Deterministic vs. Stochastic

Deterministic models consist of specific inputs and outputs for which the data is not random. In these models, the relationship among different parameters will be precisely detected and shown through mathematical terms, thus allowing no randomly generated numbers. The same specific output is expected from given known inputs (e.g., chemical reactions of known elements).

Stochastic models, on the other hand, contain at least one random input. These models make use of random number generation. The components of the systems are observed individually and their behaviour is expressed through statistical distributions. These distributions will be generating random numbers but within a certain range, which corresponds to the range of possible behaviours of the system components.

The problem defined *for this research* is the issue of patient waiting time. The system under observation is a clinical center that provides clinical services to randomly arriving patients (**stochastic** model). We want to observe the behaviour of that enter over time (**dynamic** model), while analyzing the system as a queuing system (**discrete-event** model).

2.2. Different Types of System Behaviour

Depending on the nature of their behaviour, there are generally two types of systems:

- Terminating systems
- Non-terminating systems

A terminating system starts in an "empty and idle" state and its behaviour will continue evolving until the system shuts down. An "empty and idle" state refers to a specific state of the system in which there are no entities available and all of the resources are idle and not working. The system works during the working hours defined for it and after a specific period of time it terminates and shuts down. An example of this type of systems is customer services offered in banks, which normally start from 9am, finish at 5pm, and restart from an empty and idle state the next day.

On the other hand, there are non-terminating systems that do not shut down after a specific period of time and are always in a working mode. Therefore, these systems are almost never “empty and idle” and the entities can get served at any time. Examples of this type of systems include hospital emergency departments (as in this thesis) and gas stations.

2.3. What is a Warm-Up Period?

A warm-up period is a commonly used concept in simulation studies. In a simulation model, normally when the simulation runs, the system starts from an “empty and idle” state, which is the case if the type of system under observation is terminating.

However, if the system under observation is a non-terminating system, the concept of warm-up periods arises. A *warm-up period* is the transient state that happens at the beginning of the simulation run. From the time when the system starts with an “empty and idle” state until the time the system is actually in a working mode, the simulation does *not* reflect the actual system behaviour. Since non-terminating systems never start from an empty and idle state, the information collected in a simulation model during the warm-up period is not reliable. Typically, for studying these types of systems, we will wait until the warm-up period is terminated and we start observing the behaviour of the system when it reaches a steady state. A *steady state* is believed to be a good reflector of the real-life system assuming that the variation of the parameters that are affecting the system is going to remain the same. There are several methods to figure out the duration of the warm-up period and avoid collecting data during that period, all of which start the data collection from an unknown arbitrary initial state *after* the warm-up period is over.

2.4. What is a Real-Time Simulation?

Using real *current* data as an *input* for a simulation model, which enables near-future predictions of the system behaviour in a timely fashion, is called *real-time simulation*. Basically, the core idea is to use existing information and ongoing events as an input to the system to figure out the consequences of certain actions in the near future and iteratively feeding the present data to find out about upcoming impacts. Alternative scenarios targeting different types of resource allocation combinations can also be run. Subsequently, through an optimization process, the best alternative will be chosen with respect to the results, and decisions will be made. The eventual intention is to define alternative actions (mainly about adding, moving, or de-allocating resources), if necessary, that could support us in making effective short-term operational decisions. Real-time simulations can be shown as cycles in which repetitive data feeding, data processing and output analysis take place. Figure 2 shows the cycle of a real-time simulation for a system.

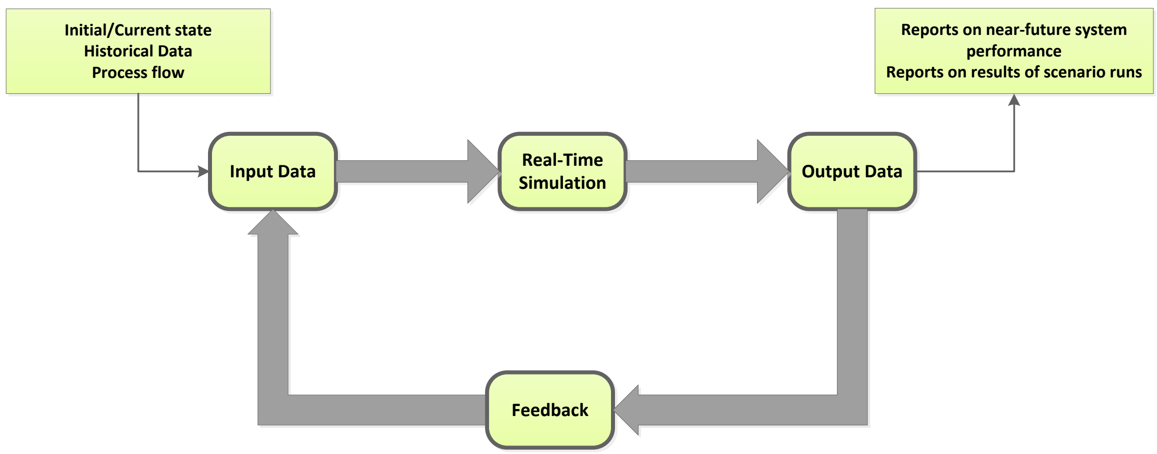

Figure 2 Real-Time Simulation Cycle

2.5. Real-Time Simulation Software Applications

There are several software applications that exhibit interesting features for the support of real-time simulations. Three leading commercial solutions are discussed here.

2.5.1 Arena

Arena is a commercial simulation tool supporting discrete event simulation[1]. It was initially developed by Systems Modelling, which was acquired by Rockwell Automation in 2000. In recent years, it has become very popular for its broad applicability in different domains, including manufacturing, packaging, supply chains, healthcare, and military applications. Arena relies on the SIMAN language for modeling. Arena is not only a simulation software package; it supports optimization aspects as well. In Arena, models are defined and constructed using different blocks. Then, through optimization, strategies are

[1] http://www.arenasimulation.com/

set and the effectiveness of each strategy is observed prior to selecting one for implementation.

Modules are used for building models in Arena. They are basically some predefined shapes that represent specific tasks or states. Modules are being lined up and connected to each other by *connectors*, which indicate the possible flows of the entities through the model. Figure 3 shows the Arena interface, including a patient care process model (not meant to be read here, but whose details can be found in Appendix A).

For importing and exporting file formats, Arena is quite flexible. It can export files in several formats like PDF, Word, Excel and HTML. AutoCAD files and Microsoft Visio files can also be imported in Arena for a 3D visual interface for simulation models.

Input data in Arena can be injected in two different ways. There is an external software in Arena called *Input Analyser*, through which importing *pure data* (i.e., raw data) and fitting it into different distributions is feasible. A variety of statistical distributions such as uniform, exponential, triangular, normal, Erlang, beta, and others exist in the Input Analyser. Given the statistical information related to each distribution, the user decides which one fits the data the best. The second approach is to directly feed the pure data into the processes with importable files, without fitting them to a specific distribution. Both of the approaches have their own pros and cons. Using actual data will improve the accuracy of real-time simulations, since it will be the exact data that we will be dealing with. However, using specific distributions for simulation runs of the model will be less time consuming than when every single data element needs to be read and processed from the imported file in each replication.

After each simulation run, Arena creates a file that can specifically be used in the *Process Analyzer* (PAN), for observing how a specific change in model values is going to affect the outcome results. The user defines a range of different what-if scenarios by changing different parameters (without directly making changes in the actual model) and runs each scenario to analyze possible outcomes.

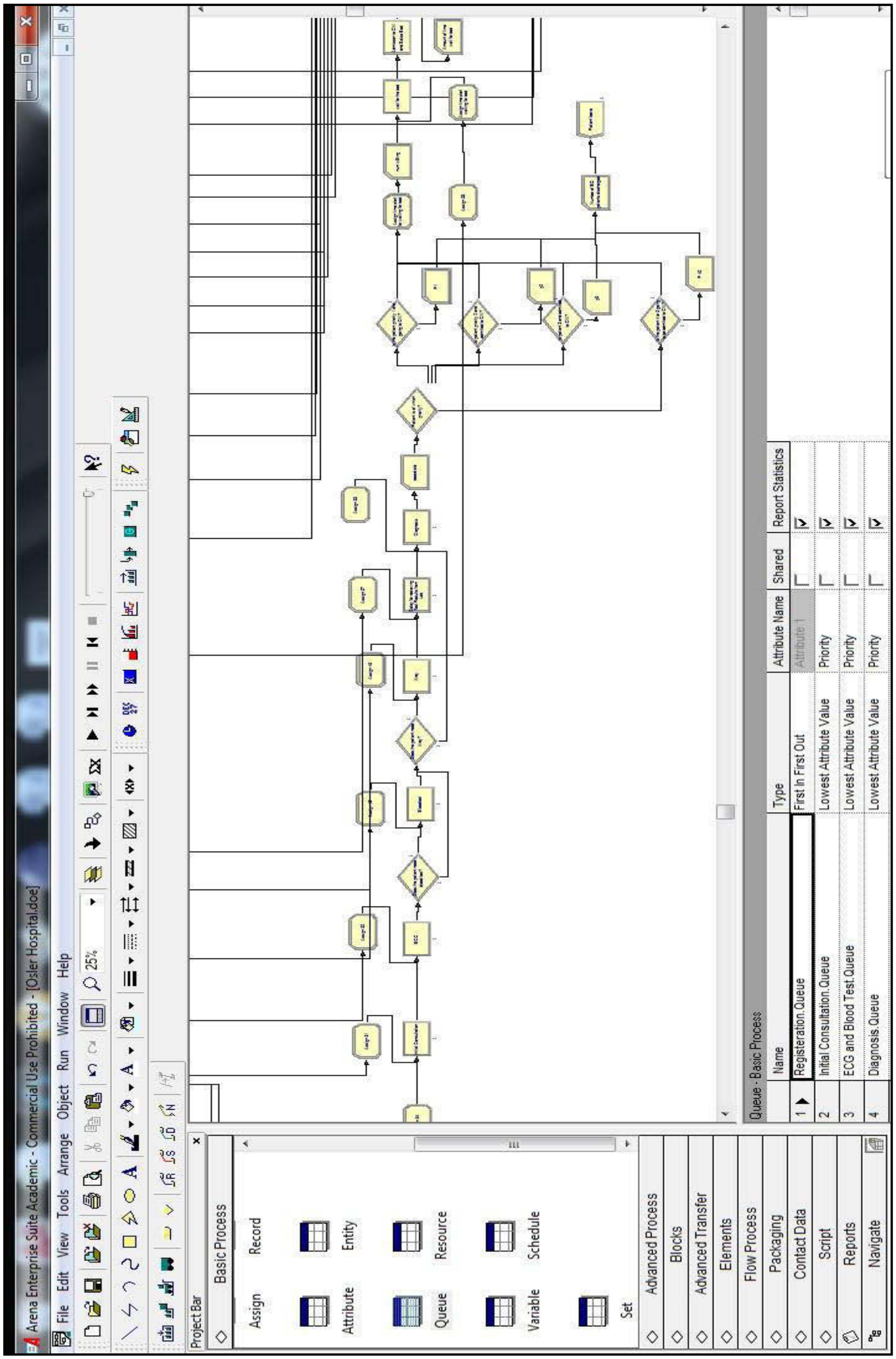

Figure 3 Arena Interface

Arena provides the user with optimization capabilities as well as simulation. There is an external package, *OptQuest*, through which a set of controls and responses are chosen for optimization. Controls are the parameters for which the values can change within a range, preventing the optimizer to use values that are out of the range (and hence reducing the state space to be explored). Controllers are used for defining linear and/or nonlinear constraints. Responses are the different types of outputs that the user is interested in minimizing or maximizing during a number of simulation runs. The *objective function* is made up of one or several responses. OptQuest uses meta-heuristic methods for optimizations. It automatically runs a number of possible scenarios to satisfy all the constraints, evaluates the objective function value with each iteration, and compares it with the previous results to determine a new set of combinations for the next iteration. If any constraint remains unsatisfied, then the provided solutions are shown as infeasible. A terminating condition should be set in the optimizer for avoiding unnecessary and time-consuming simulation runs. Finally, the alternatives are shown to the user, ranked from the best case scenario to the worst case scenario.

Arena also supports Visual BASIC for Applications (VBA) scripting for fulfilling specific needs or requirements.

2.5.2 IBM Business Process Manager

IBM Business Process Manager (BPM), formerly known as Lombardi, is a business process management software that supports real-time simulation of processes, following a discrete-event simulation approach[2]. The business process can be defined with BPMN and the historical data can be set in the model using a limited number of statistical distributions (only two). For optimization and multiple scenario runs, this tool supports profile management, which gives the user the opportunity to make changes in the model without configuring the actual base model and check for the possible outcomes while applying some configurations. A number of predefined visual reports are generated after each simulation run that can be analyzed.

Although having simulations enabled at the BPMN level is very attractive, IBM BPM is rather limited compared to Arena in our context:

[2] http://www-03.ibm.com/software/products/us/en/business-process-manager-family

- Model validation and error management are rather poor.
- The pace of simulation runs is not observable (no clock) nor adjustable (the user cannot slow down or fast forward runs).
- Report generation is fairly basic, and only one format (Excel) is available for exporting data, which limits us when feeding results to a decision support system.
- Animation is limited and insufficient for observing whether states are initialized properly.

2.5.3 AnyLogic

Anylogic is developed by XJ Technologies[3]. Three core simulation methodologies (discrete-event simulations, system dynamics and agent-based modeling) are supported. Complicated models can be built in this tool using any of the three methodologies, or a combination of two or more, whichever seems more appropriate to the context. The historical data can be set using a wide variety of statistical distributions. The simulation model can be extended through Java coding. The tool supports visual animation and real-time simulation features.

AnyLogic is almost equally good as Arena and is even better for some specific features. However, it does not have an automatic verification of a 95% confidence interval for results (something useful for real-time simulations that need to minimize their number of runs). Also, as the academic version of Arena was more easily accessible for evaluation (before buying the tool), and since the partner hospital also had plans for evaluating this tool, we decided to investigate Arena in the case study of this thesis.

2.6. Simulation Output Precision and Confidence

Stochastic models depend on random number generation. Even though the generation of numbers might follow a specific statistical distribution, these numbers are still random and as a rule, random numbers *in* will cause random numbers *out*. Therefore, the statistical behaviour of the output numbers should be observed and interpreted properly in order

---

[3] http://www.anylogic.com/

to have a correct understanding of the predicted results. In this section, several statistical concepts and theories are discussed as they have been taken into consideration for our simulation output calculations.

2.6.1 Sampling Process

Sampling is the estimation of the characteristics of a large community by selecting and observing the behaviour of a subset of that community. The larger the sampling population is, the more precise the estimations would be but the more costly the experiment would become. After the process of sampling is completed, the results are analyzed and the estimated points generated. The *sample mean* is calculated through the following formula[4]:

$$\bar{X} = \frac{\sum_{i=1}^{i=n} X_i}{n}$$

where X_i is the value observed in the i^{th} trial and n the number of experiments.

2.6.2 Central Limit Theorem

The *central limit* theorem is one of the fundamental theories is statistics. According to this theorem, if we average a particular measurable quantity through a sampling process for a number of times, then the behaviour of the averages we calculate is going to tend toward a normal distribution as the sample size and sampling trials increase[5]. This theory applies to all the random means or sums over independent and identically distributed variables, regardless of their original distribution. Therefore, in the calculation of the estimated point and the confidence interval, the t distribution is used (as we typically use few simulation runs) and appears as a constant number that relies on the number of sampling trials and the confidence level. The values of the t distribution will be close to the values of a normal distribution from 30 trials onwards. If the sampling process is taking less than 30 trials, then the t distribution should be used for calculating the related statistics.

---

[4] http://en.wikipedia.org/wiki/Sample_mean_and_sample_covariance
[5] http://en.wikipedia.org/wiki/Central_limit_theorem

2.6.3 Sampling Error Measurements

For evaluating the precision of the sampling data generated within a sampling process, there are several statistical sampling error measurements that are often used.

Standard Deviation

The standard deviation corresponds to the variation that exists around the average[6]. In other words, this indicates how much the calculated data is spread. The lower the standard deviation, the more precise the results would be considering that the data from which the mean is calculated is closer to the actual population mean. The formula for calculating the standard deviation is:

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{i=n}(X_i - \bar{X})^2}$$

where X_i is the value observed in the i^{th} trial, n is the number of experiments, and $\bar{X}$ is the sample mean.

Confidence Interval

Confidence intervals show both the precision as well as the location of measured parameters and consist of an *upper bound* and a *lower bound*[7]. This interval is a range through which the estimation point falls by the probability equal to the confidence level in further repeated trials. The common confidence levels chosen for statistical analysis are 90%, 95% and 99%. If for instance the confidence level is chosen as 95%, this means that in 95% of the different experimentations for sampling, the actual parameter of interest will exist within that confidence interval reported. The formula for calculating the confidence interval for upper and lower bounds is:

$$\bar{x} \pm t_{\alpha/2,n-1} \frac{\sigma}{\sqrt{n}}$$

where:

- $\bar{x}$ is the sample mean
- σ is the standard deviation of the sample

[6] http://en.wikipedia.org/wiki/Standard_deviation
[7] http://www.allianthawk.org/victionary/showdef.php?word=48

- n is the size of the sample
- $t_{\alpha/2,n-1}$ is a value derived from the *t* table that relies on the sample size as well as the parameter of significance (α). If the confidence level is 95%, then the value of α would be equal to 1 – 0.95 = 0.05.

Confidence intervals can be calculated for a small number of sampling trials (even for *n*=2). However, the smaller the number of the sampling trials, the wider and the less precise the confidence interval will be.

Therefore, only ensuring that the calculated data is within the confidence interval range with a certain probability is not sufficient for concluding that the precision of the final results is guaranteed. Both the interval and the variation from the mean are going to specify the precision of the sampling output data.

Margin of Error

The margin of error (MOE), or *half width*, is a statistical expression that acts as an indicator of sampling error[8]. This value is known as the radius of the confidence interval. In a sampling process, as discussed previously, a certain number of individuals from a large community are chosen for their behaviour to be observed and then the results are generalized to all the population with a certain confidence and precision. When a sampling process is occurring, there are some individuals left that remain unobserved. The lower their number, the tighter the confidence interval will become, assuming that the sampling trials remain the same. Since in a sampling process we are usually unable to avoid a non-null margin of error, the objective is to minimize this error in order to get a desirable precision in the output data. The formula calculating the margin of error is:

$$MOE = t_{\alpha/2,n-1}\frac{\sigma}{\sqrt{n}}$$

where:

- σ is the standard deviation of the sample
- n is the size of the sample

[8] http://www.allianthawk.org/victionary/showdef.php?word=48

- $t_{\alpha/2,n-1}$ is a value derived from t table that relies on the sample size as well as the parameter of significance (∝).

The MOE is sometimes calculated as an absolute number mostly if the sampling process is done only one time. However, sometimes this value is being described as a relative quantity in the context of the sample mean. Such relative MOE is:

$$Relative\ MOE = \frac{MOE}{\bar{X}}$$

Where $\bar{x}$ is the sample mean and MOE is the margin of error.

Coefficient of Variation

The coefficient of variation is the ratio of the standard deviation of a sampling output and the mean of that same sampling output[9]. This value is mostly being presented as a percentage, which in general would describe the variability of the observed sampling individuals in relation to the mean of the sample. The formula is:

$$\hat{c}_v = \frac{\sigma}{\bar{x}}$$

Where $\bar{x}$ is the sample mean and σ is the standard deviation of the sample.

2.6.4 Stochastic Simulation, a Sampling Process

Stochastic simulation involves a process of sampling within every setup. Through the period of the "run length" in each replication, the samples are collected. Simulation replication is the sampling trials that occur in a sampling process. For each trial, the average will be kept, and the average of all of the sample trials is calculated at the end of the simulation replications. The higher the number of replications, the more precise the results will be. However, this precision comes at the cost of spending more time to run the simulation model.

[9] http://en.wikipedia.org/wiki/Coefficient_of_variation

2.7. Summary

In this chapter, we have explained different simulation concepts that will be used in this research. Also, a brief introduction to some leading applications that support real-time simulation was given. In our study, because of the random nature of the events that occur in hospitals, we have decided to use a stochastic model for simulation purposes. Also since the behaviour of the hospital over time is of concern, the simulation model will be dynamic. As we are observing the care processes in a hospital as a queuing system, we chose discrete-event simulation for analysis. Eventually we chose Arena as a tool for creating our real-time simulation model.

Stochastic models are deeply integrated with statistical notions. The reason is because these models should be able to predict and evaluate the future state of the system through measurable criteria. The data generated through simulation runs is collected in a sampling process. This data should be accurate and precise enough to be relied on. We have introduced the statistical terms and theories that we will need later for making decisions on the precision of the data that the simulation model generates.

The next chapter will discuss the relevant literature in this area. We will explain how simulation and/or real-time simulation applications have been used so far as techniques for problem solving in healthcare.

Chapter 3. Literature Review

In order to understand existing work on simulations of patient care processes, a literature review has been conducted. In this chapter, conference papers and journals related to this subject have been found through a systematic approach (inspired from Kitchenham's approach for systematic literature reviews [22]) and were analyzed to extract the required data and identify strengths, limitations, and remaining challenges.

3.1. Selection Methodology

For the credibility of the papers used in the review, three trusted search engines were used:

1. SpringerLink[10]
2. IEEE Xplore[11]
3. ACM Digital Library (DL)[12]

For all three search engines, the Advanced Search section was used to clarify and specify the exact combinations of keywords that were of concern.

Since the concept of real-time simulation has a broad scope in many domains, the concentration is on the related work, where "healthcare" is the actual context. Furthermore, in the healthcare domain, there is a broad range of real-time simulation applications, for several medical objectives. For instance, in the process of tracking patient health statuses, some clinical devices are used that simulate these statuses in real time. Moreover, real-time simulations are also applied for training purpose.

The focus of this work is on the real-time simulation of the process flows involved in healthcare, which would lead to effective decision making. Consequently, the

[10] http://link.springer.com/
[11] http://ieeexplore.ieee.org/Xplore/
[12] http://dl.acm.org/dl.cfm

following keywords were chosen after several trials (previous attempts led to too many irrelevant papers or to too few papers):

- In SpringerLink: "real time simulation" AND "modeling" AND "decision" AND "wait time" AND "healthcare"
- In IEEE Xplore: "real time simulation" AND "business process" AND "wait time" AND "healthcare" AND "Patient work flow"
- In ACM's DL: "real time simulation" AND ("resource" OR "staff") AND "healthcare" AND "model" AND "process" AND "patient flow"

Note that some search engines like SpringerLink tolerate variations in the results (e.g., wait times and waiting times, real time and real-time, modeling and modeling, healthcare and health care, etc.), while other required variations and synonyms to be checked in extra queries.

3.2. Filtering Methodology

After collecting the pool of candidate papers, two filtering processes were applied in order to be more specific on selecting the most relevant papers. The first approach taken for filtering the results was to use the existing filtering options in SpringerLink and IEEE Xplore. In ACM DL, this capability does not exist. The filtering criteria were set to choose papers only if they were from journals or conference papers that were in English. The papers duplicated in search results of two or more search engines were kept only in the results of the first engine used. Also the result sorting was set to be according to relevance of the paper to the searching keywords. After the first filtering, the second filtering was done manually. The titles of the articles were inspected and the abstract and conclusion of the selected ones were skimmed through. For some papers, the content of the paper also was skimmed for making decisions on the final selection of the papers.

The content of the abstract and of the conclusion should have been as follows for a paper to be selected:

- It should be stating an issue related to patient care processes involved in any healthcare institution that offers clinical support.

- The stated issue should fixed using simulations or real-time simulations.
- It should not be related to simulations or real-time simulation for direct medical purposes, such as for medical devices that simulate the health conditions of the patients or the ones used for training purpose.

Finally any other article that was found (through other means) to be related to real-time simulation, without necessarily being applied in healthcare domain, was categorized as "other". Table 2 shows the final numbers of selected papers:

Table 2 Paper Selection

	SpringerLink	IEEE Xplore	ACM DL	Other	Total
# of papers found	124	154	114	-	392
# of papers after first filtering	65	150	114	-	329
# of papers finally selected	7	19	9	1	36

3.3. Overview

Table 3 summarizes the content of the papers according to different fields. Some atypical values are highlighted along the way (green for good, yellow for partial, red for bad).

- Source: specifies the search engine used to find the article.
- Country: indicates the country in which the hospital under observation is located.
- Scope: shows the scope of the study, including healthcare in general, a hospital, or a specific unit such as ED, an Operation Room (OR), or the Intensive Care Unit (ICU).
- Simulation type: indicates the type of the simulation (e.g., discrete event simulation, or based on queuing networks/theory)
- Real-time implementation: shows whether or not the article indicates that an actual real-time simulation has been implemented.
- Effective in decision making (DM): indicates whether the article agrees on the effectiveness of simulations or real-time simulations in decision making processes.

- Wait time issue: indicates whether the article is specifying a “wait time” issue within the scope of the research.
- Actual hospital: flags whether there was a real, actual hospital process used in the research.
- Simulation tool: shows the name of the tool used for doing simulation modeling and analysis.
- Data collection: four types of input data collection in modeling are indicated:
  - 0: is chosen when the input data is totally hypothetical.
  - 1: is chosen when the input data has been collected either manually or by doing observations and interviews.
  - 2: is chosen when the input data was collected both manually (through observation or interviews etc.) and also from an electronic database.
  - 3: is chosen when the input data collection was from electronic databases.
  - None: is chosen when the research did not need to have any data collection.
  - N/A: is chosen if the method of data collection was not mentioned.

  N.B.: In this table, the articles that mentioned they used “historical data” were also categorized as 1.
- Model V & V: illustrates if the simulation model has been validated and/or verified or not. One of 3 values are possible.
  - Yes: either verified or validated.
  - No: not verified, neither validated.
  - Partially: the article indicates that the model is validated, however it has some major inaccuracies that affect the outcome results.

Table 3 Summary of Related Work

Article	Source	Country	Scope	Type of simulation	Real-time implementation	Effective in DM	Patient wait time issue	Actual hospital	Simulation tool	Data collection	Model V & V
[4]	IEEE	Germany	OR	N/A	No	Yes	Yes	Yes	Arena	1	Partially
[5]	IEEE	Portugal	OR	DES	No	No	Yes	Yes	Arena	1	No
[1]	IEEE	-	ER	DES	No	Yes	Yes	No	Arena	None	N/A
[17]	IEEE	UK	ED	DES	No	No	Yes	Yes	DGHPSim	3	N/A
[46]	IEEE	Japan	OD	DES	No	Yes	Yes	Yes	Arena	2	Yes
[28]	IEEE	US	ED	N/A	No	No	Yes	Yes	Arena	1	Partially
[24]	IEEE	US	ED	DES	No	Yes	Yes	Yes	Arena	1	Yes
[16]	IEEE	US	ED	DES	No	Yes	Yes	Yes	Arena	3	Yes
[27]	IEEE	UK	Theatre	DES	No	Yes	Yes	Yes	Arena	2	Yes
[15]	IEEE	Sweden	ED and RD	N/A	No	Yes	Yes	Yes	N/A	2	N/A
[42]	IEEE	UK	OD	DES	No	Yes	Yes	Yes	Arena	0	N/A
[42]	IEEE	Taiwan	ED	DES	No	No	Yes	Yes	simul8	N/A	Yes
[36]	IEEE	Brazil	Surgery Room	N/A	No	Yes	No	Yes	ProModel	3	Yes
[37]	IEEE	-	ED	Agent based	No	Yes	Yes	No	Jsim	None	No
[47]	IEEE	Hong Kong	AED	DES	No	Yes	Yes	Yes	Arena	3	Yes
[12]	IEEE	Iran	ED	N/A	No	Yes	Yes	Yes	Arena	1	Yes
[14]	IEEE	Belgium	Hospital	DES	No	Yes	Yes	Yes	Arena	1	N/A
[45]	IEEE	Japan	OD	DES	No	Yes	Yes	Yes	Arena	2	Yes
[29]	**IEEE**	**Israel**	**ED**	**N/A**	**Yes**	**Yes**	**Yes**	**Yes**	**Arena**	**3**	**Yes**
[26]	SpringerLink	USA	ICU	DES	No	Yes	Yes	Yes	Process-Model	3	Yes
[44]	SpringerLink	Ireland	Hospital	N/A	No	Yes	Yes	Yes	simul8	3	N/A
[23]	SpringerLink	UK	Hospital	N/A	No	Yes	Yes	Yes	Arena	3	N/A
[38]	SpringerLink	Canada	OD	DES	No	Yes	Yes	Yes	Arena	1	Yes
[34]	SpringerLink	Sweden	OR	DES	No	Yes	Yes	Yes	N/A	1	Yes
[11]	SpringerLink	USA	Hospital	N/A	No	Yes	Yes	Yes	Arena	3	Yes
[13]	SpringerLink	Australia	Hospital	N/A	No	Yes	Yes	Yes	Extend	0	N/A
[41]	ACM	USA	Pharmacy	N/A	No	Yes	No	No	AutoMod	1	No
[25]	ACM	N/A	ED-IU	DES	No	Yes	Yes	Yes	N/A	N/A	N/A
[18]	ACM	UK	Hospital	DES	No	No	Yes	Yes	Micro Saint sharp	3	Yes
[20]	ACM	Norway	ED	DES	No	Yes	Yes	Yes	Flexsim	1	Yes
[3]	ACM	N/A	Healthcare	N/A	No	Yes	Yes	No	N/A	None	N/A
[35]	ACM	USA	ED	DES	No	No	Yes	No	N/A	None	N/A
[48]	ACM	N/A	Healthcare	All types	No	No	No	No	No	None	No
[40]	ACM	USA	Hospital	N/A	No	Yes	No	Yes	Arena	1	Yes
[10]	ACM	N/A	ER	DES	No	Yes	Yes	No	Arena	None	N/A

As can be seen from Table 3, research in this domain is conducted in many hospital locations around the world. Looking at the publication dates, simulation in healthcare has been attracting attention mostly during recent years. Nearly two thirds of the studies are using discrete event simulations for their own simulation scopes since they believed this approach reflects the real situation in healthcare domain better. 80% of the articles believed that simulations can have effective impacts on decision making. 91% of the articles were concerned with long waiting times for patients. 55% of the articles used Arena as their simulation tool, whereas 5% used Simul8[13] (the second most popular tool, well behind Arena). Among all of the simulation studies that used observations, interviews, and manual methods for collecting data, 40% were not able to verify or validate their model accurately. All of the simulation models that used electronic databases or electronic records of patients for input data, and that have mentioned their verification and validation approach, had their models accurately validated and/or verified.

3.4. Comparison and Analysis

Using simulations gives us the opportunity to identify problems and suggest approaches that can provide an effective solution without actually implementing alternatives in the real world. In the healthcare domain, there may be several reasons why simulations are applicable. It may be the long waiting time for patients that needs to be minimized, strategic decision making regarding different operational models (e.g., adding or removing units from the hospital, or layout design changes), or better scheduling for adjusting resource utilization.

Different applications of simulation or real-time simulation are being discussed in the following subsections. For a more organized analysis on the papers' content, several categories are defined according to the content relevance.

- Real-time simulations in healthcare: articles in which real-time simulations have been introduced as a solution to the identified problem in healthcare.
- Simulation modeling in healthcare: articles in which a simulation model for healthcare processes has been constructed and analyzed. Almost all of the ap-

[13] http://www.simul8.com/

proaches followed the same procedure for realizing simulation capabilities, and these steps are used as subsections:

1. Conceptual and simulation modeling.
2. Data collection and data input.
3. Simulation setup and run.
4. Model validation and verification.
5. Reports and outcome analysis.

- Challenges, limitations and restrictions in simulations for healthcare: Papers that specify the challenges and limitations in using simulations within the healthcare domain.
- Complementary papers: articles that are very relevant to real-time simulations but not necessarily applied in healthcare.

3.4.1 Real-time Simulations in Healthcare

Marmor et al. [29] mention that ED congestion is one of the important issues that most of the hospitals are facing nowadays, which causes a great increase in patient waiting time, poor service quality, patients in critical pain status, ambulance diversion, and patients leaving the hospital without seeing a doctor. The two main intentions for conducting this research were to collect real patient arrival data and to be able to predict the hospital performance in the near future. The authors injected the arrival data for a long enough period of time in the simulation model, so that they could ensure the model passed the warm-up period. The process of care was modeled in an ED with six types of incoming patients. Arena and Anylogic were used for the purpose of real-time simulation. Patient arrival times were defined as a stochastic process, however, the data related to patient discharge time did not seem to be accurate enough. The main reason was because some of the patients who were discharged from the ED had to continue their care process in the interrelated departments of the hospital and thus they ended up waiting in the ED to finally get admitted to the wards. This caused a huge congestion in the ED, which the simulation model could not reflect, since as soon as the discharge process in the simulation model was done, the patients were assumed to have left the ED.

For a proper resource scheduling in a short-term online simulation suggested by Marmor et al. [29], two mathematical approaches were considered: Rough Cut Capacity Planning and Offered-Load. These two allocation approaches have been compared in terms of their impact on resource workloads and patient satisfaction levels. InEDvance is a decision support system that simulates and predicts the near-future events of the ED and helps the manager in making decisions by presenting the related data with a graphical interface. The model was checked for validity but it is mentioned that the real-time simulation later should be implemented in the hospital for proof of effectiveness of the approach.

3.4.2 Simulation Modeling in Healthcare

Wijewickrama and Takakuwa motivated their research by pointing out the great population of elderly people in Japan [46]. *Aging issues* have created new problems in terms of overcrowding in the healthcare domain, which is a cause of long waiting times for patients. There is actually a famous saying in Japanese that states: "waiting for three hours to be seen for three minutes" [45].

Kolb et al. [25] describe "overcrowding" as a cause of *unbalanced supplies and demands*. This can either be caused by a large load of patient arrivals or by a limited number of employees in a hospital or clinic for service delivery.

White [35] mentions that congestion can result in long waiting times as well as *ambulance diversion*. An ambulance diversion occurs when a patient with a critical health status is brought by ambulance to a hospital whose ED is already full. The patient hence cannot be admitted and the ambulance should take the patient to another hospital. Kolb et al. [24] emphasize that the issue becomes more severe when the ambulance is diverted multiple times from different EDs.

Gopankumar et al. [16] introduced simulation as a tool for supporting *strategic decision making* and believed simulation is one of the best ways to observe real system behaviour when real-life experimentations are infeasible or too costly. Decision making in hospitals may not be as easy as in other organizations. The unpredictable arrivals of patients and limited capacity of the resources available make the decision making process complicated. Also, the fact that making inappropriate decisions in hospitals may result in life-threatening disasters makes the process even more complex. Baumgart et al. [4] stipulate that fast adaptation to changes that occur in a hospital is vital. Therefore, there is a need to get proper insight into an overall system performance. Gopakumar et al. [16] mention that modeling and simulation is one solution. The intent of a simulation is aligned with the required improvements in change management of the processes. As mentioned in the introduction, Blake et al. have also used simulation for strategic decision making in the context of ED patient flows [7].

Conceptual and Simulation Modeling

There are several approaches available for building simulation models. Most of them exploit *discrete event simulations* to capture the scope of the problem, as in the study conducted by Wong et al. [47]. Raunak et al. [37] used Little-JIL as a simulation tool. Unlike Arena, which supports analysis based on queuing theory, Little-JIL is based on process and resource definitions. Processes are defined in a hierarchy of steps where all types of resources are defined. There is one resource type that differs from the current resource types in processes and can be considered as an external resource, namely an "Agent". Resource instance characterization is a responsibility of the Agent. JSim is a simulation tool that uses a TimeLine for producing events. The initial state is set to zero in the TimeLine. Then, the step that should be executed at the start will be initialized in Little-JIL. When the simulation runs, each step sends event(s) in the TimeLine. The TimeLine preserves the sequence of the events, thus there is a continuous event occurrence during each simulation run.

Protil et al. [36] and Rohlder et al. [38] state that for building simulation models of patient flows, several meetings were held and the processes were monitored. However, Macdonald et al. [28], Fryk et al. [15] and Spry et al. [41] used careful observation as the only method used for defining patients process flows.

Steps that should be followed for building a simulation model can also be expressed by means of a *conceptual model*. Kolb et al. [24] and Wong et al. [47] proposed a conceptual model, which specifies the process steps. Kolb et al. [24] proposed a generic solution for all ED congestion problems and their target hospital was observed for a period of time in order enable the construction of a conceptual model. Williams and Lei [44] and Koizumi et al. [23] used a mathematical conceptual model for resource utilization evaluation.

Bayngart et al. [1] started with creating a high-level model that only reflected the major activities involved in the target processes. After a thorough analysis of the tasks, the model was built covering more details of the whole processes. For being more specific on modeling hospital performances, Komashie et al. [27] divided the hospital into sections related to the speciality of the “theatres” (units) and each section has been considered individually in terms of its performance. Also, in the studies conducted by Komashie et al. [27], Wong et al. [47], and William and Lei [44], patients are modeled according to their health situation and priority. However, in some of the studies like the one done by Bernardo et al. [5], a number of activities or entities were excluded for simplifying the model [5]. For instance, Fei et al. [14] did not include the emergency patients who came to the hospital for having urgent surgeries in the model, since the model was defined only for scheduled and planned patients’ entrance.

Raunak et al. [37] offer an agent-based simulation model at an architectural level. Modeling was done using Little-JIL and by doing some configurations with JSim.

There are different purposes for using modeling and simulation. Spry and Lawley [41] have observed a pharmacy performance for their study. The normal processes involved are: taking orders of medication, checking for conflicts with other medications that the patient is using, and delivering the medication. Apart from this routine, at some specific times of a day, the staff will be busy answering phone calls regarding requests for medication. The authors are using simulation for scheduling the resources. Sepulveda and Cahoon [40] proposed a model that evaluates the impact of different kinds of hospital layout on the overall performance.

Data Collection and Data Input

The input data should be sufficient in quantity and accuracy for creating a concrete simulation model that provides sufficiently precise output. The more the input data is accurate, the more reliable the final results will be.

In the study done by Eskandari et al. [12], data collection was done manually for 8 weeks, 24 hours a day, 7 days a week. The data was related to 6500 patients. 70% of the data was used for simulation purposes and 30% was kept for validating the model.

Rohlder et al. [38] initially used self-writing methods for collecting patient information. In this method, the patients were asked to write down their timing information on a piece of paper on their own as they got involved in each of the activities. This resulted in *imprecise data collection*. Afterwards, an observation method was chosen for patient data [38]. Also, Fei et al. [14] collected 2230 records of patients in a period of one year by using an observation method. In the data collection information shared by Kolb et al. [24], it was mentioned that the data was related to 8525 patients of the hospital and was collected in three consecutive months. However, Baumgart et al. [4][1] have announced that although there was a shortcoming in their data collection (only 92 records manually collected and approximated for some activities), they were still able to figure out some valuable information through simulation. Yu et al. [49] did the data selection in an arbitrary way.

In the studies done by Komashie et al. [27] and Wijewickrama and Takakuwa [45][46], data collection was through observing and interviewing the administrators and clinical people as well as through electronically records of patients. The input data was partially taken from an existing database that stored patients' tracking records. The rest of the data was collected manually. Macdonald et al. [46][28] had their data collection done manually for all the study.

Fryk and Steins [10] discussed the benefits of using *information systems* for storing patient information [15]. The main intention was to use the existing data in information systems as input data for simulation. However, the data collection was not accurate enough, since the hospital was using two separate information systems for recording patients information, and also since the information systems were designed mostly for billing purposes rather than process flow data. Marmor et al. [29] collected most of the data using information systems. For instance, for collecting the arrival data, they used a single hour of one specific working day of the week (Tuesday), for the past 50 Tuesdays at the same hour. They used a "Moving Average" methodology on the data collected in each hour of the day.

Bernardo et al. [5] and Macdonald et al. [28] got their input data from the historical data that the hospital had recorded. Where incomplete data was faced, interviews were done with the ED staff to collect supportive data.

Gunal and Pidd [17] used two sets of data from two different information systems that were collected and integrated in one file in order to be fed to the DGHPSim simulation tool (used in some of the general hospitals in the UK for building and observing possible improvement alternatives). The first set of data came from the hospital's Patient Administration System and the second set from Health Episode Statistics developed by the UK Department of Health. A tool was developed that could transform the data to a format that the simulation tool, DGHPSim, could use.

Baumgart et al. [1][4], Bernardo et al. [5], and Wijewickrama and Takakuwa [46] utilized an external application in the Arena package called *Input Analyzer*, and the input data was fit to statistical distributions. The distributions represented the behaviour of entities and were used instead of actual pure data. Centeno and Ismail [10] considered patient types differentiation, and separate distributions were defined for each type of patients by collecting related data. Moreover, through the studies conducted by Komashie et al. [16] and Rohlder et al. [38], data was gathered from information system databases and "Expert fit" was used for fitting the data to statistical distributions. Cochran and Bharti [11] used BestFit[14] to fit the data into known distributions.

Alvarez and Ceneteno [1] used *VBA in Arena* and developed an interface where input data was asked from the user and where the VBA code fits the given data to a statistical distribution for process timings or arrivals. This is configurable through database transformation libraries and settable durations and numbers of replications.

Wong et al. [47], Williams et al. [44], Koizumi et al. [23], and Holm and Dahl [20] discussed the *warm-up period* and excluded the data collected during that period. They made sure that the outputs were sufficiently accurate and reliable. Holm and Dahl [20] mention that the approach taken for figuring out the duration of warm-up period was to plot the outcome results for a specific duration of run time and decide visually on the period.

Simulation Setup and Run

Baumgart et al. [4] and Bernardo et al. [5] used *Arena for modeling*, just like many others (see Table 3). However, Spry and Lawley [41] did not use a simulation modeling tool. Instead, they developed written code that not only executed the simulation runs, but also

[14] http://www.palisade.com/risk/

helped make decisions according to multiple scenario results. Since their software has been developed specifically for a certain pharmacy, generalizing and porting the software for use in other pharmacies is a difficult task.

Raunak et al. [37] stated that *adjusting the resource allocation* is one of the possible approaches taken for simulation scenario runs. Adding or omitting resources should be decided carefully, knowing that omitting may cause overutilization of the remaining resources and adding may lead to underutilization of resources at some points.

Kolb et al. [24] proposed five ways of using buffers for EDs, each of which with a number of different scenarios to run. Wijewickrama and Takakuwa [45] proposed different appointment schedules as different *alternatives* to run.

However, Gopakumar et al. [16], Rohlder et al. [38] and Protil et al. [36] proposed a limited and *pre-planned* number of scenarios that needed to be checked for their outcomes. Gopakumar et al. [16] were only interested in running two specific scenarios for choosing one as a strategic decision. Protil et al. [36] checked three specific scenarios to see which one would be the most beneficial and would prevent the long duration of surgical centers idleness in the long run.

Model Validation and Verification

In some of the approaches, like the research conducted by Komashie et al. [27] or by Rohlder et al. [38], "verification" has not been done for the model and only "validation" of the process is covered. For verifying their models, Gopakumar et al. [16], Eskandari et al. [12] and Sepulveda et al. [40] used the *animation features* of the simulation software, which provided the researchers with insight about the logic of the system, and the system was deemed to be working as expected. Sepulveda et al. [40] verified the model by using a feature in the simulation software Arena. The "Trace" option gave the user this opportunity to verify the model according to simulation rules. Marmor et al. [29] validated their model by running simulations with 100 replications and comparing the results with existing data in a database. Afterwards, they used the input data as the data of a specific day until a certain hour of that day. Then ran the simulation model to check whether the outcome results were close enough to the actual data they collected.

A basic inspection approach has been taken for validating the model in the study conducted by Baumgart et al. [4] and Komashie et al. [27]. They validated their model by running the system with current data and comparing the model results with the actual data. Although they were facing a considerable difference in both patient throughput time as well as operation room running time, they announced that their model met the "overall credibility" criterion. They verified the model by showing the flow to clinical people and obtaining their confirmation. In another study conducted by Macdonald et al. [28], in spite of inaccuracies in four areas of the research scope, the model was said to be validated and verified and all the inaccuracies were considered as simulation limitations. The first issue was the total throughput time of the patients, which had around 100 minutes difference from the actual value. The second issue was the waiting times of patients for some specific processes that differed by about 13 minutes from the actual data. The third issue was the assignment of the patient to a specific doctor until the end of the care process, which was not modeled properly. A fourth issue was that at the beginning of each day, the number of patients in the ED was set to zero, omitting all the existing patients from the system.

Wijewickrama and Takakuwa [46] observed that the model validation was the most time-consuming step of making their simulation model. Eskandari et al. [12] validated their model by calculating the patient throughput time during simulation runs and comparing it to the 30% of the input data that was kept unused for validating the system. Eventually the model was said to be valid and verified.

In a different paper, Wijewickrama and Takakuwa [45] also validated their model by using a "data generator" coded in VBA, through which the user is allowed to modify the input data (without making any change to the model itself) and observe the outcome. The authors verified the model by setting different patient arrival rates and increasing the rate percentage in each run to figure out the effectiveness in the average waiting time of the patients. Also, the number of patients observed was raised by increasing the arrival rate.

Reports and Outcome Analysis

Macdonald et al. [28], Kolb et al. [24], Gopakumar et al. [16], Komashie et al. [27] and Kolker [26] illustrated the output results of simulation runs using different charts and tables for a quicker revision and evaluation.

Baumgart et al. [4] defined a specific and limited number of resource scheduling configurations to run. The resource utilization is under observation for its quantity and percentage of workload. By increasing the number of resources, the utilization of resources decreases. However, at a certain point, this approach seems to have a negative impact on performance.

Wijewickrama and Takakuwa [46] used another approach for finding the optimized alternative. Arena's *OptQuest* is used as an optimization application to automatically search for the best solution, given an objective function and constraints. Calculations were done for waiting times of each patient category as well as for an overall patient waiting time. The outcome results were showing that 70% of the total waiting times were related to two specific categories of patients among ten. The intention was to figure out a proper scheduling for reducing the patient times without adding any resources, which in the end was accomplished. Weng et al. [42] used the "National Emergency Department Overcrowding Scale" (NEDOCS) in addition to the OptQuest results and came up with an optimized scenario. NEDOCS is calculated with a specific formula[15] and if the value is higher than 100, then this indicates that the ED is overcrowded.

Eskandari et al. [12] showed that by applying new strategies and with a really low cost, the patient waiting time was reduced by over 40%. In this study, fourteen *scenarios* were specified by experts. For choosing the best scenario performances, the Analytic Hierarchy Process (AHP) was used followed by a technique called Technique Ordered Preference by Similarity to the Ideal Solution (TOPSIS). After criteria were set for choosing the best scenario, AHP was used for quantifying the importance of each of the (sub)criteria and to be able to prioritize the most important ones. After each criterion got weighted, each scenario had a certain weight. The strategy behind TOPSIS for choosing the best scenario is that the eventual alternative should have the shortest distance from the Positive Ideal Solution, and should have the furthest distance from the Negative Ideal Solution. In this study, the alternative with the maximum score for the TOPSIS measure indicator was chosen as the best scenario. Centeno and Ismail [10] integrated Arena with a

[15] http://www.nedocs.org/

linear programming software (LINGO[16]) using VBA coding. The decisions made were optimized and the best solution was considered.

3.4.3 Challenges, Limitations and Restrictions of Simulations for Healthcare

Several issues were reported in the study conducted by Marmor et al. [29]. The first one was the approach taken to make sure the simulation model was reflecting the real ED system. Another issue addressed was input data accuracy. One of the reasons for these inaccuracies is that the patients who were done using ED services sometimes needed to be transferred to other units of the hospital. However, if the destination unit was already full, the patient should have to continue waiting in ED for a free spot in the next unit.

Baumgart et al. [4] announced that their simulation model did not reflect proper *resource sharing* in parallel processes. Another limitation in their model was that in reality, two types of resources could *interchangeably* do the same task, a situation that the model could not handle. The third problem mentioned was the missing factor "quality of performance", which could not be evaluated through a simulation model.

One of the challenges reported in the study done by Raunak et al. [37] is that their simulation tool did not support *staff quantity changes* during working hours of a day.

Weng et al. [42] made several assumptions for creating their model, which are basically the limitations faced:

- The staff productivity quality of work was assumed to be uniform.
- Departments were operating individually.
- Patients LWBS (Leave Without Being Seen) was not considered in creating the simulation model.
- ED was in a normal situation and no external events were interfering with the current state.
- Trainees were excluded from the model.

With all these challenges involved, this study concluded that adding one resource would significantly impact the outcome results.

[16] http://www.lindo.com/products/lingo/

Barjis [3] considered the data collecting procedure to be one of the most challenging parts of modeling and simulation for healthcare. The required information is often not included in the historical data and therefore should be collected through an observation approach or through discussions with clinical people. Using the latter methods, the data collection error will increase and may even result in incorrect output results. Daily decision making seems to be a very effective way of improving system performance, only if the simulation model is flexible enough to reflect accurately the real system. If this constraint is not met, the simulation will be useful only for a captured scene of the current system and not for constant improvement. Furthermore, Barjis [3], Preston White [35] and Young et al. [48] mentioned that healthcare is among the domains where *human factors* interfere the most. The human interactions that are not systematic and can be subjective on certain occasions will amplify the complexity of modeling patient care processes. The models should reflect the system with all its complexities, which is basically a challenge in simulation for healthcare. Also, Barjis [3] added that regarding limitations, not much attention has been given to conceptual modeling, which is the first step of a simulation and modeling process.

These many challenges are echoed by Carter and Blake, who shared their views on different issues related to the use of simulation in healthcare based on many years of experience and experiments [9].

3.4.4 Complementary Papers

Rozinat et al. [39] believe that real-time simulations can be used for short-term operational decision making in organizations. They also mention that until now, most of the simulation studies have focused of approaches for strategic and long-term decision making. They used real-time simulations for modeling a credit card process to define the usability of the real-time simulation approach for decision making. They have defined their own language and have used a coloured Petri-net for simulation modeling. They have integrated a workflow manager with a simulation model, they *loaded* the simulation model with the initial state data related to the credit card applications in process, and they predicted the state of the system for near-future events. This is very close to what we propose in this thesis, but they do it for a domain where accurate and timely data is easier to get. We will revisit their approach in Chapter 7 in a more detailed manner.

3.5. Summary

According to papers reviewed in this chapter, simulation is a technique that enables obtaining significantly valuable information about a current system as well as future system performance, even in a healthcare context. Simulation can be used to forecast upcoming events and related consequences. Accessing this information can help make appropriate decisions prior to predictable event occurrences. We have also learned that Arena is a popular simulation tool in healthcare and that not much consideration has been devoted in creating a conceptual model prior to building a simulation model in healthcare.

Most of the studies focus on simulation capabilities for facilitating strategic decision making. In these studies, the scope was assumed to be in a normal situation without any interfering external events occurrences during the prediction horizon. Less consideration has been given to using simulation for operational decision making in healthcare.

The benefits of using concurrency of simulation runs have been discussed in some of the studies. However, most papers did not mention any actual trials for experimenting on how effective a real-time simulation would be in terms of performance analysis and decision making. In fact, there was only one paper really discussing real-time simulations in healthcare [29], and very few are actually published in other domains, except for the research conducted by Rozinat et al. [39].

In many approaches, the scope of the simulation modeling was limited to one specific department of a hospital. For example, ED is one of the most popular departments for simulation research. Although each of the departments of a hospital is assumed to be working separately from one another, there is still an interaction among them as the patients are being frequently transferred from one department to another. Every decision each department or section makes somehow affects the rest of the departments. For instance, the ED manager can make a decision that would result in less congestion in ED but would also lead to an overcrowding situation in some other department (sometime resulting in a worse patient flow performance). Therefore, the larger the scope the more complex the model becomes, so the most appropriate modeling way should consider the related consequences of decision making for several departments or sections simultaneously.

Based on this knowledge and on the understanding of the domain built from this systematic literature review, the next chapter will present the general phases and steps our methodology for real-time simulation of healthcare processes, which will then be instantiated for our case study in Chapter 5 and Chapter 6 (one for each of the two phases).

Chapter 4. Real-time Simulations in Healthcare

This chapter presents a new methodology that enables real-time simulations in healthcare. It contains two phases that should be completed for realizing real-time simulation capabilities using a simulation tool: model creation and model execution. Each of these phases is divided into detailed steps that should be followed in order to ensure a real-time simulation model is built that reflects the actual flow under observation. These steps are discussed in the following sections.

4.1. Real-Time Simulation Steps – Overview

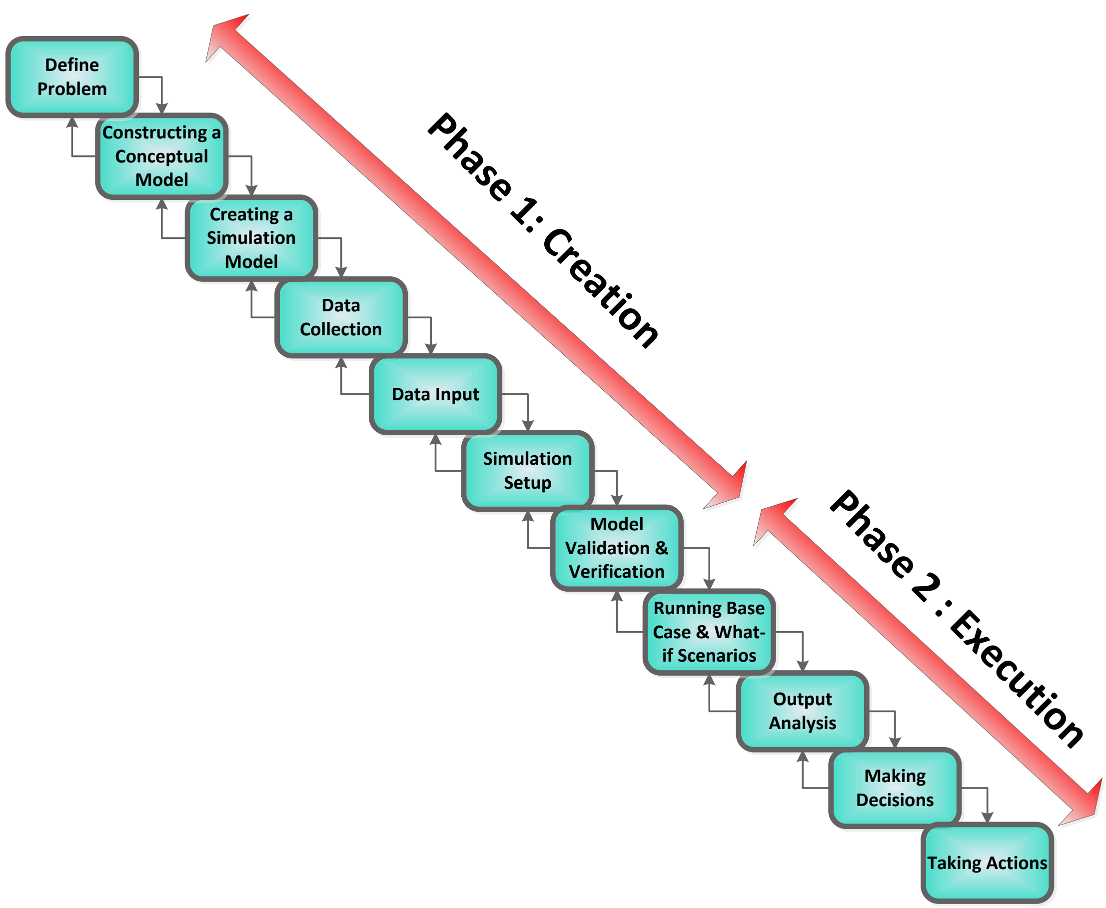

Figure 4 Real-Time Simulation Steps

Certain steps should be followed for prototyping a real-time simulation of processes. If the real-time simulation model is being created for the first time, the steps illustrated in Figure 4 should be followed. After the model is created and has run once, there is no need to repeat all of the steps again, and only some of the steps mentioned in Figure 4 should be followed. Figure 5 shows these steps as an iterative cycle.

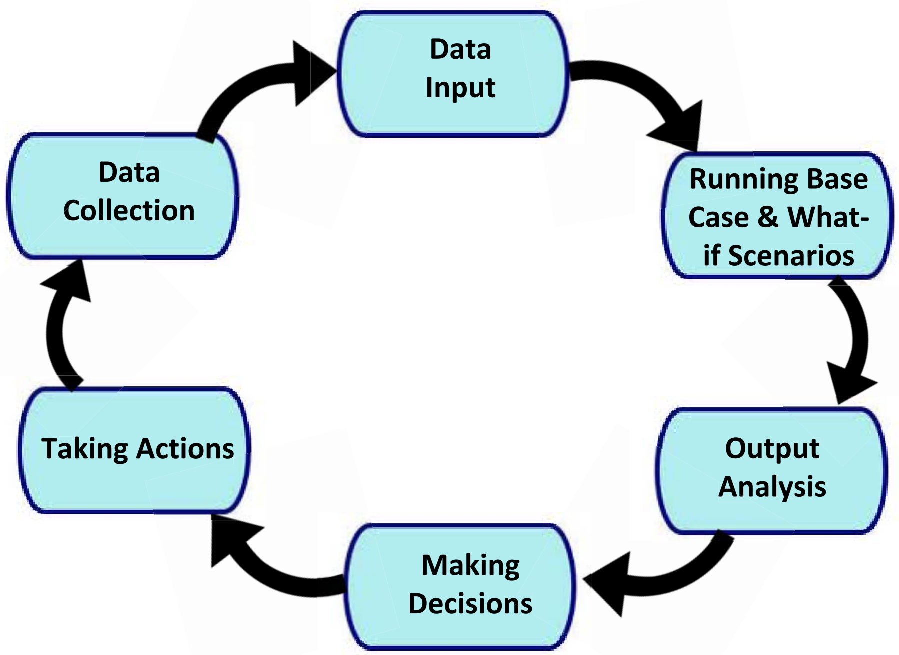

Figure 5 Real-Time Simulation Cycle

Note that the steps in this chapter focus on real-time simulations in healthcare, but are defined independently from process modeling and simulation technologies (e.g., BPMN and Arena).

4.2. Steps of Phase 1 – Model Creation

The first phase is composed of seven general steps.

4.2.1 Defining the Problem

Simulation is a technique used for solving problems by emulating real-world experiences and proposing subsequent actions after prediction of probable future outcomes. Therefore, at the initial stage, the objectives should be specified and the problem should be clarified along with its scope. This step should be done carefully in order to be able to formulate the statement of the problem properly. Not all types of problems can be solved through a simulation approach.

Queuing management is one field of knowledge for which the simulation approach seem to be usable and convenient. Queues are defined as a lineup of entities waiting in order to get a certain service. Queuing systems are the type of systems that follow queuing disciplines. Concepts like entity arrivals, service time and server capacity are definable in these systems. The common problem faced in queuing systems, as the name indicates, is the existence of the queues, which is a considerable factor from a business point of view. Queues exist because of the growth in the demand that should be responded with a limited available supply. The length of queues and the time that the entities spend in the queues are counted as business performance criteria. Discrete event simulation is one simulation methodology to model queuing systems that can help reasoning about the cause of waiting times and find alternative actions that can help reduce such values.

4.2.2 Creating a Conceptual Model

Often, systems and their behaviour are very complex, and an appropriate level of abstraction should be used for a sufficiently detailed model representation that can address the identified problems. A major requirement for developing an accurate simulation model is to be able to look at the actual system from the right perspective. A good understanding of the system under observation can help reduce ambiguity. If a system is about to be observed for further analysis, the behaviour of the system should be monitored and captured. A *conceptual model* is used as a prior step before creating simulation models. It specifies the precise descriptions of the information required for building the model as well as an explanation of the structure and behaviour of the system under observation.

A conceptual model can be captured with a graph, pseudo code, a state machine or even a mathematical formulation. Depending on the depth of precision required for defining the system, different types of conceptual models can be used. Clarifying the scope of the work and defining the detailed description of the steps that should be followed are essential in order to create a valid simulation model.

4.2.3 Creating a Simulation Model

For developing a simulation model, all of the system components within the scope of the defined problem captured by the conceptual model should be modeled. The conceptual model should be translated to the simulation model. The statistical input data required by the simulation model but that is not present in the conceptual model, has to be identified so that it can be acquired in the data collection step. The sequence of the activities and events involved within the scope of the defined problem should be modeled with a required level of abstraction. If a simulation tool is being used for this purpose, the real-time implementation (the initialization of the system from a predefined state) becomes a necessity. Some information needed for constructing the simulation model might not be captured through a conceptual model, but can be handled through data collection.

4.2.4 Data Collection

A real-time simulation requires real data to be collected from a database. This database (or multiple databases) should include the historical data as well as the initial state data.

- **Historical data**: This is the accumulated data related to the previous performance of the (hospital) events and activities. Past activity trends can be extended to future performance of the activities, and hence this information is one of the substantial ingredients of a simulation model. The service time for entities at each location, the arrival of the entities to the system, and the data related to the availability and capacity of the resources involved in the flow are part of the historical data.
- **Initial state data**: This data should also be fed to the simulation model. A real-time simulation starts from the real state of the organization/hospital and not from an “empty and idle” state. Therefore, the initial state data should be loaded in the simulation model prior to the start of every run. This requires the initialization of the state and capacity of each resource for a simulation run and the present number of entities waiting in the queues for getting served.

The injection of the historical data is commonly done in discrete event simulation studies. One distinctive feature of real-time simulation is the collection and injection of *initial state* data, in order to bypass the warm-up period.

4.2.5 Data Input

The collected historical data should be processed and fit into the mathematical distribution that best describes the behaviour of the activities and events involved in the process. These statistical distributions will be generating random data during at run time, however this data will be following the distribution's characteristics. This input data should be provided through an interface or an external file feeding the simulation model prior to every simulation run.

The data input should be done in a timely basis (e.g., every 8 hours, every 12 hours or every day, depending on the needs of the organization) for enabling the short-term decision making process to be continuous and efficient.

The precision of the initial state information is of significant importance, since the warm-up period is to be avoided by loading the initial state of the simulation at each run. The historical data related to the duration of the activities is less likely to change in a short period of time. However, the distribution of the arrival rate should be calculated and fed to the model as often as the initial state of the hospital. This is because of the unpredictable and random nature of the patient arrivals in ED during different days of the week and hours of the day.

4.2.6 Simulation Setup

The setup for the simulation runs needs to be done carefully since real-time simulations target near-future predictions rather than long-term predictions. We first explain some differences between a simulation setup for long-term prediction and for short-term prediction.

Warm-Up Period in Long-Term Simulations

As discussed in Section 2.3, the warm-up period is a transient period after which the system reaches a steady state and the output values become stationary. If a simulation runs from an empty and idle state and for a long enough period of time, then the steady state starts from an autonomous state of the system. Figure 6 shows the formation of the transient state and the steady state of a simulation run for an output variable.

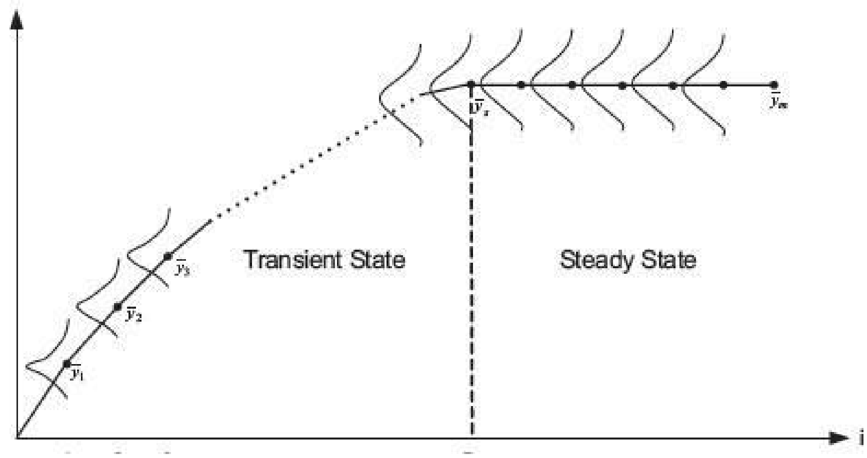

Figure 6 Warm-up Period and Steady State [6]

The steady state occurs because the system will reach an equilibrium state and will get stabilized after a while, when reflecting the future state of the system based on the historical data.

In most of the studies from the literature review, the application of simulation has been observed for strategic decision making. The uses of simulation for long-term or strategic decision making generally requires that non-terminating systems always reach their steady state for observations to become reliable. It is then assumed that in the long run, the system is only affected by the same number of elements as is being affected currently.

Warm-up Period in Short-Term Simulations

The warm-up period is bypassed in a short-term simulation process in order to provide results quickly. Since the current information of the system is loaded in the simulation model, the simulation does not start from an empty and idle state anymore and therefore the data collection that occurs even at the very beginning of the simulation runs is reliable. Thus, in real-time simulation for short-term predictions there is less complications in terms of defining the warm-up period and avoiding it while collecting the output data.

Run Length Setup

The run length and number of replications in a simulation model should be set in accordance with the objective of the study. In this study, the objective is to enable reliable short-term decision making process using real-time simulation. The *run length* is the horizon for which we would like to base our decision making process. This value is equal to the period of time in a near future that we would like to predict for, repeatedly. For a real-time simulation to be effective, this length should neither be very small (to get sufficient accuracy and confidence in the results) nor very large (as we want real-time, short-term predictions). For instance if in a process all of the activities take at least 2 hours to accomplish, running the simulation for 1 hour ahead of time will not provide us with useful results. Also, as quick as the process of setting up and running the alternative scenarios might be, it still takes some time and this time should always be a very small proportion of the time horizon for which we are doing the short-term predictions.

The terminology that has been used above as “not too large” and “not too small” cannot be quantified to specific values that are context independent. The reason is because the chosen length of run is totally dependent upon the nature of the activities involved in the system and this value differs from system to system. If the behaviour of the system is observed, the appropriate value for run length can easily be chosen for the simulation runs.

Replication Setup

After the run length is set, the *replication number* should also be decided. Normally, both the run length and the number of replications are the two determiners of the precision and reliability of the output data. However, since in real-time simulations the run length should be chosen as a non-changing quantity in a given context, then the only factor that determines the precision of the output data is the number of replications. There is no specific precision that is common or typical along the sampling process. Determining the precision required for the output results of simulation is more or less the choice of the decision maker while considering the time constraint for running the simulation model. Thus, even though a higher number of replications would result in a much better estimation of the actual output results, the time required for running the simulations should not be ignored and in fact should be analyzed. The decision maker should come up with a certain precision that is both affordable and useable for providing solutions for the defined problem presented by a simulation model. The precision of the output data can be managed through the statistical tests mentioned in Section 2.6.

4.2.7 Model Validation and Verification

Model *verification* is meant to ensure that the model is behaving as intended. For verification, one can change the parameters inside the model and check if the outputs are as expected or not. Verification, as important as it is, is not sufficient for expecting accurate output results. The model should also be validated before any actual implementation occurs. *Validation* is the process of ensuring that the model sufficiently reflects the *real* system and that there is not any considerable difference between the model and the reality. This process can be done through confirmations with the experts who are directly involved with the activities in the real-life process.

4.3. Steps of Phase 2 – Model Execution

By following the steps of Phase 1, the creation of the model is completed. The second phase of the methodology is composed of four additional steps targeting execution and analysis of the model.

4.3.1 Running the Base-case and What-if Scenarios

The next step is to execute simulation runs and analyze the results. Since the final objective of the real-time simulation is to propose an effective solution that could solve the defined problem, there should exist some performance criteria against which we will be evaluating the future state of the system through each what-if scenario run. Therefore, firstly the performance metrics should be defined. *Performance metrics* allow us to measure the effectiveness of the suggested solutions via quantifiable values. In most of the systems, there is more than one metrics that should be considered. At least one of the performance metrics should directly be in relation with the problem definition.

The first scenario is known as the *base-case scenario*, which gives a view of the state of the system in a near future if it keeps performing the same way as it was normally planned to. The *what-if scenarios*, as the name indicates, are alternative scenarios that show the predicted state of the system if at the present time some effecting parameters are changed. These scenarios show the different configurations through the flow that may affect the values of the measurable criteria.

4.3.2 Output Analysis

A simulation model collects the tracking information of all of the events and activities that take place while a simulation is running. Any type of statistics related to the performance of the system can be generated for a precise analysis. However, the major parameters to focus on are the performance metrics set previously. Usually, the output responses to a specific change in the parameters would not be in the same direction (better or worse) for all the outputs. Within a change, the value of a performance metric might become much more desirable while at the same time causing the value of another metric to become very undesirable. The *trade-offs* between all the metrics should be analyzed as an initial step for decision making.

4.3.3 Making Decisions

After analyzing the outputs generated from each scenario, we should look for the *best scenario* in terms of the measurable criteria set for the comparison among different alternative scenarios. Decision makers should follow an approach for deciding on the best implementation scenario.

One algorithm that is used to find the best alternative among a set of alternatives is to use *pair-wise comparisons*. In this approach, an alternative will be compared to another alternative by the measurable criteria set previously and the better alternative will be kept. This process should continue until there is only one scenario left, which is believed to be the best alternative among all the alternatives considered. Most of the times in a multi-criteria decision-making approach, it is not possible to have all the feasible alternatives evaluated to find the best solution among those. Therefore, the decision maker will be surrounded with a countable (and likely small) number of options to choose from.

In this approach, even though the criteria are set previously and the most desirable values would be extreme minimums or maximums, there will mostly be trade-offs among the criteria. The decision maker should consider the variability among the different criteria and pick the one that would result in the most reasonable and possible outcomes.

4.3.4 Taking Actions

As the real-time simulation runs occur in a timely way, the output of every scenario is analyzed and the best decision is made. Sometimes the best decision may be to take no action and stay with the base-case scenario (status quo) rather than following one of the what-if scenarios. Otherwise, appropriate actions should be taken according to the previous step results.

4.4. Summary

The methodology is meant to be used in a closed loop fashion, meaning that the simulation data will be updated with real-time data of the hospital so the decision maker can run simulations to predict the outcome state of the process in a near future (e.g., 4 or 8 hours ahead). Although certain steps are suggested to be followed for using simulations and solving a problem, it is possible that some new, previously hidden aspects of the system under observation might become visible. In this case, the backward arrows shown in Figure 4 will be used, and changes should be applied to the previous steps to be able to move forward to the next steps again. Healthcare processes are much more complex than processes where the activities are automated and predictable. Completing the above steps can help ensure a representation of a consistent and accurate model that sufficiently reflects the actual system while enabling real-time simulation capabilities. The first three steps should be done carefully and precisely since these steps will be done once in a long period of time; afterwards only the injection of the data and analysis will be repetitively occurring.

The next two chapters focus on the application of each of the two phases of the methodology to a validation experiment on a real hospital process.

Chapter 5. Creation of a Real-Time Simulation Model

This chapter instantiates and validates the first phase of our methodology on a cardiac patient flow at a community hospital in Ontario. Arena is used for the simulation.

5.1. Selected Care Process

William Osler Health Center (or *Osler hospital* for short), located in Ontario, is one of the largest high-volume community hospitals across Canada. The hospital currently provides 830 beds with 4700 employees. The average annual number of people visiting the hospital for care services is about 180,000. Almost 75 % of the people who get admitted to the hospital wards enter from the ED of the hospital. Osler hospital currently has two campuses: Etobicoke General Hospital (EGH) and Brampton Civic Hospital (BCH). Figure 7 shows the annual Osler report of patient length of stay in ED according to patients' severity of health condition in 2011/2012. As seen in Figure 7, for some of the months, the patient length of stay is higher than the provincial target. Provincial targets are 8 hours for patients with complex health conditions and 4 hours for patients with less severe health conditions).

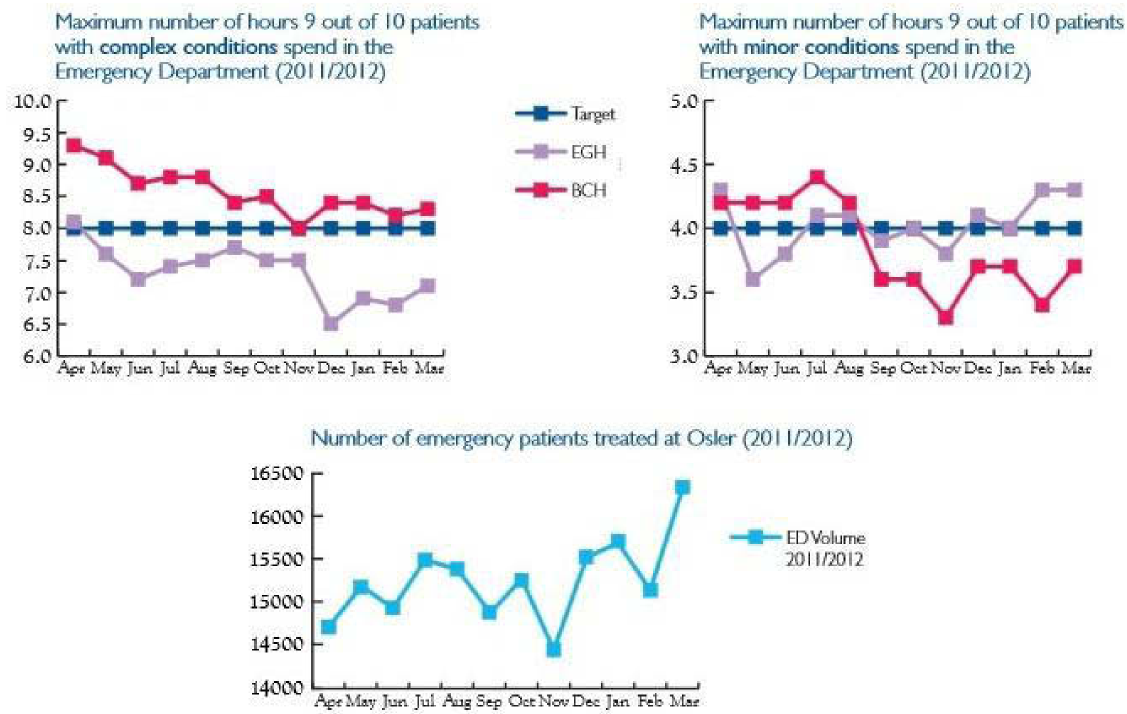

Figure 7 Length of Stay in Osler Hospital 2011/2012 [43]

Table 4 shows a more recent report on patients' length of stay in emergency department of Osler hospital.

Table 4 Length of Stay in the Emergency Department of Osler's Brampton Civic Site (March 2013) [32]

March 2013					
Hospital Site	Average Time Spent in Emergency Room		Total Time Spent in Emergency Room [9 out of 10 patients]		Hospital Type
	Complex conditions (Hours)	Minor or uncomplicated conditions (Hours)	Complex conditions (Hours)	Minor or uncomplicated conditions (Hours)	
Provincial Target	-	-	8	4	
Provincial	**5.4**	**2.1**	**10.1**	**4**	
WILLIAM OSLER HEALTH CENTRE - BRAMPTON CIVIC SITE	**6.3**	**1.9**	12.1	3.5	Very-High Volume Community Hospital

This case study covers the implementation of a simulation-based short-term decision making process for the cardiac patient flow of Osler hospital. The entire clinical process for cardiac patients of the hospital was chosen for this study. The healthcare process selected involves several units, including the emergency department (ED), the cardiology ward (CW), the cardiac catheterization laboratory (CCL) and the test laboratories (TL). There are 5 types of cardiac patients with different CTAS levels that are dealt with in the cardiac patient flow. Therefore, the patients will receive clinical services based on the severity of their health condition.

Figure 8 highlights the cardiac patient flow of this case study. The physical operation of the patient pathway starts with patient arrival to the ED and continues with triage, then a physician consults with the patient and orders tests from TL. TL provides test results, and then the physician does a second consultation. If the patient is of a high acute level and must be hospitalized for further care processes, he/she will be moved to CW, then undergoes surgery in CCL, and finally returns to CW until he/she is discharged. Various delays can occur and hence influence different wait times. Several alternatives and potential concurrency are also considered along the way. We are taking this specific patient pathway for our study. There exist some other steps that might be followed for some special cardiac patients; however, we do not focus on them in this research.

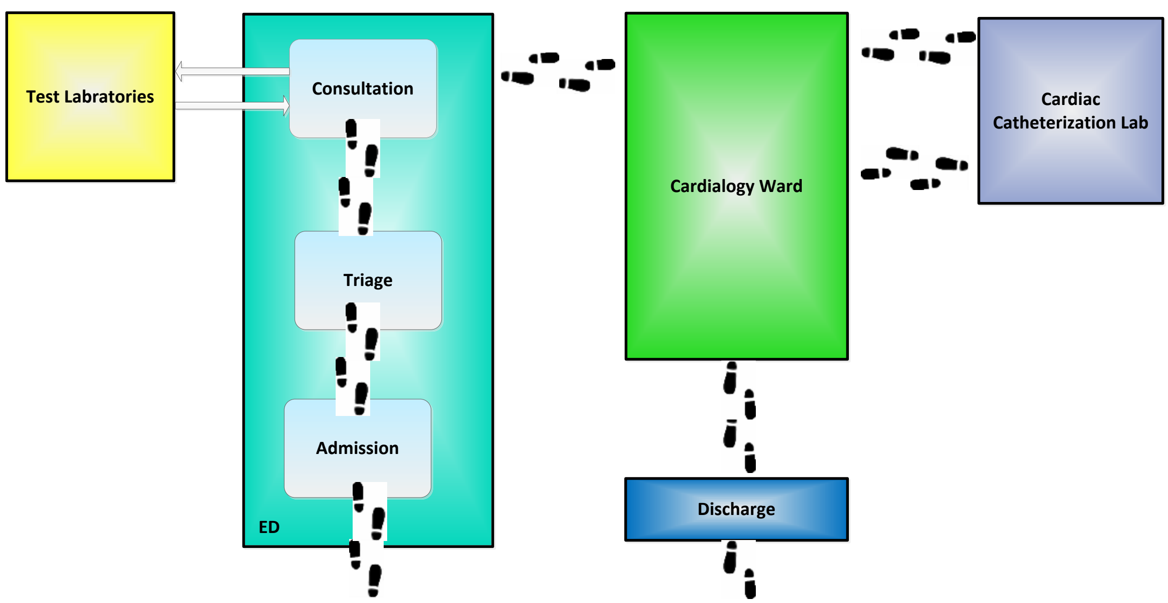

Figure 8 Cardiac Patient Flow in Osler Hospital

Note that this simulation is part of a larger project that relies on business process management and real-time location of staff and equipment to enable real-time patient flow monitoring and analytics (Figure 9).

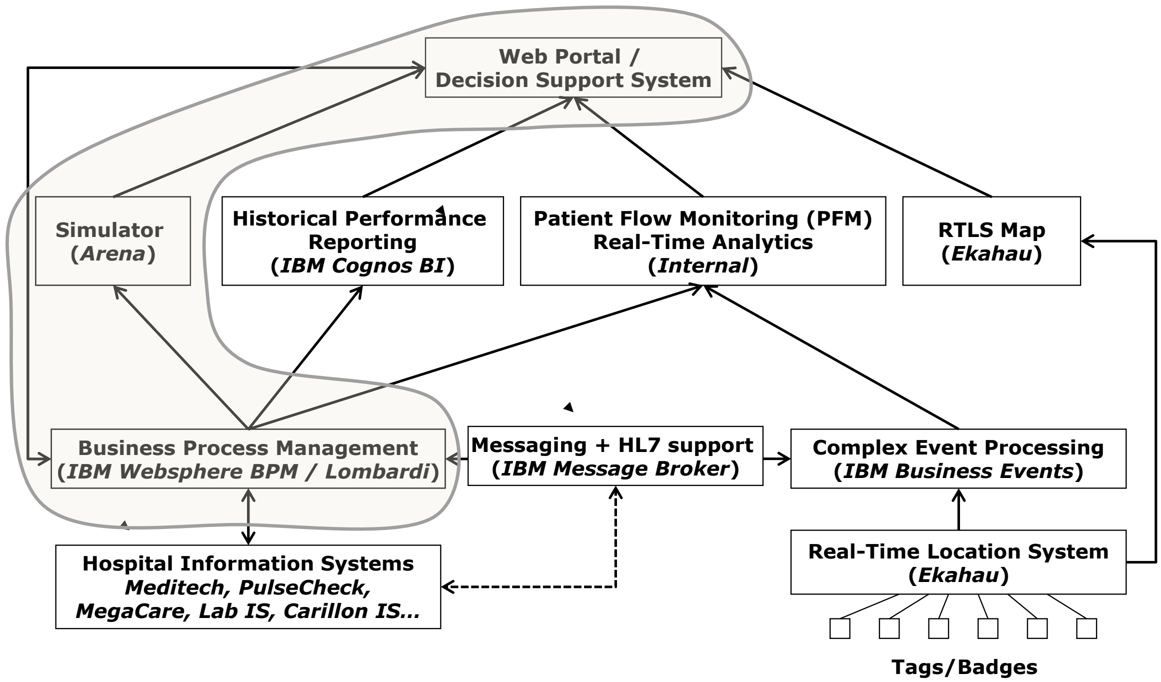

Figure 9 Use of Simulation in the Larger Osler Project

The Arena-based simulation (highlighted in the figure) takes advantage of this infrastructure to extract the *initial state* of the simulation with high accuracy. The results of the simulation are eventually expected to be used in a larger decision support system currently being designed by a Ph.D. student in his thesis.

The real-time simulation methodology steps discussed in Chapter 4 are followed in this case study. The creation phase is discussed in this chapter and the execution phase is covered in Chapter 6.

5.2. Define the Problem

The cardiac patients of Osler hospital are currently experiencing high waiting times. Waiting time can be caused by several reasons and there might not be a complete solution available to eliminate it. However, it is believed that a real-time simulation can help reduce the patient waiting time by providing the decision makers with feedback on the effectiveness of short-term decision alternatives based on adjustments in resource scheduling. Thus, a simulation model is created for analyzing the mentioned problem with the scope of the most common cardiac patient pathway at the hospital. Consequently, at the initial step, it is expected from the simulation model to be able to provide the following information regarding the defined problem:

- The total waiting time of patients for all activities in the near future.
- Consequences of possible resource scheduling adjustments on patient total waiting time in the near future.

5.3. Constructing a Conceptual Model

A conceptual model has been partially defined prior to the creation of the simulation model. In this research project, the cardiac patient care process of Osler hospital has been previously modeled in BPMN (see an extract in Figure 10) using a commercial business process modeling tool, namely IBM Business Process Manager (BPM) [2]. This model is itself based on a care flow modeled more informally by senior nurses at the hospital. This flow was defined to capture the engagement of the physical clinical processes with a real-time patient flow monitoring system [8]. Thus, this *executable* business process model reflects the interactions among different actors with a precise consideration for the sequence of activities that take place in the flow process, from the patient's perspective.

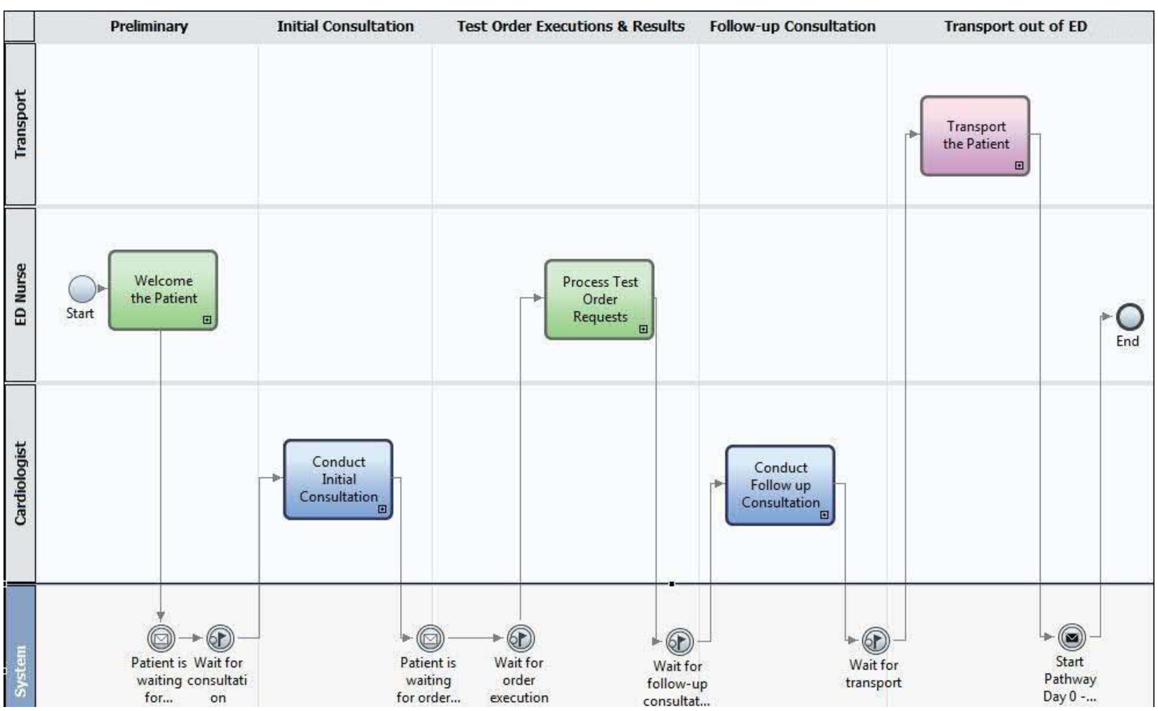

Figure 10 Overview of BPMN Model of Cardiac Care Process

With the business process model and the simulation model following separate objectives for reflecting the reality, all of the information captured in one is definitely not reflected in the other. There is specific information of interest for creating a simulation model that is captured in a business process (BP) model. However, the information represented in the business process might not be sufficient for creating a complete simulation model.

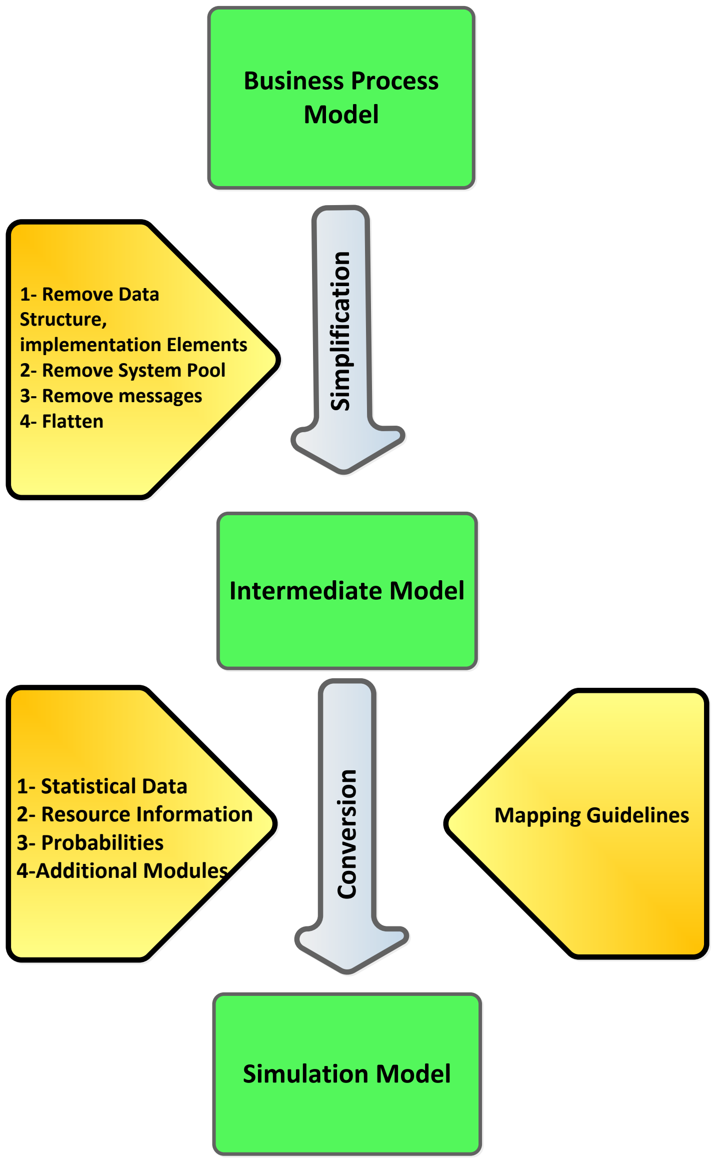

Figure 11 Guidelines for Mapping a Business Process Model to a Simulation Model

A mapping approach for creating an intermediate model is used for this case study. We have removed the unnecessary information that a BP model presents, due to its own specific objectives, to create an intermediate model that essentially captures the conceptual model. Afterwards, we have added the information not initially captured in the BP model that should be reflected in a simulation model. Figure 11 shows the complete approach that was taken to create a simulation model from a BP model partially representing the conceptual model. The mapping approach is supported by simplification guidelines (to produce the intermediate, conceptual model) and by mapping guidelines.

5.3.1 Intermediate Model

An intermediate model has been created that is a simplified version of the business process model. This intermediate model is made up of business process modeling elements that do not necessarily follow the BPMN language rules. One assumption was made while building the intermediate model: the business process model that was defined had not considered the role of the "admission and discharge clerks" and had assigned their tasks to the rest of the resources. While creating our intermediate model, we created these roles and assigned these tasks to them. There were several points that were considered while building the intermediate model represented next as simplification guidelines.

5.3.2 Simplification Guidelines

Four important guidelines helped simplify the source executable BPMN model and convert it to a conceptual-level BPMN model to be transformed later to a simulation model. A simple example of how we created the intermediate model is illustrated in Figure 12.

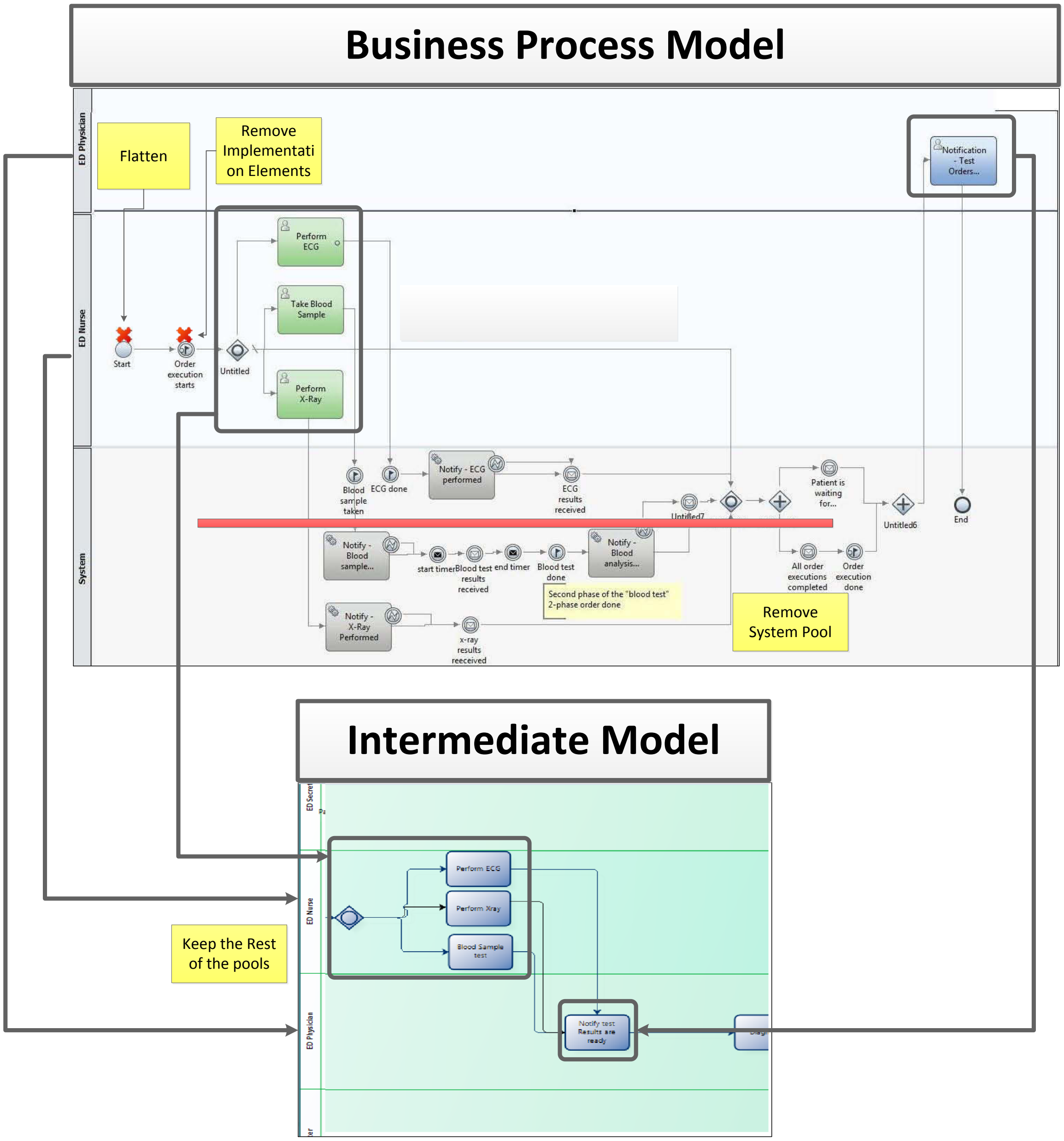

Figure 12 Creating an Intermediate (Conceptual) Model from a BPMN Process Model

1) Remove Data Structures and Workflow Implementation Elements

The original BPMN model is an executable business process model for monitoring the patient flow. Thus, there are several elements that exist in this model for realizing and exacting the execution of the flow. For instance, a *time tracker* node is set in a business process model to track the timings of the activities and events in real-time. All the data structures and (user interface) implementation elements were simply removed when creating the intermediate model. Their existence is subject to the specific implementation objectives and is not required for simulations.

2) Remove System Pools

Since the business process captures the care flow from a system's point of view, information captured in a BPMN model reflects two types of interactions:

- Interactions between the system and the participants.
- Interactions within the system itself.

Since Arena is an entity-driven simulation tool, only the elements that directly impact the entities are modeled. Therefore, the first type of interactions will be modeled indirectly in the simulation model whereas the second type of interactions will not be reflected at all [2]. *System* is one of the main actors in the business process model that sends, receives and records required information (see the bottom of Figure 10). The interactions within the system are reflected in the system pool (where a BPMN pool is a horizontal region containing activities). This actor, as important as its role is in the business process model, is not covered in the simulation model. This means that all of the elements existing in the system pool of a business process model should not be included in the intermediate model and, therefore, the system pool should be removed. The other pools were kept as is.

3) Remove Messages

There are several types of event nodes used in BPMN. The event nodes model the occurrence of events that might happen at the beginning, at the end, or during a process flow. The mapping guidelines related to the event nodes that are being used at the beginning or the end of the process flows, will be discussed in the next section. One of the common types of event nodes used in our business process model is the *message node*. The message nodes that exist in the middle of the process flows are often used to indicate the initiation or completion of an event, or to send and receive messages in order to communicate the result of a task or a process. This information transfer is not a visible physical task from the patient's point of view, and therefore shall be ignored in the process of building the intermediate model.

4) Flatten the Business Process Model

BPMN supports hierarchical decomposition. Due to the complexity of the cardiac care process model, this process has been modeled using sub-processes. However, for simplicity, the intermediate model is flattened and shows one process flow at one level that requires all of the start and end nodes to be removed, and only the first start node in the first business process and the last end node in the last process are to be kept.

5.4. Creating a Simulation Model

The Arena simulation model was created using the following *guidelines* for converting the intermediate model to the simulation model and finally for adding required information specific to the simulation model.

Before discussing the conversion guidelines, some of the most typically used modules in the Arena simulation tool are briefly defined:

- *Create Module*: This module is used to initiate the discrete arrival of the entities to the simulated process.
- *Process Module*: This module corresponds to an activity or a task involved in a process in entities point of view.
- *Decide Module*: This module is used whenever an entity decides among different possible branches through a process logic according to a specific condition or by a certain probability.
- *Dispose Module*: This module is used to remove the entities from the process after the process is completed.
- *Assign Module*: This module is used to assign an attribute to entity, to assign a value to a variable, etc.
- *Record Module*: This module is used to collect and record the required statistical values of interest.
- *Delay Module*: This module delays the entity by a certain amount of time.

5.4.1 Conversion Guidelines

As mentioned in Figure 11, all the information captured by the intermediate model is *mapped* to the simulation model following certain guidelines.

Activity to Process

In the intermediate business process model, the concept known as *activity* is used to present a duty assigned to an individual. As Table 5 indicates, the activities that exist in the intermediate model can be categorized into three groups. The first type is the starting activities, corresponding to the arrival of entities in the system (in this case study, patients who arrive at the emergency department). The entrance of patients in the hospital is shown as a task in the intermediate model. In the simulation model, the entities need to be created at the beginning of the process, letting the model get populated with the entities. The equivalent of these types of activities is a *create* module in the Arena simulation model. The additional information that should be added for this module is the rate of arrivals and the number of entities entering at each time step.

Table 5 Concept Conversion – Activity

Business Process Model Concepts	Condition	Arena Simulation Modeling Concepts	Additional Information required?
Activity	Arrival of entities to the system	Create module	Yes
	Sending or receiving notifications or requests	Included in a process module	Yes
	Tasks assigned to participants excluding system related tasks	Process module	Yes

The second group of activities that exist in the intermediate model are the ones corresponding to the activities through which a participant is using the system for communicating with the other users. These activities are the ones that we indirectly modeled (previously discussed in the “2) Remove System Pool” section. These activities are always part of a physical activity. For example, after the ED nurse does the triage and requests for initial consultation to be done by the ED physician, in the patient’s point of view, the activity of “requesting for initial consultation” is part of the triage activity. Therefore, the detailed break-down of the activity components is not necessary to be reflected in the simulation model and these activities can be merged with the related physical activity. These tasks are being modeled in a simulation through one “process module” that captures both the physical activity and the participant’s related communications via the system.

The third type of tasks in an intermediate model contains the physical activities involved in the flow that will be directly modeled as *activities* in the simulation model using a *process module*.

The supplementary information that should be added to the second and third groups of tasks includes the duration of the activities, as well as the resources assigned to each activity.

Pool and Lane to Resource

An intermediate model contains pools and lanes, which show different participants in a process flow, regardless of their capacity and availability.

Table 6 Concept Conversion – Lane vs. Resources

Business Process Model Concepts	**Condition**	**Arena Simulation Modeling Concepts**	**Additional Information required?**
Pool	-	-	-
Lane	-	Resource	Yes

In the intermediate model, a pool shows the organization for which the flow is being modeled and the lanes show different organizational roles. These roles are shown mainly to specify task authorization. In Arena, tasks require to be authorized and assigned as well. The authorization of each task will be done inside each implemented related module. Figure 13 shows the conversion of a lane to the resource. However, this information is not sufficient for a simulation model. Therefore, additional detailed information related to the capacity and the availability of the resources for a more precise analysis of each resource behaviour and its specific workload is required.

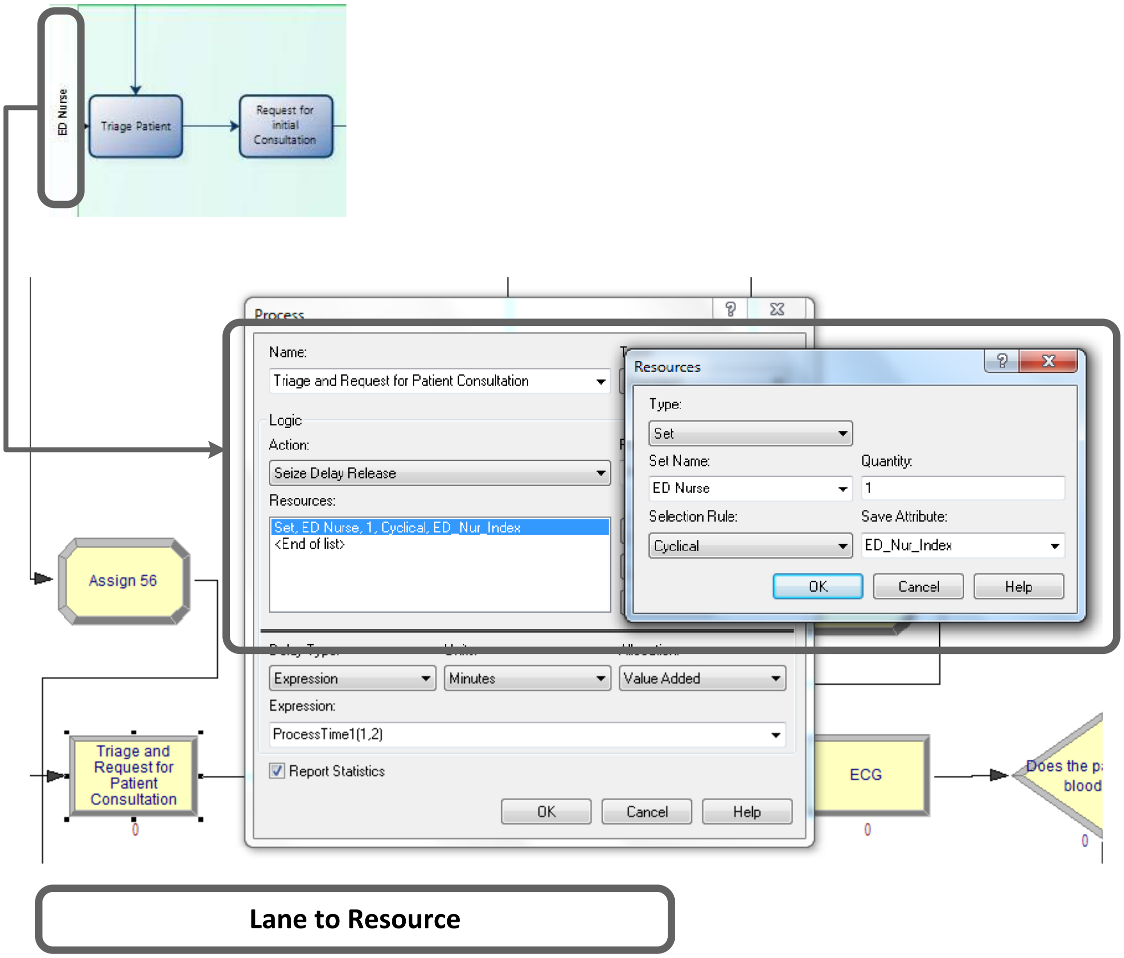

Figure 13 Conversion of Lane to Resource

Sequence Flow to Connector

In the intermediate model, the order of the activities and events is shown by means of sequence flows. In Arena, the sequence of the activities is declared by the use of connectors.

Table 7 Concept Conversion – Connectors

Business Process Model Concepts	Condition	Arena Simulation Modeling Concepts	Additional Information required?
Sequence flow	-	Connector	-

Gateway to Decide Module

Table 8 shows the two different types of gateways that were used in the intermediate model.

Table 8 Concept Conversion – Gateway vs. Decision Module

Business Process Model Concepts	**Condition**	**Arena Simulation Modeling Concepts**	**Additional Information required?**
Gateway	Exclusive Gateway	Decide Module	Yes
	Inclusive Gateway	Decide Module	Yes

In the intermediate model, sometimes the patient path has to branch through an exclusive or inclusive gateway. The inclusive gateway is used when the entity is allowed to choose one or all of the branches. An exclusive gateway is used when an entity can only choose one branch among all of the possible branches. For both types of gateways, the equivalent concept in Arena is the *decide* module. However, exclusive gateways are shown by one decide module, whereas several decide modules are used for mapping the inclusive gateways properly. The number of decide modules we use for the latter is equal to the number of branches of the inclusive gateway in the intermediate model. A set of conditions or probabilities are used as additional information to show the possibility of taking each branch.

Start Node to Create Module

A start node is used to trigger a process in the intermediate model. This node refers to the arrival of patients to the hospital. As long as there is an activity referring to the patient arrivals to the ED for which an equivalent concept of *create module* is used, the start node could be ignored and not be included in the simulation model. In our intermediate model, the specific activity related to the arrival of the patients to the hospital already existed, therefore we simply ignore the start node. An example of this conversion is shown in Figure 14.

End Node to Dispose Module

By the end of a process flow, the end node is used in the intermediate model to show the termination of the flow. This node is equivalent to the *dispose module* in Arena (Table 9), which discards the entities from the model after they are finished with the flow. An example of this type of conversion is also shown in Figure 14.

Table 9 Concept Conversion – Ending the Flow

Business Process Model Concepts	Condition	Arena Simulation Modeling Concepts	Additional Information required?
End Node	-	Dispose	No

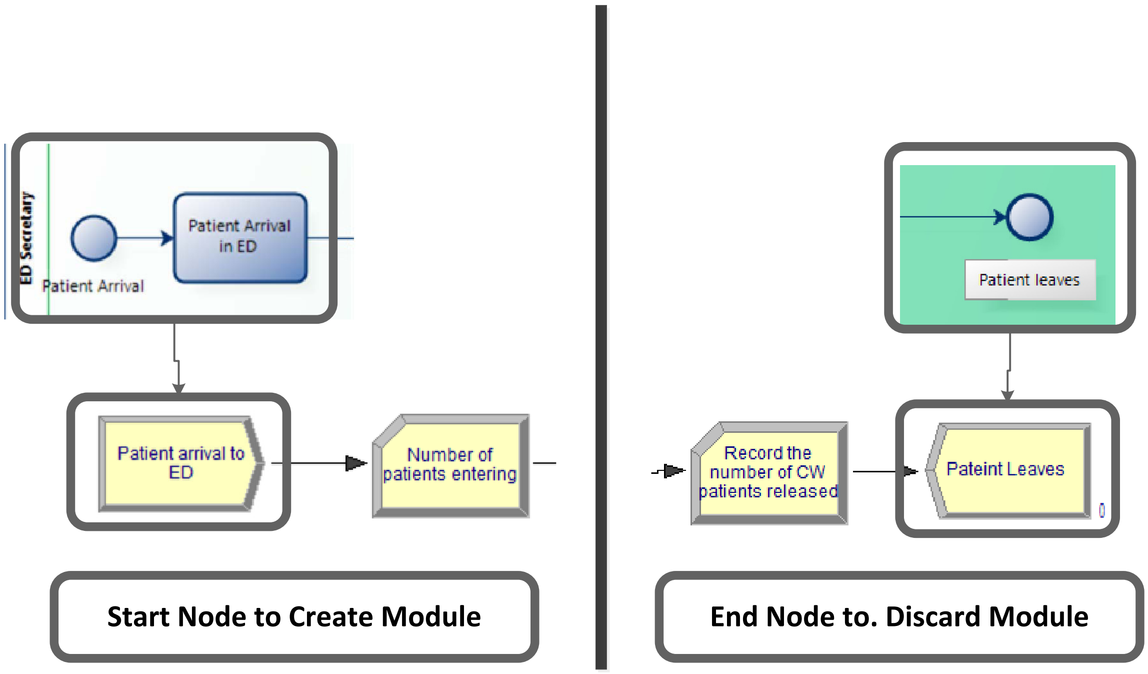

Figure 14 Start Node to Create Module - End Node to Discard Module

5.4.2 Additional Modules

Eventually, we added some modules to our model due to the simulation objectives:

- Since the patients were of different CTAS levels, and because we modeled such levels as an attribute for the entity, this has required us to use some probabilities related to the percentage of each priority by using a *decide* module or an *assign* module. Also we setup the queues to follow a specific discipline through which the patients are served according to their health condition priority. The patients with more severe conditions are placed in front of the activity queues to get served sooner.

- We used a *record* module to record the statistics we need to have at the end of the simulation runs.
- We also have used *seize* and *release* modules only for specific cases, for an entity to seize and release a resource. We have set one specific resource from a resource group to be assigned to a patient until the end of the patient's journey in hospital. For example, if the patient is being triaged by EDnurse1 among a group of ED nurses, then he/she will have his/her blood test or ECG done by EDnurse1. This is reflecting the procedure followed at Osler hospital.
- After interviewing an expert who was directly involved with the cardiac patient flow of Osler hospital, we had to add some other modules to the flow as well. We realized there were some activities that impacted the patient duration of stay in the hospital, like a delay in patient flow before receiving medical test results. We showed these activities through *delay* modules. Also, the activity "post procedure round" was broken down into a set of smaller activities for increased precision when assigning the tasks involved.
- The activities involved in patient clinical pathways are mostly human-related and the issue of capturing human-related behavior has been a concern in prior studies. However, in this work, we tried to mitigate its impact by introducing small randomized variations to the simulation model behaviour whenever appropriate. For example, while choosing the appropriate distribution for an activity such as performing the ECG, we have considered a small proportion of time added to each of the pure data values that was retrieved. This delay was taken into consideration to reflect the situations when no ECG device is idle to be used and the ED nurse unwillingly waits for one device to become idle. For realizing an important real-time simulation feature (the initialization from a predefined state), not supported out-of-the-box by Arena, a number of mod-

ules were added to the simulation model (e.g., see the Activating Real-Time Simulation part in Figure 15) that caused the size of the model to almost double.

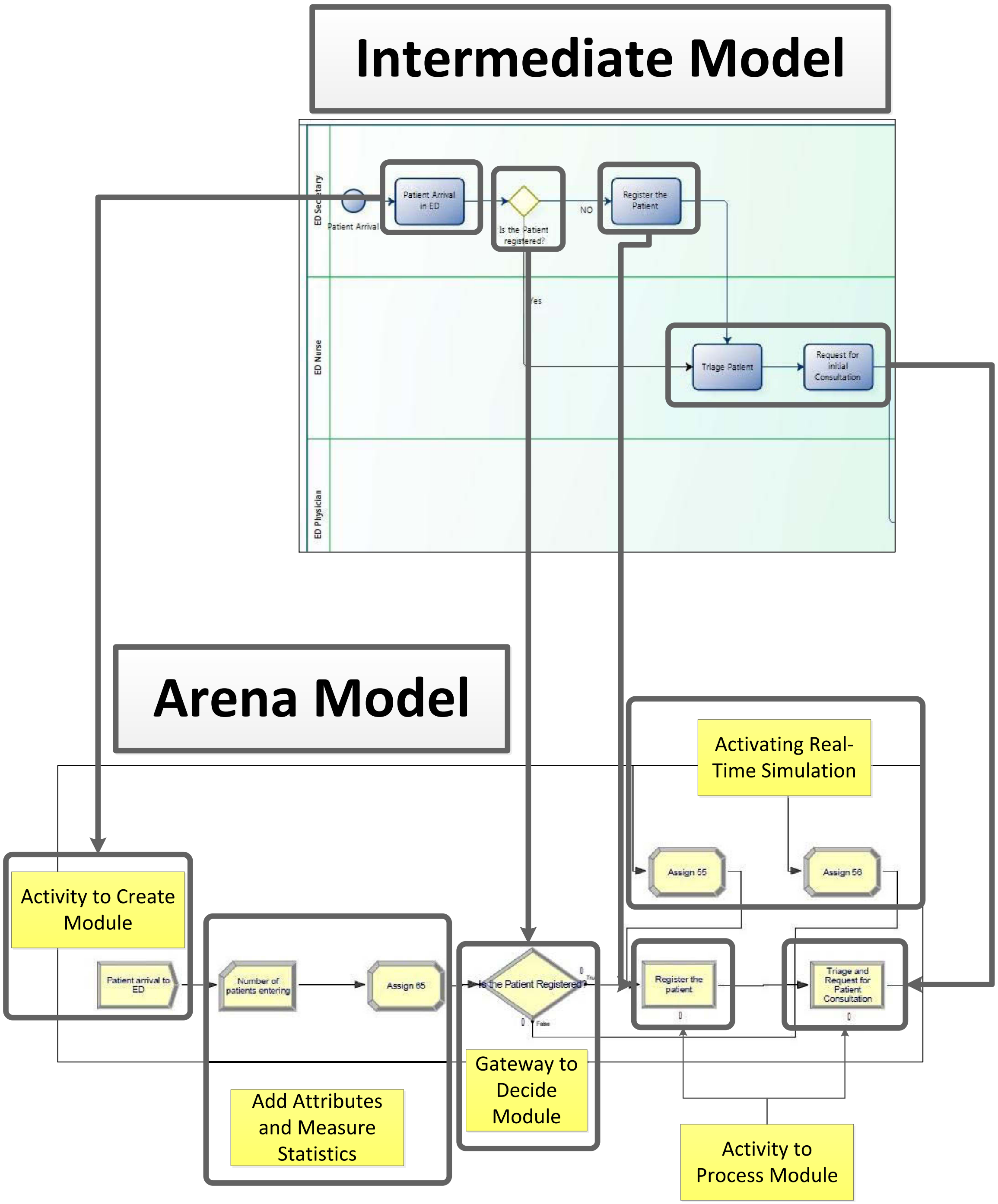

Figure 15 Mapping the Intermediate Model to an Arena Simulation Model.

- We eventually added a fully separate process to our simulation model for automating the number of replications for each scenario run. We will discuss the reason for adding this process later.

Figure 15 shows a sample example of how we used the conversion guidelines.

5.5. Observations about Mappings

Business process models are built for different purposes, such as:

- Process documentation
- Process reorganization
- Process monitoring and controlling
- Process performance improvement
- Workflow management

Based on different objectives expected from the model, the modeling styles and methodologies often differ. However, there is one common point among all different purposes: the flow will be looked at from a business point of view and the activities and events related to a business process model will be captured. There is much practical information regarding the *sequence* of the activities and events that is captured in a business process model and that could be reused for building a corresponding simulation model. Using the information already captured avoids starting to model from scratch. This approach has however one main drawback. That is, even if business process modelers use the same methodology for building models (e.g., BPMN), they may have different ways of modeling the same process and the styles might vary according to the points of view. This can limit the generality of the guidelines expressed in the previous section but this cannot deny the valuable information that can be extracted from the business process model for creating a simulation model.

5.6. Data Collection and Data Input

As discussed in the previous section, although the sequence of activities in the simulation model could be captured through mapping from the business process model, there is still some additional information that is required to be injected to the model to form a complete simulation model. This data is a necessary component for forming quantifiable parameters that can be measured and compared (e.g., probabilities, statistical distributions and resource schedules). In this study, external access to some of the input data of the simulation model was provided. An Excel file is populated with the required data and is *imported* in the simulation model. This enables the user to have access to the input data without changing the implemented model and to make changes externally if required. Eventually, the input data from the real-time business process monitoring system (see Figure 9) will be automatically injected into the external file (hence acting as an interface) in a timely manner so the simulation model can run with the real current data of the system. In this study, the focus was on trying to access and configure the data externally without making any direct change in the simulation model, and less engineering effort was invested in integrating the real-time monitoring system and the simulation model.

There are two types of data required to be injected into the simulation model:

- Historical data
- Initial state data (present state of the hospital)

5.6.1 Historical Data

The historical data was partially taken from the hospital current database and the rest was acquired by interviewing hospital employees. The data received from the database was the real data related to one month performance of the hospital (January 2013). The data collected through the interviews represented sufficient approximation of the real data according to the observed behaviour of the cardiac pathway at Osler hospital confirmed by the employees. Due to the confidentiality of the patient data, not all of the data could be acquired from the database. Also, sometimes the data was not entered properly by the clinical staff and was unavailable for use.

5.6.2 Initial State Data

The initial state of the hospital should consist in all of the information below in order to inject an accurate current state of the system:

- Number of patients in each queue related to each task.
- Number of patients involved with a certain task while the simulation starts.
- Starting state of each resource type (nurses, physicians, beds, etc.)

Several assumptions were made for feeding such data to the model:

- Before the initialization, resources are assumed to be idle. Work in progress is not initialized explicitly. Only the queues need to be initialized as the first thing the simulation will do when its starts is to allocate patients to resources.
- If a patient is involved with a specific task when the simulation gets started, the patient is assumed to be starting from the *beginning* of that task and not from in the middle of getting served.
- All patients who enter wait until they receive their complete service, and there is no patient Leave Without Being Seen (LWBS).

Therefore, only the number of patients in each queue was set to a specific number in the Excel file, whose structure and sample data are illustrated in Figure 16. For instance, if there is *task1* for which *resource1* has been assigned, Arena reads the number in the queue related to *task1*, checks for the availability of resource1 at the beginning of the simulation run and, if *resource1* is available, takes one of the entities in the queue inside the process module and makes the rest of the entities stay in the queue. If *resource1* is not available or is not being scheduled to *task1* at the beginning of the simulation run, all of the entities will stay in the queue and wait for *resource1* to become available. Thus, setting the number of patients in the queue is indirectly setting the state of the resources as well as the number of patients already being processed.

Resource schedules (e.g., for nurses and physicians) in terms of their capacity were also set to be read from an external file, in which the capacity of the resources can change per shift (4 hours) for a 24 hour period. Figure 17 shows the structure of this second Excel file filled with sample data.

	1	2	3	4	5	6	7	8	9	10
	Arrival Rate	Tirage	Registeration	initial con	ECG test	Blood test	Xray	Delay Lab results	diagnosis	admission to cw/seize bed
1	0	9	9	8	13	9	9	8	0	0
2	EXPO(0.2352)	UNIF(15,17)	UNIF(5,10)	TRIA(10,15,25)	UNIF(10,13)	UNIF(5,8)	UNIF(15,18)	UNIF(60,70)	UNIF(10,13)	

	11	12	13	14	15	16	17	18	19	20
	clean bed	transport fm ED to cw	Patient Stay in bed	transfer to OR	Operation Angiogram	Operation PCI	ccl observation	transfer to CW	Stay in bed day1	give medication 1
1	1	1	1	1	3	2	1	1	2	1
2	UNIF(0.66,6)	UNIF(5,7)	TRIA(12,14,16)	UNIF(5,7)	TRIA(40,45,50)	TRIA(30,45,60)	TRIA(15,30,45)	UNIF(5,7)	UNIF(240,300)	UNIF(5,10)

	21	22	23	24	25	26	27	28	29	30
	Stay in bed day 1	checkpatient status1	give medication2	stay in bed day 2	check patient status2	give medication 3	stay in bed 3	check health state 3	discharge process	wait for the bed
1	2	2	2	2	1	1	2	1	3	7
2	UNIF(240,300)	UNIF(3,4)	UNIF(5,10)	UNIF(240,300)	UNIF(3,4)	UNIF(5,10)	UNIF(240,300)	UNIF(3,4)	UNIF(40,50)	

Figure 16 Real-Time Simulation Initialization for Queues and Process Duration

ED admission Secretary

Time	Capacity
1st shift	1
2nd shift	1
3rd shift	2
4th shift	2
5th shift	1
6th shift	2

ED Nurse

Time	Capacity
1st shift	8
2nd shift	8
3rd shift	7
4th shift	6
5th shift	5
6th shift	4

ED Physician

Time	Capacity
1st shift	4
2nd shift	4
3rd shift	2
4th shift	2
5th shift	3
6th shift	4

Discharge Secretary

Time	Capacity
1st shift	2
2nd shift	2
3rd shift	3
4th shift	3
5th shift	1
6th shift	2

CW Secretary

Time	Capacity
1st shift	2
2nd shift	2
3rd shift	2
4th shift	2
5th shift	1
6th shift	2

CW Nurse

Time	Capacity
1st shift	6
2nd shift	6
3rd shift	4
4th shift	5
5th shift	3
6th shift	5

CCL Surgeon

Time	Capacity
1st shift	4
2nd shift	4
3rd shift	2
4th shift	4
5th shift	3
6th shift	2

Transporter

Time	Capacity
1st shift	2
2nd shift	2
3rd shift	2
4th shift	2
5th shift	1
6th shift	2

Beds

Time	Capacity
1st shift	32
2nd shift	32
3rd shift	30
4th shift	34
5th shift	25
6th shift	26

Figure 17 Real-time Simulation Initialization of Resource Scheduling

Real-time simulations, as discussed before, target the prediction of the system performance for the near future, in a timely manner. Changing resource schedules can be done for a 24 hour period in separate 4 hour shifts (aligned with the hospital's practice), so that the user has the option to set the current schedule of the resources externally (without changing the model) and check for the simulated near-future performance of the hospital.

Having these two files makes the Arena simulation much more flexible (through parameterization of the current state of queues and of resource scheduling) and amenable to handling real-time contexts. These files, together with the additional modules for state loading and the additional process for the automatic determination of the number of replications needed to achieve suitable precision (Section 5.4.2), represent *original and essential improvements* over typical usages of Arena.

5.7. Simulation Setup

As discussed in Section 4.2.6, the warm-up period is bypassed by loading the initial state of the simulation from Excel files. All of the initialization process takes place at time zero of the simulation run, preparing the model to start from a state that represents the actual present state of the hospital. Also, the run length and the replication number should be decided according to the nature of the system under observation and the desired precision level.

5.7.1 Run Length

The length of run chosen for this case study was 8 hours, i.e., two 4 hour shifts. Arena reports calculate the waiting times per activity after an entity is done performing the activity and leaves the related module. The reports might show a zero value for the waiting time of an activity for two reasons:

- Either the resources were perfectly scheduled for the task, resulting in no wait time, or
- Some patients have started getting served, but have not finished being served within the simulation run length, so there is no record of waiting.

Therefore, by choosing a run length longer than the maximum total duration of the physical activity (waiting time plus the value-added time of that activity due to the experience of the patients), we can ensure that if any patient is receiving a specific service, he/she will be done by the time that the simulation run length terminates and the related statistics will be recorded. The run length should exceed the maximum total duration of the activity that involves seizing a resource in the patient flow process.

In this case study, the maximum activity duration was 3.7 hours. Therefore, any run length to be chosen should exceed this number. At the same time, the objective of short-term decision making limits us from picking long run lengths (e.g., 4 days or 1 week). 8 hours was confirmed to be an appropriate time for setting a run length in this cardiac patient flow.

5.7.2 Number of Replications

The number of replications is the main factor that determines the precision of the output data in a real-time simulation (Section 4.2.6). In addition, the precision of the output data is the decision maker's choice. The higher the replication number is, the more precise the data will be, but the longer the execution time becomes. Two statistical tests have been implemented inside the simulation model to decide on the precision of the data. One test is the confidence interval of 95%. The output that Arena generates is by default within a confidence interval of 95% [21], meaning that in 95% of the repeated trials, the mean of a specific replication would be reported within the interval of the mean of all replications ± the half width. If the length of run is too short and the number of replications is insufficient or if the collected data is significantly dependent upon the other output means, then the reports would be missing some data since the data collection would not be adequately sufficient. After three replications (this is our minimum), this interval can be computed for the outputs.

The second implemented statistical test is the relative margin of error test. This value is calculated through the average relative MOE of the outputs. This value should not exceed 5% of the average MOE error. A specific condition has been set *in the simulation model* that checks for the average relative MOE after each replication and if the calculated statistics are sufficiently precise (with less than 5% error), then the replications will terminate. Basically, the number of the replications does not need to be set to a specific value. Instead, this number will be *adjusted automatically* according to the precision required for the output data. This is achieved in Arena by having an observer process (Figure 18) that monitors the MOE dynamically and decides whether to do another replication or stop the simulation

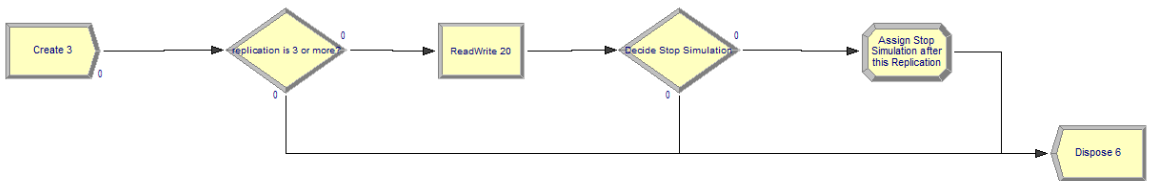

Figure 18 Arena Process for Monitoring the Required Number of Replications

A maximum number of replications has however been implemented (to constrain long durations as well as situations where the error level cannot be less than 5%) and set to 20. If any defined scenario reaches this number of replications, then there is a possibility of less precise output data for that scenario. Using this method, we can ensure that the desired precision of the data is realized through the minimum number of replications set with the lowest run time for scenarios.

5.8. Model Validation and Verification

Several methods were used to verify the simulation model. We first augmented the number of arriving patients and, as expected, the waiting time increased when the same resources were scheduled. We also made changes in the simulation initialization file in terms of the number of patients in the queues and the duration of the activities, and the expected behaviour was observed when running the simulation. Animation of the simulation model, where one can see the patients waiting at particular tasks (Figure 19), also proved to be useful for verification.

The data fed to this model is not directly from the real data (as the real-time patient flow management system described in Figure 9 is not yet in place), and approximate but realistic data is used. This research does not cover the actual implementation of the real-time simulation for the mentioned hospital. The model was only validated by one member of the clinical staff of the Osler hospital and it was evaluated to be reflecting the real cardiac pathway. The objective was to use a generic model from a real hospital and show how real-time simulations can be effective for making decisions in an environment dealing with waiting time issues. For truly validating the model, the real-time monitoring system should work in parallel with the real-time patient flow management system [2].

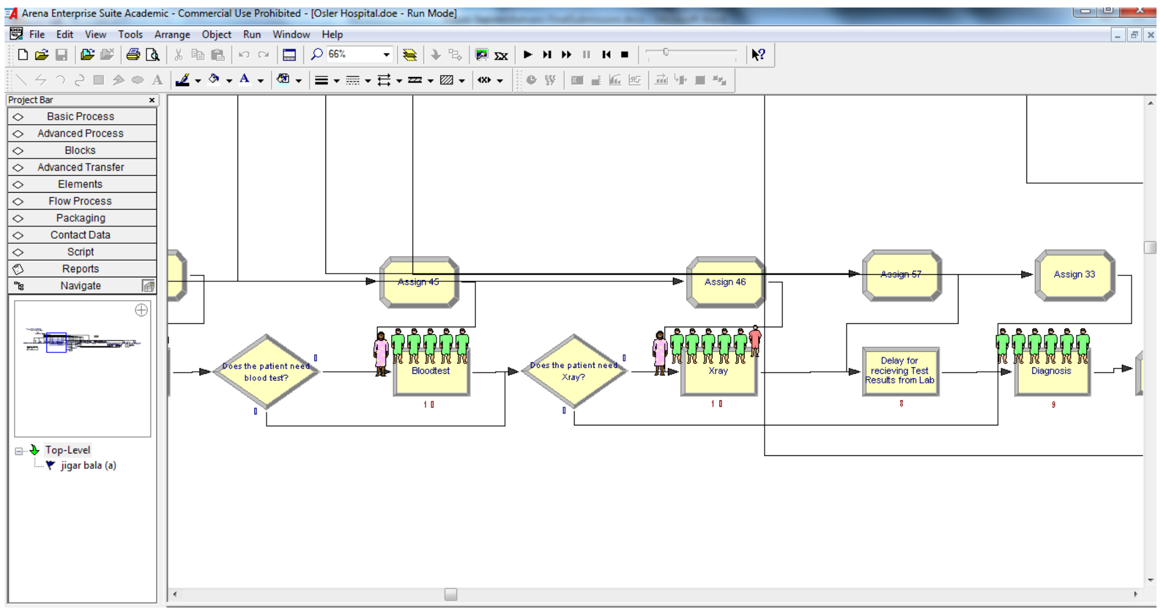

Figure 19 Animation of the Simulation Model for Verification in Arena

5.9. Summary

In this chapter, the creation of the simulation model and the detailed steps of this first phase have been discussed and illustrated. A problem was defined and a simulation model was developed through the use of a conceptual model. The conceptual model has been created by simplifying an executable business process model of the patient flow. The two types of input data (historical and initial state) have been collected and set into external files that feed the simulation model. Appropriate run length and replications were decided for the simulation model and finally the model has been verified and, to some extent, validated.

Beyond the methodology steps themselves, there are important and novel features demonstrated here:

- Simplification guidelines to go from an executable BPMN patient flow model to an intermediate conceptual model in BPMN.
- Construction guidelines to go from a conceptual model to an Arena model that can support real-time simulation (e.g., with additional modules for state loading).
- Design of a simulation infrastructure and of templates (Excel files) for loading a current state (including queues and resource scheduling) into the simulation model.
- Dynamic computation of the required number of replications to obtain a particular confidence interval and margin of error, which minimizes the simulation duration (something important to enable real-time decision making).

The next chapter covers the steps of the second phase of the methodology, which focuses on the execution and analysis of the simulation model.

Chapter 6. Execution of Simulation Model

With the first phase completed, the second phase involves the execution of the simulation model. What-if scenarios will be defined and the results from the scenarios will be analyzed in order for the best alternative to be chosen as a potential decision, leading to actions from the decision maker.

6.1. Running the Base-Case and What-if Scenarios

The simulation model is run and the alternative scenarios are defined. The creation of different candidate alternatives is subject to the configuration of resource allocation combinations. All of the scenarios proposed for realizing short-term decision support involve adjustments in resource scheduling.

There must be criteria against which the evaluation and comparison of the alternative results should be done. For this study, the main objective is the reduction of patient wait times, therefore, this parameter has been taken into account as one of the comparison criteria. In order to suggest more realistic alternative actions (as a reduction in wait times cannot be achieved at any cost), other criteria were also selected. These performance metrics are relevant as they will be influenced by the patient waiting times. Alternatives are compared against these defined criteria and ultimately the best one will be proposed as the most effective and feasible one.

6.1.1 Defining Criteria

The criteria should be set considering all the important factors that impact decision making. These criteria should be approved by the decision makers themselves. For this study, criteria were chosen for meeting the main objective of reduction of patient wait time as well as two other main factors related to *budget issues* for staff scheduling. These measurable criteria should be calculated for the base-case scenario and the alternative scenarios, and eventually compared.

Average Patient Waiting Time

There are several definitions that should be clarified in this study:

- *Length of stay*: The total time that a patient spends from the moment it enters the flow until the moment it is discarded from the flow.
- *Ideal length of stay*: The total time that the entity travels an entire flow, going through all the activities and events in the flow, assuming that this entity is the only one in the whole flow, and there is no bottleneck.
- *Average patient waiting time*: This value corresponds to the waiting time that a patient will typically experience during the entire process of the patient flow. It can be calculated in two possible ways:
    - Find the waiting time for each of the patients, for each of the activities, and average this quantity over the total number of patients who went through all the process.
    - Find the average waiting time per patient for each activity and then sum these quantities to find the total patient wait time.

In a short-term decision making process, the intention is to be able to run short-term simulations or real-time simulations and decide in view of near-future predictions. Therefore the run length should be chosen for a short period in the future. However, if the run length is less than or equal to the ideal length of stay, then no entity will traverse the entire flow during one run length. Moreover, to be able to calculate the average waiting time from the first approach introduced above, one must have waiting time information for all the entities that traversed the process and average these values.

In this case study, the run length has been chosen to be 8 hours, which is also smaller than the ideal length of stay of 3 days. Therefore, no entity will completely finish the entire cardiac pathway in one simulation run. In order to calculate the average waiting time through the first proposed calculation method, the entities that are initialized at the beginning of the simulation run should contain historical data related to their waiting time experience in the previous activities. This would allow having the waiting time information of the patients who get discharged in the 8 hour period and to be able to calculate the average value. However, there is one drawback to this approach. The number of patients who go through all the process and get discharged from CW in a short period of time (in our case 8 hours) is low, and therefore the average waiting time will be calculated based on a *few* patients, which will require a larger number of replications in order to become reliable enough to base our decisions on.

Hence, the second method for calculating patient waiting time has been used. In this method, the average waiting time for each activity is first calculated, and then these values are summed up to calculate the total waiting time per patient. If all of the entities go through an entire flow, then both of the calculation methods proposed before should show the same result. This means that the average of the sums is equal to the sum of the averages, and this can be demonstrated mathematically.

Let us assume that there is a matrix defined with X_{nm} as the elements indicating the waiting time of patient n for activity m, with P_1 to P_n referring to the different patients from 1 to n.

$$\begin{cases} P1 \\ P2 \\ Pn \end{cases} \begin{bmatrix} X11 & x12\cdots & X1m \\ \vdots & \ddots & \vdots \\ Xn1 & \cdots & Xnm \end{bmatrix}$$

As discussed previously, one way to find the average waiting time is to sum up all of the elements in each row (waiting time of each patient spent during an entire patient flow) and then divide the sum of those values with n, which is the number of patients.

$$P1\text{: } Total\ wait\ time = X11 + X12 + \cdots . + X1m$$

$$P2\text{: } Total\ wait\ time = X21 + X22 + \cdots + X2m$$

$$\cdots.$$

$$Pn\text{: } Total\ wait\ time = Xn1 + Xn2 + \cdots . + Xnm$$

Average waiting time = $\frac{Total\ wait\ time\ P1 + Total\ wait\ time\ P2 + \cdots \ldots + Total\ wait\ time\ Pn}{n}$

$$= \frac{X11 + X12 + \cdots + X1m + X21 + X22 + \cdots + X2m + Xn1 + \cdots + Xnm}{n}$$

Another method is to sum up each column values (the waiting times per activity) and divide the value from each column by the number of patients (n) and finally add all the values.

$$Actvity\ 1: Average\ wait\ time = \frac{X11 + X21 + \cdots + Xn1}{n}$$

$$Activity\ 2: Average\ wait\ time = \frac{X21 + X22 + \cdots + Xn2}{n}$$

$$\ldots\ldots.$$

$$Activity\ m: Average\ wait\ time = \frac{X1m + X2m + \cdots + Xnm}{n}$$

Average waiting time $= Average\ wait\ time\ Activity\ 1 +$

$$Average\ wait\ time\ Activity\ 2 +$$

$$\ldots$$

$$Average\ wait\ time\ Activity\ m$$

$$= \frac{X11 + X12 + \cdots + X1m + X21 + X22 + \cdots + X2m + Xn1 + \cdots + Xnm}{n}$$

Mathematically, these two results show the same value due to the calculations above. However, if among the elements of this matrix, any of the values *is missing*, then the quotients will differ. For instance, if one of the elements in the first column cannot be calculated, then the average of the values will be over *n-1* patients rather than *n* patients. If there is any waiting time value per patient per activity that cannot be calculated in this matrix, then the sum of the averages may not be equal to the average of the sums. Therefore, the two methods of calculating the waiting time would have different results.

The first method of calculation of the waiting times was not applicable to our case due to the time restrictions explained above. Thus, the second method of calculation was chosen for calculating the waiting time. However, this value does not present the total average patient waiting time as its definition represents. Instead it sums up the average waiting time of all of the activities per patient.

Total Cost

The total cost is an important factor for making decisions for staff scheduling in a hospital. The patient waiting time may be decreased with a proper resource scheduling and appropriate allocation of clinical staff. But there is always a limitation to the amount of supply that can be provided for serving the patients. Although, reduction of patient waiting time is the ultimate objective of this case study, expenses and budget limitations were also chosen to be a criterion in the decision making process.

We are categorizing the payment regulations for all types of resources into two groups. For some of the resources (beds, physicians, etc.) the cost would be per each resource use. As for the rest of the resources (ED nurse, clerks, etc.), they are assumed to get paid on an hourly basis. The salaries have been approximated and are not from the real hospital data.

Percentage of Patients Discharged

Finally, for the budget issues to become more meaningful, the last criterion chosen is the percentage of the patients who get discharged. This value can *indirectly* represent the prediction of hospital revenues in a short-term future period. The profit that the hospital makes is calculated through subtraction of the costs from the revenues. The more patients are discharged in a given period of time, the higher the revenues of the hospital. The scheduling of more staff might however result in higher expenses. A decision maker should consider the revenues that the hospital gains in a short-term period, and check whether these revenues outweigh the costs spent for resource scheduling. The calculation for this criterion was done through the following formula:

$$\text{Percentage of Discharged Patients} = \frac{Number\ of\ Patients\ Discharged}{Number\ of\ Patients\ Arrived + Number\ of\ Existing\ Patients}$$

6.1.2 Base-Case Simulation Run and Results

The base-case scenario is the initial case that should be observed. The present performance of the hospital was set into the simulation model, and the impact of current operation of the hospital as it was, on the 8 hour future time frame, was observed. Evaluating alternative scenarios should be done when the base-case leads to unsatisfactory performance results. These what-if scenarios should be compared with the base-case scenario. Some predefined and customized statistical reports are generated using the Arena simulation tool. Although the detailed calculations of each of the performance criteria is not necessary for defining scenarios and only the output results are of concern, we will briefly discuss the detailed results from which the output data was calculated.

Table 10 shows the final performance metrics for the base-case scenario. The average wait time is in hours, while the total resource costs are in dollars.

Table 10 Final Average Values of the Outputs

Output	Average	Half Width	Minimum Average	Maximum Average
Average Waiting Time of Activities per Patient	7.7793	0.10	7.6279	7.9519
Percentage Discharged	56.6106	4.64	47.1154	64.4231
Total Resource Costs	17788.16	128.64	17537.00	17968.97

The average waiting time of the patients per activity will be 7.77 hours in the next two shifts (8 hours). The average waiting time per each individual activity per patient is given in Table 11, where the activities that are mainly causing the waiting times are i) waiting for bed, ii) delays in receiving the lab test results, iii) diagnosis, and iv) ECG.

Table 11 Average Waiting Time of Patients per Each Activity

Wait Time Per Entity	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
Bloodtest	0.3807	0.07	0.2876	0.5293	0.00	2.8470
Check for patient Health state day 2	0.1468	0.03	0.1114	0.2320	0.00	0.8042
Check for patient Health state day 3	0.2069	0.07	0.1603	0.4204	0.00	0.8385
Check patient health state day 1	0.3305	0.03	0.3076	0.4214	0.00	0.8632
Delay for recieving Test Results from Lab	1.0834	0.01	1.0735	1.0958	1.0005	1.1661
Diagnosis	1.0624	0.10	0.8795	1.2201	0.00	4.3743
ECG	0.5926	0.05	0.5324	0.6908	0.00	2.5535
give medication to patient day 2	0.00	0.00	0.00	0.00	0.00	0.00
give medication to patient in 3rd day	0.00	0.00	0.00	0.00	0.00	0.00
Give medication to the patient day 1	0.00	0.00	0.00	0.00	0.00	0.00
Initial Consultation	0.4184	0.12	0.3133	0.7604	0.00	3.6523
Operation Angiogram	0.2702	0.03	0.2197	0.3369	0.00	1.5232
Operation PCI	0.3330	0.07	0.2279	0.4827	0.00	1.5056
Patients discharge process	0.4267	0.09	0.2725	0.6095	0.00	1.8022
Registeration	0.3404	0.02	0.3089	0.3696	0.00	1.1853
Transfer Patient from CCL to CW in bed	0.00141016	0.00	0.00	0.00333495	0.00	0.04773395
Transfer to Operation room in CCL	0.00	0.00	0.00	0.00	0.00	0.00
Transport the patient from ED to CW	0.01885156	0.00	0.01496526	0.02127492	0.00	0.1141
Triage and Request for Patient Consultation	0.2086	0.02	0.1779	0.2525	0.00	0.7077
Xray	0.3484	0.06	0.2837	0.4901	0.00	2.5798
wait for bed	3.8277	0.40	3.2279	4.4273	0.00	6.6318

The percentage of the patients discharged from ED and CW is approximately 57% (i.e., (50.75 + 8.125)/103). Individual discharge counters were set for ED and CW, and Table 12 shows their results.

Table 12 Number of Patients Discharged from ED and CW

	Average	Half Width	Minimum Average	Maximum Average
Number In Patient	103.00	0.00	103.00	103.00
Number of ED patients discharged	50.7500	4.51	41.0000	58.0000
Record the number of CW patients released	8.1250	0.94	6.0000	9.0000

There are different types of costs associated with resource usage. 0 shows the detailed costs associated with different resources.

After the base-case scenario is set and the criteria have been evaluated for this case, the next step is to figure out different possible alternative scenarios for resource allocation in order to compare the candidate alternatives against the base-case.

6.1.3 "What-If" Scenario Runs and Results

What-if scenarios are different alternative actions that could be taken for resource scheduling within the next few hours, in the near future. The alternative scenarios are run to predict possible outcomes of the hospital performance under certain circumstances. The trade-offs between the different business outcomes will be assessed and the best option will be chosen among the possible alternatives (including status quo represented by the base-case). Arena provides two external applications for supporting the decision-making process: Process Analyzer (PAN) and OptQuest.

Defining "What-if" Scenarios using PAN

The Process Analyzer is an external application in Arena that is used to manually set and run what-if scenarios to be compared by their outcome results.

Table 13 Total Resource Costs

Busy Cost	Average	Half Width	Minimum Average	Maximum Average
Beds	0.00	0.00	0.00	0.00
CCL Surgeon 1	0.00	0.00	0.00	0.00
CW admission Secretary	15.7236	1.40	12.5893	17.8380
CW Nurse 1	159.49	10.82	137.53	181.15
Discharge secretary	125.09	10.35	114.20	149.73
ED admission Secretary	29.7059	0.86	27.9615	31.1170
ED Nurse 1	1123.24	25.11	1086.10	1183.12
ED Physician 1	0.00	0.00	0.00	0.00
Transporter	0.00	0.00	0.00	0.00

Idle Cost	Average	Half Width	Minimum Average	Maximum Average
Beds	0.00	0.00	0.00	0.00
CCL Surgeon 1	0.00	0.00	0.00	0.00
CW admission Secretary	208.28	1.40	206.16	211.41
CW Nurse 1	1759.00	10.08	1738.85	1779.45
Discharge secretary	97.4666	11.95	74.2681	109.80
ED admission Secretary	82.2941	0.86	80.8830	84.0385
ED Nurse 1	1434.25	24.93	1376.16	1467.57
ED Physician 1	0.00	0.00	0.00	0.00
Transporter	0.00	0.00	0.00	0.00

Usage Cost	Average	Half Width	Minimum Average	Maximum Average
Beds	401.25	9.42	380.00	410.00
CCL Surgeon 1	4175.00	153.20	3900.00	4400.00
CW admission Secretary	0.00	0.00	0.00	0.00
CW Nurse 1	0.00	0.00	0.00	0.00
Discharge secretary	0.00	0.00	0.00	0.00
ED admission Secretary	0.00	0.00	0.00	0.00
ED Nurse 1	0.00	0.00	0.00	0.00
ED Physician 1	7959.38	52.33	7875.00	8025.00
Transporter	218.00	16.68	192.00	248.00

We have chosen different resource capacities for different types of resources that were available (e.g., ED nurses and beds) and we checked the impact of changing their capacity on the performance metrics set previously. The capacities are changeable within each shift. Therefore, since our prediction horizon has been set to 8 hours (equal to two shifts), we can have two capacity values (first and the second shifts) set for each resource type.

Table 14 What-if Scenario Run Results

	Scenario Properties			Controls																		Responses		
	S	Name	Reps	ED sec_Ca	ED sec_Cap	ED_Nu_Cap(1)	ED_Nu_Cap(2	ED_Ph_Cap(1)	ED_Ph_Cap(2)	CW_Sec_Sch(	CW_Sec_Sch(2)	CW_Nu_Cap(1)	CW_Nu_Cap(2)	CCL_Su_Cap(1)	CCL_Su_Cap(2)	Sec_Dis_Cap(1)	Sec_Dis_Cap(2)	Transport_cap(1	Transport_cap(2	Bed_Capacity	Bed_Capacity(	Average Waiting Time	Total Resource Costs	Percentage Discharged
1		Base Case	8	1	1	8	8	4	4	2	2	6	6	4	4	2	2	2	2	32	32	7.779	17788.158	56.611
2		Add ED Nurse	6	1	1	9	9	4	4	2	2	6	6	4	4	2	2	2	2	32	32	7.584	18231.724	56.891
3		Add 1 ED Physician	6	1	1	8	8	5	5	2	2	6	6	4	4	2	2	2	2	32	32	7.365	17948.243	56.891
4		Add 2 ED Physicians	8	1	1	8	8	6	6	2	2	6	6	4	4	2	2	2	2	32	32	7.050	17817.083	53.726
5		Add ED Staff	4	1	1	9	9	6	6	2	2	6	6	4	4	2	2	2	2	32	32	6.860	18186.993	55.288
6		Remove 1 CW Nurse	8	1	1	8	8	4	4	2	2	5	5	4	4	2	2	2	2	32	32	7.779	17468.158	56.611
7		Add 2 CCL surgeon	7	1	1	8	8	4	4	2	2	5	6	6	6	2	2	2	2	32	32	6.259	17574.066	54.808
8		Open 5 more beds	12	1	1	8	8	4	4	2	2	6	6	4	4	4	2	2	2	37	37	7.508	19004.282	57.933
9		Add Clerks and Secretaries	6	2	2	8	8	4	4	2	2	6	6	4	4	3	3	2	2	32	32	7.575	18208.238	55.128
10		Add 1 ED Phy & 1 CCL Sur	7	1	1	8	8	5	5	2	2	6	6	5	5	2	2	2	2	32	32	6.835	17677.122	57.830
11		Add 2 ED Physicians & 2 CCL Su	6	1	1	8	8	6	6	2	2	6	6	6	6	2	2	2	2	32	32	5.634	17843.356	57.051
12		Add 1 ED phy & 1 CCL & 8 beds	8	1	1	8	8	5	5	2	2	6	6	5	5	2	2	2	2	40	40	6.170	19172.416	57.332
13		Change in ED & CW staff	8	1	1	9	8	6	4	2	2	5	6	6	4	3	2	2	2	35	35	4.743	18663.200	56.731

Interpretation of Some of the Scenario Results

Table 14 shows the results of various what-if scenarios that we have defined. Different numbers of replications were automatically assigned to the scenarios depending on the precision of the output data needed to meet the desirable error level of less than 5% on average.

The first scenario is the *base-case* scenario, which is the “no-action” scenario. In the third scenario, *2 ED physicians have been added*, and the waiting time has decreased by approximately 0.7 hours (42 minutes). However, this configuration has caused the percentage of discharged patients to be lower. Even though the ED tasks related to the ED physicians will lead to smaller waiting times, the medical tests (ECG, blood test and X-ray) done by the ED nurse will cause a delay in the discharge of the patients from ED. Scenario 6 shows that the *removal of a CW nurse* will not cause any change in terms of the waiting time and will decrease the cost of the resources while keeping the number of patients discharged at the same level (compared to the base-case). Scenario 10 shows that by *adding one more physicians and one CCL surgeon*, the waiting time will be decreased by almost 1 hour while causing the total resource costs to decrease and the percentage of discharged patients to rise by almost 1%. This configuration has caused the idle cost of the nurses to decrease, since they will be busier. Eventually the last scenario shows the possibility of a configuration that can cause the waiting time to be 4.7 hours at the cost of almost 850$ more than the base-case scenario.

Advantage of Defining Scenarios through a Holistic Point of View

The main goal while defining the scenarios was to reduce the patient waiting time by adjusting the number of resources. The defined scenarios can be run very quickly (every replication takes about 0.6 second), which enables the decision maker to evaluate and compare many alternative scenarios in advance. The decisions are made through a more *holistic view* of the different hospital departments involved in the care process. Making decisions at a holistic level prioritizes the overall performance improvement and prevents one-dimensional local optimizations. For instance, in this case study, the waiting time for beds is approximately 3.8 hours, which in theory is the time that patients spend in ED before getting admitted to CW. Opening more beds might remove the bottleneck from this activity, but causes the bottlenecks to be pushed to the activities in CW and eventually causing little change in the overall waiting time. All of the addressed issues within the scope of the defined problem could be managed through analysis of the predicted output of the scenarios.

The output criteria should be the important factors upon which the decision maker needs to evaluate the scenarios and decide. These criteria could be as numerous as the decision maker requires and feels comfortable with while making decisions. However, the process of decision making may take more time if a large number of criteria are taken into consideration. In this case, using OptQuest would be more appropriate and beneficial.

Optimizing the Simulation Results using OptQuest

In case of large number of criteria for assessing the scenario results, an automated search approach can be used via OptQuest, an external application in Arena. This optimizer uses meta-heuristic methods to find the best solution according to an objective function and a set of constraints defined prior to optimization runs. An objective function can be, for example, a minimization of waiting times. Constraints can be the limitations faced in terms of the resource capacity available. This optimizer searches through possible alternatives and considers different combinations of resource allocation automatically until a terminating condition occurs. The more runs are executed, the more alternative scenarios can be evaluated. However, time is an important factor in running the scenarios and shall not be ignored in the process of short-term decision making.

Table 15 shows the result of the search with OptQuest. The scenarios are listed with the best scenario at the top, i.e., the one with the most desirable value for the objective function that satisfies all the constraints (i.e., that is *feasible*).

The advantage of using the optimizer is that the number of possible scenarios to be checked is higher than when the scenarios are manually set into PAN. There is also the possibility of coming up with a solution that is feasible and affordable that normally might not occur to mind. The drawback is the time it takes for the optimizer to run and provide the results. Also, using the optimizer would not allow the user to check a specific candidate scenario.

Table 15 OptQuest-Optimizing Output Results

Simulation	Objective Value	Status	Bed	Bed	CCL_	CCL	CW_N	CW_Nu	CW_S	CW_Se	ED sec	ED se	ED_Nu_	ED_N	ED_Ph_C	ED_Ph_	Trans	Transp	Sec_Dis
57	5.665979	Feasible	40	25	6	6	6	3	3	3	3	1	9	4	6	3	3	3	1
98	5.845903	Feasible	35	33	6	6	5	3	2	3	3	2	8	6	5	5	2	2	2
93	5.897286	Feasible	35	34	6	6	5	3	2	3	3	2	8	5	5	5	2	2	2
92	6.775626	Feasible	35	35	5	6	5	4	2	3	3	2	8	5	5	4	2	2	2
42	6.851226	Feasible	40	37	6	4	3	5	3	2	3	2	7	7	3	6	3	1	1
80	8.408361	Feasible	30	35	5	5	3	4	1	1	2	2	7	9	3	3	1	1	2
43	6.809885	Infeasible	40	37	6	5	3	5	3	2	3	2	7	7	3	6	3	1	1
65	7.599898	Infeasible	25	25	3	4	3	3	1	1	1	1	5	4	3	3	1	1	1
69	7.590986	Infeasible	25	25	3	4	3	4	1	1	1	1	5	4	3	3	1	1	1
95	6.537264	Infeasible	36	31	5	5	5	5	3	2	2	2	8	7	5	4	3	2	2
10	7.007001	Infeasible	40	37	6	3	3	5	3	2	3	2	7	7	3	6	3	1	1
100	6.969115	Infeasible	34	35	6	6	4	3	2	3	3	2	8	5	4	5	1	2	1
8	8.582718	Infeasible	29	29	4	4	4	4	2	2	2	2	5	5	4	4	2	2	2

6.2. Output Analysis, Making Decisions and Taking Actions

From the list of feasible scenarios, the two or three best ones are presented to the decision maker. The decision maker should make the final decision considering the trade-offs between the best scenario outcomes. The chosen scenario is the scenario that leads to a reduction in the total patient waiting time, with a reasonable resource cost that will be compensated with a larger percentage of the patients discharged from the hospital.

6.3. Summary

The execution of the real-time simulation model of the cardiac pathway at Osler hospital was discussed, and alternative scenarios were defined and evaluated. We have specifically focused on a holistic optimization approach rather than a local optimization by considering the defined criteria calculated for all of the units involved in the patient care process. Using this approach, the average waiting time of activities per patient is minimized without pushing the bottlenecks to other units of the hospital.

PAN will give the decision maker an opportunity to develop specific scenarios and evaluate their results compared to the base-case scenario. This can be done quickly. One of the disadvantages of using PAN is that the number of scenarios that can manually be set is limited to the user's selection. OptQuest, however, gives the user a wider variety of options by automatically searching through the possible feasible scenarios. However, OptQuest's main drawback is the time required for setting and running all the scenarios using an optimization mathematical model, and the fact that the user will not be able to evaluate a specific scenario he/she has in mind. Using either of the approaches for finding the best scenarios, the effects of implementing a number of alternatives is evaluated and the user is allowed to think about the consequences of alternative actions prior to occurrence of bottlenecks and make the right decision in advance.

Note also that PAN and OptQuest are however not mutually exclusive. Several specific scenarios could be investigated with PAN while OptQuest could be used in a complementary way to see whether adding feasible scenarios should be considered.

The next section will discuss the overall evaluation of this research.

Chapter 7. Comparison

Chapter 3 highlighted two existing approaches that aimed to support real-time simulations to enable short-term decision making: Marmor et al. [29] and Rozinat et al. [39]. The following sections present a detailed discussion of these related approaches and compare their strengths and the weaknesses with the one developed in this thesis. General threats to the validity of this research are also discussed.

7.1. Objectives and Scopes

Marmor et al. [29] focus on online prediction of near-future states with the objective of adjusting resource scheduling in emergency departments for 6 types of patients in the ED of a hospital. They mentioned that after the patients are done with their care process in ED, some need to be transferred to wards. However, most of the times the wards are full and patients end up waiting in the ED area. This waiting time was not recorded and it also caused further issues like congestion in ED that could not be captured in the model. The approach is known as real-time simulation because the simulation model is fed with real-time arrival data of the patients to the ED, and this information is taken from the hospital electronic database with every patient registration.

Rozinat et al. [39] integrate a workflow management system and a simulation system to enable operational decision making, rather than long-term strategic decision making. A credit card application process has been modeled in this study. The consequences of making changes in the resource scheduling have been checked and the throughput time of the credit card application was calculated. The study does not have a specific focus on the healthcare context where human-related activities are much more complex.

In our study, the real-time simulation model is within the context of healthcare and unlike both of the above studies, the scope of the defined problem is on interrelated departments of a hospital for one type of patients (cardiac patients) with 5 priority levels in terms of how severe their health problems are. An entire clinical pathway was chosen for this study from the time the patient enters the ED to the moment he/she is moved to the CW and finally discharged. This broad scope of the study prevents some inaccuracies found in Marmor et al. Moreover, the credit card process that was modeled by Rozinat et al. is not an exact real-life process, whereas in this study we have used the real-life clinical pathway of cardiac patients of a community hospital.

7.2. Warm-Up in Real-Time Simulations

In the study done by Marmor et al. [29], the initial state of the simulation is the state *after* the simulation run passes the transient state. In order to attempt to overcome the warm-up period, this study feeds the simulation model with real-time arrival data of the patients and runs the simulation model for a long period of time (almost continuously).

Rozinat et al. [39] use a Petri-net based simulation model. They inject the "current state" information reported from a workflow management system to overcome the warm-up period. The event logs and the simulation logs are in the same format.

Marmor et al. suggest the normal overcoming approach for the warm-up period for their real-time simulation model. The warm-up period differs from run to run depending on the setting of the model and the data input. This period is required to be calculated before each run unless the simulation is running automatically, non-stop. Another drawback is that, although the purpose is to be able to make short-term predictions, the simulation model should run for a sufficiently long period of time to be reflecting the current state; then the data collected will be reliable. Rozinat et al. bypass the warm-up period by starting the simulation from a predefined state that is basically the current state of the process in reality.

In our approach, with the objective of enabling short-term decision making, we could not afford the existence of a warm-up period for every simulation run. We overcame the warm-up period by having a simulation model whose system's current state can be loaded dynamically. For implementing real-time simulation features, Rozinat et al. have developed their own language and finally have used Petri-nets for modeling their simulation with the CPN tool[17]. We have used discrete event simulations using Arena

[17] http://cpntools.org/

v14, which did not support loading the initial state capabilities out of the box, and we came up with a novel approach for loading the initial state to our model.

7.3. Run Setup and What-If Scenarios

Both of the studies done by Rozinat et al. [39] and Marmor et al. [29] highlight the importance of short-term decision making, which require short-term predictions of near-future situations. However, Marmor et al. must first run the simulation for a long enough period of time to overcome the warm-up period, which is contradicting the objective of short-term decision making. They used 100 replications for running their simulation model. They have also not mentioned any specific statistical tests for calculating the precision of the output data. Marmor et al. use two mathematical resource allocation strategies and their results have been set as different possible (fixed) scenarios. Rozinat et al. defined four simulation scenarios and run them for two weeks into the future, with 50 replications each. The results were observed with a confidence level of 95%, while initially keeping the same number of resources and afterwards adding more resources.

We have used the terminology "short-term" in a different manner as opposed to the ones described in the above studies. Short-term predictions should occur for a short period of time in the future depending on the type of systems simulated. For instance, in our case (cardiac pathway), the short-term prediction could start from 4 hours minimum and 1 day maximum to ensure the possibility of adjusting resource scheduling in the near future. In healthcare environments, the dynamic nature of the activities and events involved would continuously change the state of the hospital, and for a long period of time in the future, the output data would not be useful anymore. Moreover, in our study, there is an emphasis on executing the real-time simulation on a cyclic basis (e.g., every 8 hours) for maximizing the objective of the study (decrease patient waiting time).

Rozinat et al. mentioned they have defined the 95% confidence interval for specifying the variation in their output results, but without discussing the margin of error. In our study, we could not afford long simulation runs caused by a large number of replications. At the same time, the precision of the output data was of concern as well. Apart from the confidence level of 95% calculated for the output data, we implemented an automated output precision calculator, which terminates the number of replications if the average relative MOE of the outputs is less than 5%. With this level of precision, between 4 and 12 replications were required for 13 different scenarios that we defined, with every replication taking around 0.6 seconds. Using this approach, we can make sure that we have sufficient precision with the minimum number of replications required to make the process of running scenarios as quick as possible.

7.4. Validation of the Model

Marmor et al. [29] validated their simulation output data against the hospital database, comparing the real performance of the hospital with the performance predicted by the simulation model in a "no-action" scenario. Rozinat et al. [29] could not get a real-life process to compare their simulation model against.

In our study, the model was confirmed by an expert from the hospital to be reflecting the hospital performance, but a detailed validation approach was not taken due to privacy-related limitations in accessing the patient information database. Such validation could however be done inside a hospital environment.

7.5. Discussion

Table 16 specifies the comparison of our work with the two most related work as an overview. We have taken a clinical process for implementing real-time simulation model. Marmor et al. [29] also use a clinical process however, they do not take a complete end-to end clinical process. Moreover, their process is limited to the boundaries of an ED. In our study as well as the study done by Rozinat et al. [39], the initial state of the system is being loaded prior to each simulation run so that the warm-up period is bypassed. The period of 8 hours in future has been considered as a short-term objective in the study conducted by Marmor et al. [29], whereas a period of two weeks is known as a short-term for the credit card process under observation in the research done by Rozinat et al. [39]. Each of the two related studies perform 100 and 50 replications for each scenario which definitely takes much more time rather than when the replications are automated to satisfy a specific precision. The impact of limited number of scenarios is checked as what-if scenarios in both of the related studies. We have performed 13 scenarios and our approach is more flexible to perform as many scenarios as required. None of the three studies has actually implemented the what-if scenarios in the real-world process. Both our simulation model and the one from Marmor et al. [29] were validated to be sufficiently reflecting the real-world process.

Table 16 Comparison with Related Work

		Clinical Process	Entire Process	Load Initial State	Bypass Warm-up	Run length	Replication Number	Fixed Scenarios	Implement scenarios	Validate Model
1	Marmor et al. [29]	Yes	No	No	No	8 Hrs	100	Yes/2	No	Yes
2	Rozinat et al. [39]	No	Yes	Yes	Yes	2 Wks	50	No/4	No	No
3	This thesis	Yes	Yes	Yes	Yes	8 Hrs	Automated	No/13	No	Yes

7.6. Limitations and Threats to Validity

There were several limitations and threats to validity in conducting our research. The threats can be reviewed in three different categories, as suggested by Perry et al. [33]:

- *Construct validity*: Specifies how well the case study answered the research question.
- *Internal validity*: Examines any bias and confounding factors.
- *External validity*: Specifies how much of the results can be generalized.

7.6.1 Construct Validity

We have not used the exact real data for feeding our simulation model, and instead we have used approximate (but realistic) data. The main reason is because of the confidentiality of the hospital's patient information. The data collected was partially taken from the hospital database and the rest was obtained by indirectly interviewing a domain expert of the hospital who has been involved in the care process we were interested in.

The actual implementation of the created simulation model was not done in the mentioned hospital. Therefore, even if the real-time data was available for injection in the simulation model, we still need to implement the model and actually run it in the hospital and validate the results.

However, note that the focus of the thesis was more about the process of creating and exploiting real-time simulations in healthcare and about demonstrating their potential than in using them to solve a specific problem in a real environment.

Also, the processes involved in clinical environments are usually of a complex nature because a majority of their tasks are human-related. In most simulation studies, it is assumed that the resources are fully dedicated to the patient flow process and if they are scheduled and are idle, they will start their task right away. In real-life, however, this is not the case and hence the exact behaviour of the resources is complex to capture in a simulation model. In our case study, we could not address this problem but we tried to adjust the model to the reality by setting *delays* that usually occur in our patient flow process.

7.6.2 Internal Validity

The conceptual model was built from a business process model and was then mapped to a simulation model. Additional data was injected to the simulation model afterwards. Since the simulation model has been created based on a BPM model, we have assumed that the BPM model is itself a valid and reliable model. Note that the source BPM model was not created by the thesis author, but by fellow students based on an earlier version developed by Osler nurses. Some bias was hence mitigated at that level.

7.6.3 External validity

We have selected only one patient flow in our case study to explore the feasibility of our research question. We cannot yet generalize our work (including the simplification and conversion guidelines) to all the systems that provide clinical services and are dealing with random arrival of the patients. However, this process is believed to be representative of many other types of healthcare processes, and the selected hospital (Osler) is also representative of many Canadian healthcare institutions. Yet, this work should be replicated with more processes in different hospitals to augment the level of confidence in its conclusions.

7.7. Summary

We have compared our work with two closely-related approaches (and we are not aware of other such approaches), and their weaknesses and strengths have been discussed. The contributions of this research are unique in their depth and efficiency for the proposed scope (real-time simulations for healthcare processes). Yet, several threats to the validity of our research exist and have been discussed.

The next and final chapter will conclude the thesis and identify future work items.

Chapter 8. Conclusions

This chapter goes back to the initial research question and recalls the main contributions of this thesis. Suggestions for future work are also discussed.

8.1. Contributions

The research question addressed in this thesis (Section 1.2) was: *Is the real-time simulation of clinical processes feasible in the context of a patient care process, with wait time predictions represented in a way that supports short-term operational decision making related to resource allocation?*

We believe that the work done in the thesis and the contributions made along the way result in a positive answer to this question.

In this thesis, the application of real-time simulations in healthcare for short-term decision making has been investigated. An entire clinical pathway for cardiac patients of a community hospital has been modeled. A conceptual model has been created for facilitating the process of simulation modeling. Some helpful guidelines were used to transform a BPMN operational model of the patient flow to a simulation model. Also, other information was added to the simulation model to enable loading of the current state of the hospital into the model. An external file populates the simulation model with historical data and the initial state data. The model starts executing from the initial state loaded to the simulation model at the beginning of every replication. The model is set up to make predictions over a short period of time in the future and collects statistical data of interest regarding the near-future state of the hospital. Afterwards, some performance metrics are defined for a quantitative comparison among different possible alternative actions that could be taken. The alternative scenarios, also known as what-if scenarios, are run and will show the predicted impact of any decision made in the present time on the future state of the hospital. The trade-offs related to the variety of performance metrics should be monitored and analyzed. The decision maker (possibly a senior nurse) is then able to make quick, feasible and timely decisions for the near future, based on the information provided through running what-if scenarios. The complementary use of an optimizer was also explored along the way.

Our approach performs well compared to closely-related approaches, and also against most healthcare simulation approaches in general. Table 17 provides one additional row for Table 3 (the summary of related work) that captures the characteristics of the work done in this thesis.

Table 17 Thesis Approach

Article / Source	Country	Scope	Type of simulation	Real-time implementation	Effective in DM	Patient wait time issue	Actual hospital	Simulation tool	Data collection	Model V & V
This thesis	Canada	ED & CW	DES	Yes	Yes	Yes	Yes	Arena	2	Partially

The main contribution of this thesis is the application of real-time simulations for supporting short-term decision making related to patient care processes in a timely way. There are also several minor contributions that were provided through this research work:

- We have defined a methodology that can help realize real-time simulation capabilities for supporting decision making.
- We have also used the real-time simulation methodology in a case study for assessing its validity. Through this practical work, we added real-time simulation capabilities (e.g., initial state of queues and resource scheduling) to a leading simulation tool that did not support the required real-time features in the first place.
- We have developed and tested guidelines that explain how we can reuse a BPM model to construct a conceptual model and map that conceptual model to a simulation model.
- Due to short-term decision making requirements, we have suggested an approach through which the number of simulation replications is adjusted automatically for each set of simulation runs, based on the desirable precision of the output data.

- Using real-time simulations, we showed that we can have a more frequent performance evaluation of the system under observation (in our case, a hospital) by running the simulation in a timely manner. We have also defined what-if scenarios with different resource scheduling combinations, and we have observed the impact of each scenario on the future state of the care process, so support decision making.

This thesis has hence advanced knowledge on the applicability of real-time simulations for healthcare by providing an approach, supported by guidelines and tools, for model construction and analysis, with a sample case study illustrating both the usage of these guidelines and the overall feasibility of the approach.

8.2. Future Work

Future work items include an integration of BPM-powered monitoring system with the simulation tool to automatically feed the historical and initial state data. The historical data should be collected from the database that collects real-time process information about patients and should be transformed to statistical distributions prior to loading. The collection and injection of the real data to the simulation model should be triggered periodically followed by an automated activation of scenario runs. In addition, usability studies should be conducted to better evaluate the work.

An automated decision support system that can define and evaluate different scenarios and present the best alternatives would make the process of decision making quicker and easier. Manually setting the scenarios does not necessarily result in the best feasible implementation alternatives; thus, using an optimizer can broaden the variety of scenario evaluation. Also, the time required for setting up and running an optimization should be considered because of the constraints imposed by short-term decision making.

In the experiment, we have only considered the walk-in cardiac patients of the hospital. The *scheduled patients* should also be considered while modeling to reflect the reality of the care processes better and to increase the reliability of the output results of simulations.

Finally, although a real cardiac clinical pathway has been used, with realistic data and a model validated by a domain expert, additional validation needs to be done before generalizing the results and deploying this approach in hospitals.

References

[1] Alvarez, A. and Centeno, M.: Enhancing Simulation Models for Emergency Rooms Using VBA. *31st Winter Simulation Conference (WSC 99)*, vol.2, ACM, 1999, pp.1685-1693.

[2] Bahrani S., Bougueng R. T., Mouttham A., and Amyot D.: Real-time Simulations to Support Operational Decision Making in Healthcare. *Summer Computer Simulation Conference and Work in Progress (SCSC 2013 and WIP 2013)*, Simulation Series, Vol 45 #11, Curran Associates, Inc., July 2013, pp. 768-775.

[3] Barjis, J.: Potentials and Challenges of Healthcare Modeling and Simulation. *SCSC '09 Proceedings of the 2009 Summer Computer Simulation Conference,* Society for Modeling & Simulation International, 2009, pp. 173-176.

[4] Baumgart, A., Zoeller, A., Denz, C., Bender. H., Heinzl, A., and Badreddin, E.: Using Computer Simulation in Operating room management: Impacts on process engineering and Performance. *40th Annual Hawaii International Conference on System Sciences, HICSS 2007*, IEEE CS, 2007, p.131.

[5] Bernardo, A., Borges, J., Brito, A., Morteo, A., Sperandio, F., and Gomes, C.: Simulating a Portuguese Hospital Master Surgery Schedule. *1st Int. Conf. on Serious Games and Applications for Health (SeGAH)*, IEEE CS, 2011, pp.1-4.

[6] Birta, L. and Arbez, G.: *Modelling and Simulation*. Springer, 2007.

[7] Blake, J.T., Carter, M.W., and Richardson, S.:An evaluation of emergency room wait time issues via computer simulation. *INFOR*, Vol. 34, 1996, pp. 263-273.

[8] Bougueng Tchemeube, R.: *Location-Aware Business Process Management for Real-time Monitoring of Patient Care Processes*, Master of Computer Science thesis, University of Ottawa, Canada, July 2013.

[9] Carter, M.W. and Blake, J.T.: Using Simulation in an Acute-Care Hospital: Easier Said Than Done. *Handbook of Operations Research in Healthcare*. New York: Kluwer, 2005, 191-215.

[10] Centeno, M.A and Ismail, A.M.: A Simulation-ILP Based Tool for Scheduling ER Staff. *Proc. of the 2003 Winter Simulation Conference.,* vol.2, IEEE CS, 2003, pp.1930-1938.

[11] Cochran, J.K. and Bharti, A.: Stochastic Bed Balancing of an Obstetrics Hospital. *Health Care Management Science Journal*, Kluwer Academic Publishers, 2006, Volume 9, Issue 1, pp. 31-45.

[12] Eskandari, H. Khosravi, S., and Geiger, C.D.: Improving the Emergency Department Performance Using Simulation and MCDM Methods. *43rd Winter Simulation Conference (WSC 11),* IEEE CS, 2011, pp. 1211-1222.

[13] Everett, J.E.: A decision Support Simulation Model for the Management of an Elective Surgery Waiting System. *Health Care Management Science Journal*, Kluwer Academic Publishers, 2002, Volume 5, Issue 2, pp. 89-95.

[14] Fei, H., Meskens, N., and Darzi, E.: Evaluating Alternative Surgery Plans With Discrete Event Simulation Model. *Health Care Management (WHCM), 2010 IEEE Workshop*, IEEE CS, 2010, pp. 1-6.

[15] Fryk, P. and Steins, K.: A Modern Process Perspective, Process Mapping and Simulation in HealthCare, Opportunities and IT Infrastructural Needs. *Health Care Management (WHCM), 2010 IEEE Workshop on*, IEEE CS, 2010, pp.1-6.

[16] Gopakumar, B., Wang, Sh., Khasawneh, M., Srihari, K, Gandhi, T., and Saundres, N.: A Systems Approach for Hospital Inpatient Bed Capacity Planning. *IEEE Int. Conf. on Industrial Engineering and Engineering Management*, IEEE CS, 2009, pp. 744-748.

[17] Gunal, M. and Pidd, M.: SGHPSIM: Supporting Smart Thinking to Improve Hospital Performance. *40th Winter Simulation Conference (WSC 08)*, IEEE CS, 2008, pp.1484-1489.

[18] Gunal, M. and Pidd, M.: Interconnected DES Models of Emergency Outpatient and Inpatient Departments of a Hospital. *39th Winter Simulation Conference (WSC 07)*, IEEE CS, 2007, pp. 1461-1466.

[19] Hevner, A.R., March, A.T., Park, J., and Ram, S.: “Design science in information systems research”. *MIS Quarterly*, Vol. 28, No.1, 2004, pp. 75-105.

[20] Holm, L.B. and Dahl, F.A.: Simulating the Effect of Physician Triage in the Emergency Department of Akershus University Hospital. *41st Winter Simulation Conference (WSC 09)*, IEEE CS, 2009, pp. 1896-1905.

[21] Kelton, W.D., Sadowski, R., and Swets, N., 2010. *Simulation with Arena: 5th edition*. Macgraw-Hill, New York, USA.

[22] Kitchenham, B.: *Procedures for performing systematic reviews*. Technical Report, Keele University and NICTA, Staffordshire, UK, 2004.

[23] Koizumi, N., Kuno, E., and Smith, T.E.: Modeling Patient Flows Using a Queuing Network with Blocking. *Health Care Management Science Journal*, Springer US, 2005, Volume 8, Issue 1, pp. 49-60.

[24] Kolb, E.M.W., Peck, J., Schoening, S., and Lee, T.: Reducing Emergency Department Overcrowding Five Patient Buffer Concepts. *40th Winter Simulation Conference (WSC 08)*, IEEE CS, 2008, pp. 1516-1525.

[25] Kolb, E.M.W, Lee, T., and Peck, J.: Effect of Coupling between Emergency Department and Inpatient Unit on the Overcrowding in Emergency Department. *39th Winter Simulation Conference (WSC 07)*, IEEE CS, 2007, pp. 1586-1593.

[26] Kolker, A.: Process Modeling of ICU Patient flow: Effect of Daily Load Leveling of Elective Surgeries on ICU Diversion. *Journal of Medical Systems (J Med Syst)*, 2009, Springer US, Volume 33, Issue 1, pp. 27-40.

[27] Komashie, A., Mousavi, A., and Gore, J.: Using Discrete Event Simulation (DES) to Manage Theatre Operations in Healthcare: An Audit-based Case Study. *Tenth Int. Conf. on Computer Modeling and Simulation (UKSIM'08)*. IEEE CS, 2008, pp. 360-365.

[28] Macdonald, S., Karkam, I., Al-Shirrawi, N., Chowdhary, R., Escalante, E., and Afandi, A.: Emergency Department Process Improvement. *Systems and Information Engineering Design Symposium,* IEEE CS, 2005, pp.253-262.

[29] Marmor, Y.N, Wasserkrug, S., and Shtub, A.: Towards Simulation-Based Real-Time Decision-Support Systems for Emergency Departments. *41st Winter Simulation Conference (WSC 09)*, IEEE CS, 2009, pp. 2042-2053.

[30] Nance, R.E.: *A History of Discrete-Event Simulation Programming Language*. TR 93-21, Department of Computer Science, Virginia Polytechnic Institute and State University, USA, 1993.

[31] OMG: *Business Process Model and Notation (BPMN)*, version 2.0, January 2011. http://www.omg.org/spec/BPMN/2.0/

[32] Ontario Ministry of Health and Long-Term Care, http://edrs.waittimes.net/En

[33] Perry D. E., Porter A. A., and Votta L. G.: Empirical studies of software engineering: a roadmap. *Conf. on the Future of Software Engineering*, ICSE'2000, ACM, 2000, pp. 345-355.

[34] Persson, M.J. and Persson, J.A.: Analysing Management Policies for Operating room Planning Using Simulation, *Health Care Management Science journal*, Springer US, 2010, Volume 13, Issue 2, pp. 182-191.

[35] Preston White, Jr. K.: A Survey of Data Resources for Simulating Patient Flows in Healthcare Delivery Systems. *37th Winter Simulation Conference (WSC 05),* IEEE CS, 2005, pp. 926-935.

[36] Protil, R.M., Stroparo, J.R., and Bichinho, G.L.: Applying Computer Simulation to Increase the Surgical Center Occupation Rate at a University Hospital in Curitiba, Brazil. *40th Winter Simulation Conference (WSC 08)*, IEEE CS, 2008, pp.1609-1616.

[37] Raunak, M., Osterweil, L., Wise, A., Clarke, L., and Henneman, P.: Simulating Patient Flow through an Emergency Department Using Process Driven Discrete Event Simulation. *2009 ICSE Workshop on Software Engineering in Health Care (SEHC'09)*, IEEE CS, 2009, pp. 73-83.

[38] Rohlder, T.R., Lewkonia, P., Bischak, D.P., Duffy, P., and Hendijani, R.: Using Simulation Modeling to Improve Patient Flow at an Outpatient Orthopedic Clinic, *Health Care Management Science Journal,* Springer US, 2011, pp. 135-145.

[39] Rozinat, A. Wynn, M.T, Van der Aalst, W.M.P, Ter Hofstede, A.H.M and Fidge, C.J: Workflow simulation for operational decision support, *Data & Knowledge Engineering,* Elsevier Science Publishers, 2009, Vol. 68, Issue 9, pp. 834-850.

[40] Sepulveda, J. and Cahoon, L.E.: The Use of Simulation for Process Improvement in a Cancer Treatment Center. *31st Winter Simulation Conference (WSC 99),* ACM, vol. 2, pp. 1541-1548.

[41] Spry, C.W. and Lawley, M.A.: Evaluating Hospital Pharmacy Staffing and Work Scheduling Using Simulation. *37th Winter Simulation Conference (WSC 05)*, IEEE CS, 2005, pp. 2256-2263.

[42] Weng, S.J., Cheng, B.C, Kwong, S.T, Wang, K.M and Chang, C.Y. Simulation Optimization for Emergency Department Resources Allocation. *43rd Winter Simulation Conference (WSC 11)*, IEEE CS, 2011, pp. 1231-1238.

[43] William Osler Health Center: *William Osler Health Centre's emergency department wait time information now available online*, 2009, http://www.williamoslerhc.on.ca/workfiles/MediaRelease-EDWaitTimes.pdf

[44] Williams, P., Tai, G., and Lei, Y.: Simulation Based Analysis of Patient Arrival to Healthcare Systems and Evaluation of an Operations Improvement Scheme, *Annals of Operations Research*, Springer US, 2010, Volume 178, Issue 1, pp. 263-279.

[45] Wijewickrama, A. and Takakuwa, S.: Simulation Analysis of Appointment Scheduling in an Outpatient Department of Internal Medicine. *37th Winter Simulation Conference, (WSC 05)*, IEEE CS, 2005, pp. 2264-2273.

[46] Wijewickrama, A. and Takakuwa, S.: Simulation Analysis of an OutPatient Department of Internal Medicine in a University Hospital. *38th Winter Simulation Conference (WSC 06)*, IEEE CS, 2006, pp. 425-432.

[47] Wong, S.Y., Tsui, K.L., Chin, K.S., and Xu, M.: A Simulation Study to Achieve Healthcare Service Quality Improvement in Accident & Emergency Department (AED). *Quality and Reliability (ICQR), 2011 IEEE International Conference on*, 2011, IEEE CS, pp. 259-263.

[48] Young, T., Jahangirian, M., and Lilford, R.: Three Critical Challenges for Modeling and Simulation in Healthcare. *41st Winter Simulation Conference (WSC 09),* IEEE CS, 2009, pp. 1823-1830.

[49] Yu, Y., Lifeng, W., Huang, H., and Perera, T.: Modeling & Simulation of an Outpatient Service System. *Computer Application and System Modeling (ICCASM), 2010 International Conference on*, 2010, vol.2, pp. 691 -693.

Appendix A: Arena Simulation Model of Patient Flow

The Arena simulation model built in this thesis is available online in two complementary formats:

- A diagram (in PDF) that shows the simulation modules and their links: http://www.eecs.uottawa.ca/~damyot/pub/Bahrani/model.pdf
- A report that describes the characteristics of all the Arena modules in this model, including default values and conditions: http://www.eecs.uottawa.ca/~damyot/pub/Bahrani/model.html

In order to give an idea of its size and complexity, Figure 20 gives a bird's eye view (not meant to be read) of this simulation. Note the presence of the MOE monitoring process (expanded in Figure 18) on the top-left corner.

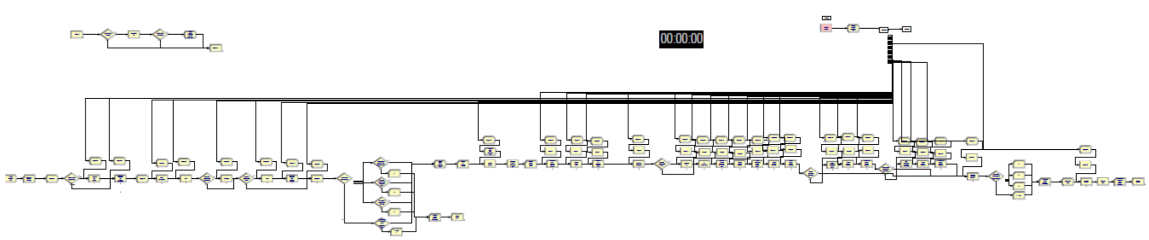

Figure 20 Bird's Eye View of the Simulation Model in Arena