Process Improvement in a Private Hospital using Discrete Event Simulation

The Case of Companion and Complementary Diagnostics

Jorge Daniel Carvalho Mira

Thesis to obtain the Master of Science Degree in

Industrial Engineering and Management

Supervisor: Prof. Diogo Filipe da Cunha Ferreira

Examination Committee

Chairperson: Mónica Duarte Correia de Oliveira

Supervisor: Diogo Filipe da Cunha Ferreira

Supervisor: Alexandre Manuel Martins Morais Nunes

Members of the Committee: Amílcar José Martins Arantes

December 2022

Declaration

I declare that this document is an original work of my own authorship and that it fulfils all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Declaração

Declaro que o presente documento é um trabalho original da minha autoria e que cumpre todos os requisitos do Código de Conduta e Boas Práticas da Universidade de Lisboa.

– Murphy's Law is real.

É com este presente trabalho que termino o maior ciclo da minha vida. Ficou o sentimento de missão e objetivo cumpridos ainda que o maior ainda esteja por vir.

Com isto, começo por expressar o meu maior agradecimento aos meus pais, que estiverem sempre disponíveis para mim desde o primeiro dia. Que me deram todas as bases para chegar até aqui, para além de todo o apoio. Da mesma forma, agradeço à minha família, nomeadamente tios, primos e padrinhos, e amigos que estiveram presentes durante este percurso e que me puderam proporcionar os melhores momentos.

Profissionalmente, não podia deixar de agradecer ao meu orientador, o Prof. Diogo da Cunha Ferreira, que aceitou guiar-me na dissertação com o tema que propus.

Como não podia deixar de faltar, o meu muito obrigado à equipa do Hospital da Luz, nomeadamente ao Eng. Gonçalo Luís, por me ter fornecido todos as bases para que este estudo fosse realizado e por todo o seu tempo dedicado ao projeto, e à Dra. Madalena Castelo Branco por toda a recetividade, compreensão e motivação.

Por último, mas não menos importante, o meu agradecimento ao P. Carlos Azevedo pela sua maior compreensão, pelas suas palavras e força que me deu.

Abstract

In recent years, the need for health units has increased owing to the ongoing introduction of new health problems and the intensification of existing kinds of illnesses. As research on this area progresses, knowledge becomes the primary motivation for patients to adopt a better quality of life. Consequently, a number of health units have struggled to develop solutions to manage excessive demand while maintaining a high quality of service in order to enhance patient satisfaction and boost revenues. One of these tactics is the use of simulation models as a method of operational research to analyse the flow of patients and predict potential solutions to enhance resource efficiency while decreasing waiting time.

The purpose of this work is to develop and demonstrate a simulation model that replicates the functioning of a hospital unit. For this purpose, data derived from the health services of the Imaging Department of Hospital da Luz, located in Lisbon, Portugal, were used. This model will be able to duplicate the daily flow that this department is subjected to by its patients and personnel, as well as depict how the system evolves as a result of the management of the department's number of resources and examination rooms. At the end of the analyses, several solutions are presented and discussed, from level to level, with regard to improvements in the level of performance of services, resources and patients, although it appears that only the increase in one unit of examination rooms translates into a huge positive impact on waiting times in the respective queues.

Keywords: healthcare simulation, process improvement, simulation models, operational research methods, discrete event simulation.

Resumo

Nos últimos anos, a procura por unidades de saúde aumentou devido ao aparecimento constante de novos problemas de saúde e intensificação de tipos de doenças já existentes. À medida que as investigações nesta área progridem, a informação torna-se a principal motivação para que os pacientes adotem um estilo de vida mais saudável. Consequentemente, várias unidades de saúde têm-se esforçado para desenvolver estratégias de forma a gerir a alta procura, tentando manter uma boa qualidade de serviço para aumentar a satisfação dos pacientes e os lucros. Uma dessas estratégias é o uso de modelos de simulação como método de investigação operacional para analisar o fluxo de pacientes e prever possíveis soluções para combater tempos de espera mais elevados e aumentar a eficiência dos recursos.

O objetivo deste estudo é construir e apresentar um modelo de simulação que traduza o funcionamento de uma unidade de saúde de um hospital real. Para este efeito, foram utilizados dados dos serviços de saúde do Departamento de Imagiologia do Hospital da Luz, situado em Lisboa, Portugal. Este modelo será capaz de reproduzir o fluxo a que este departamento está sujeito pelos seus pacientes e funcionários todos os dias durante o seu funcionamento e ainda demonstrar de que forma o sistema evoluiu provocado pela gestão do número de recursos e salas de exames através de várias experiências realizadas. No fim das análises, várias soluções são apresentadas e discutidas, de nível a nível, no que toca às melhorias do desempenho dos serviços, recursos e pacientes, ainda que se verifique que apenas o aumento numa unidade de salas de exame se traduza num enorme impacto positivo nos tempos de espera nas respetivas filas.

Palavras-chave: simulação na saúde, melhoria de processos, modelos de simulação, métodos de investigação operacional, simulação por eventos discretos.

Table of Contents

List of Illustrations

List of Tables

List of Acronyms

ATT – Arrival to Treatment Space

ATP – Arrival to Provider Time

CAT – Computerised Axial Tomography

CTAS – Canadian Triage and Acuity Scale

ED – Emergency Department

ESI – Emergency Severity Index

IS – Imaging Service

KPI - Key Performance Indicator

LOS – Length of Stay

MRI – Magnetic Resonance Imaging

MTS – Manchester Triage System

PD – Process Duration

US – Ultrasound Service

WT – Waiting Time

1. Introduction

In this chapter, the dissertation is introduced, the issue is explained, the study's objectives are stated, the research methodology is presented, and the dissertation's structure is described. The backdrop of the subject under investigation is provided in *section 1.1.*, the dissertation's objectives are presented in *section 1.2.*, the research methodology is described in *section 1.3.*, and the dissertation's structure is defined in *section 1.4.*.

1.1. Problem Contextualization and Background

Patients are increasingly curious and apprehensive about receiving healthcare due to global competitiveness in a rising industry (Li et al., 2015). Manzoor et al. (2019) correlate that increasing health consciousness and affluence in the contemporary world have dramatically increased the demand for healthcare and moved demographic trends toward a healthier way of life. As a consequence, it has been producing a demanding environment that has an impact on local businesses, including medical services. Due to the growing rivalry between hospitals, the healthcare relationship has evolved to emphasise the provision of outstanding healthcare services (Bleustein, Clifford et al., 2014), leading patients to pick the most suitable healthcare unit.

To contend with the competition, hospitals and healthcare units must provide patients with better healthcare services to meet their demands due to heightened consumer expectations and more demand for standard services (Zarei et al., 2015). As a result, hospitals confront significant obstacles. As the need for medical treatment develops, hospitals may dissatisfy patients with prolonged LOS caused by excessive wait times (Manzoor et al., 2019). Consequently, Wang et al. (2012) agree that a number of them have made substantial efforts to improve hospital efficiency to reduce LOS and other issues such as patient wait times, patient satisfaction, levels of spending, capacity, resource consumption, working conditions, staff morale, accessibility to error-free treatment, and medicines all need to be efficiently managed while the quality of care is continuously improved (Barjis, 2011).

Hospitals and other healthcare institutions often base their continuous improvement activities on human experiences and qualitative notions. However, such strategies may not deliver the big improvement one would have hoped for, and it is impossible to make a quantitative forecast of the outcome of such an effort (Wang et al., 2012).

A. R. Andersen & Plesner (2022) remark that governments often seek to alleviate the problem by raising the proportion of their gross domestic product allocated to healthcare while concurrently minimising the duration of patients' stays. For example, Portugal's public healthcare expenditures surpassed 15 billion euros in 2021.[1] Healthcare executives and decision-makers are increasingly

[1] https://pt.countryeconomy.com/governo/despesa/saude/portugal

debating healthcare efficiency (Zeng et al., 2012), spending about 11% of its gross domestic product (GDP) on healthcare.[2] To achieve this efficiency, computer simulation to aid in efficient decision-making in the healthcare industry has become increasingly popular in recent years to enhance operations (Doğan & Unutulmaz, 2014). A simulation model may mimic the process and its dynamics under certain random distributions, show patient flow and care delivery techniques, and provide forecasts for performance evaluation (Gaba, 2004). With the support of such a tool, healthcare management can assess the effectiveness of existing procedures, enable the studying of potential modifications, experience circumstances that would not otherwise be possible without spending a significant amount of money on system development, training, and equipment acquisition, or look into the relationships or trade-offs among system variables (Manzoor et al., 2019).

Furthermore, the typical functions of comparing situations or visualising processes may be expanded upon by using healthcare simulation. A simulation model may be used as a continuous initiative to assess and enhance performance and boost efficiency.

1.2. Dissertation Purpose

Motivated by the contextualisation described in the preceding section, the primary objective of this research is to develop a simulation model capable of portraying the functioning of an Imaging Department. This model will be used to test various combinations of internal factor changes, such as the number of personnel resources and facilities of a particular service, to examine the development of system performance under different scenarios. Thus, for each tested combination, it will be possible to predict how the unit will behave, determining the number of resources sufficient to reach a state that already reflects a good performance for the decision-maker.

In addition, this study complements the evaluation of the effectiveness of human resources in each scenario, as assessed by the occupancy rate, which indicates the proportion of execution time that a particular resource was dedicated to an activity. It should be observed that the occupancy rate increases when the amount of a given resource allocated to a given number of activities decreases, but this does not equate to more excellent performance in terms of waiting times, thus indicating a trade-off.

In conclusion, the constructed model should offer the department helpful information to enhance its resource management. This model's precision matches demand and supply, allowing for the optimal allocation of resources in time and space while reducing queueing times.

[2] https://www.pordata.pt/en/Portugal/Current+expenditure+on+healthcare+as+a+percentage+of+GDP-610

1.3. Methodology

Figure 1 depicts a potential research methodology based on five phases to attain the objective outlined in the preceding section.

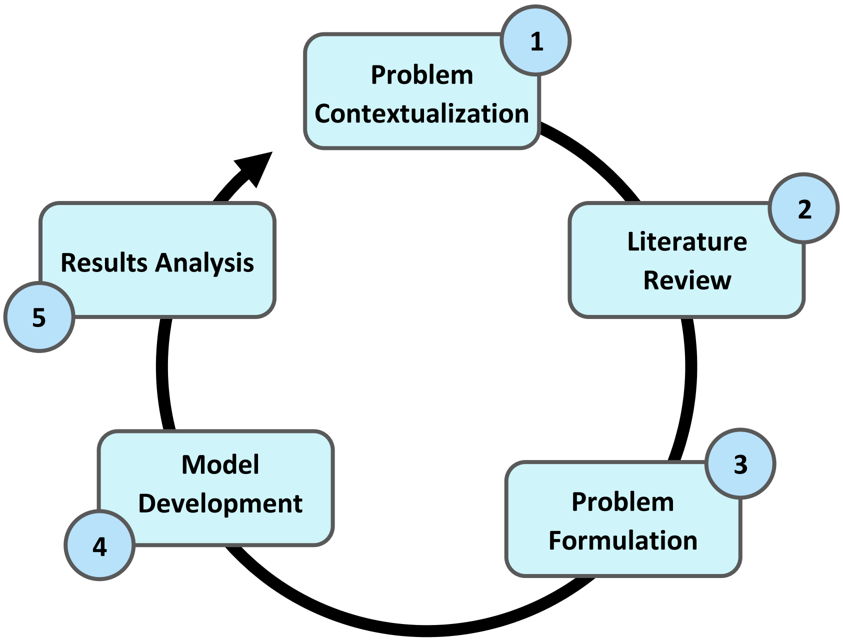


Figure 1 *– Proposed Methodology.*

Step 1 – Problem Contextualization

The contextualisation of the problem guides the reader from a broad subject area to a specific topic of study. It outlines the study's scope, context, and importance by summarising existing knowledge and background information on the issue. It outlines the study's objective in the form of a research topic supported by a collection of questions or a hypothesis[3]. Additionally, it briefly explains the methodological approach used to investigate the research problem, highlights the potential outcomes the study may reveal, and outlines the remaining structure and organisation of the paper. It is inserted in chapter 1 of this document.

Step 2 – Literature Review

A literature review examines relevant sources pertinent to acquiring knowledge on healthcare simulation subjects. It gives a description, summary, and assessment of these works in connection to the explored research problem in a way that serves as a basis to develop it. It consists of chapters 2, 3, 4 and 5 of this document.

Step 3 – Problem Formulation

This step consists of providing a concise description of the whole issue and emphasising the need for a more profound knowledge of the most crucial aspects to conduct a more thorough inquiry. In this

[3] https://libguides.usc.edu/writingguide/

manner, the formulation of the problem brings the reader to the significance of developing the study to establish the research areas. In addition, it assures the research questions, hypotheses, or assumptions to be followed and presents the framework that will guide the whole process until the analysis of the results. This step entails chapter 6 of this document.

Step 4 – Model Development

This stage, described in depth in chapters 7 and 8, focuses on creating the model formulated in the previous step. It begins by examining the data given through papers and files to find trends and comprehend patient behaviour. Given the nature of the model, the dataset must be fitted several times as a probability distribution before it can be incorporated into the model. The final step is configuring the model within the SIMUL8 software, in which the model presented in the previous section will be reproduced by linking the activities to the queues, associating the required resources with each activity, and inserting the previously obtained inputs with each activity.

Step 5 – Results Analysis

This step involves analysing the results acquired from the preceding step to determine the influence of each testing scenario on the system's performance and the advantage of utilising such a combination of activities and resources to support the operations of the healthcare unit. The recommended configurations of the model are evaluated in terms of average queuing time, maximum queuing time, and efficiency in terms of resources' occupancy rate; a comparison analysis is undertaken to determine the benefits of each combination performed and produce recommendations. Chapters 9 and 10 are responsible for addressing this stage.

1.4. Dissertation Structure

This dissertation consists of ten chapters. This first chapter attempts to explain the topic and the rationale for its research, as well as its goals, methodology, and structure of the document.

The second chapter 2 starts with a concise description of what it means to be a system, identifying the variables that may affect it. The primary distinctions, benefits, and drawbacks of studying a system via various approaches are outlined, depending on the context of the problem: the first relates to the difference between studying a system through experiments and researching it in a real-world setting or by modelling. The second describes the nature of the models, whether they are physical or mathematical, providing an example of each. The third refers to the search for solutions, either by analytical methods or by simulation, with an emphasis on the latter, providing the categorisation of simulation models along three dimensions.

The objective of the third chapter is to expose the reader to the three most common and widely used simulation techniques: discrete-event simulation, agent-based simulation, and dynamical systems. As noted in the first chapter, this dissertation is based on the first method's application. The whole internal functioning of this method is described in depth, including a description of all the system's components and how they interact to transit the interstate system. This transition is governed by the simulation clock, which may move time to the next event in the list of events. In addition, a brief description of the other two techniques is given.

The fourth chapter uses information from multiple case studies on the same issue to develop a simulation model for healthcare systems. This chapter groups the proposed methodology in several case studies, beginning with understanding the problem's context, collecting data, debating the variables of interest, selecting the tool where the model will be developed, and concluding with the validation of the conceptual model and experiments with decision factors in order to find an optimal solution or set of solutions. Thus, this chapter is divided into four stages: understanding the problem scenario, determining the problem objectives, selecting the model content, and constructing the model.

The fifth chapter is a continuation of the preceding chapter in which the actual case study of Wang et al. (2012) using DES at Central Baptist Hospital is presented. The purpose is for the reader to comprehend in practice the whole procedure described in the previous chapter.

The sixth chapter presents all the knowledge about the problem. Here, the entire scope and functioning of the health unit are specified in great detail to be implemented in the same way during the model's construction phase.

The seventh chapter describes all operations performed on the data required to create the model, from its gathering through its inclusion. In addition, it represents the processes performed for each dataset and how they were categorised. This chapter begins by justifying the significance of the General Data Protection Regulation (GDPR) legislation and how it was considered in the dissertation.

The eighth chapter focuses on model construction. Here, all software operations are specified to reproduce the conceptual model accurately. This chapter is, therefore, broken into two sections. The first section describes how each model object was defined, including the simulation clock, arrival shifts, resources, labels, and the list of activities and queues. Thus, all configurations essential for the system to function, as described in chapter 6, were adequately stated for each component. The second section presents four verifications necessary to validate the model and make it reliable for drawing conclusions.

The ninth chapter discusses the whole set of results derived from running the system in the simulation tool. The initial conditions for running the system are provided first, followed by the termination conditions, i.e., when it is no longer essential to run the system to retrieve the data for analysis. Then, a technique for solution search is proposed based on the analysed values. The final step is the interpretation and presentation of the results to the decision-maker.

Chapter 10 focuses on finishing the process covered in the preceding chapters to summarise and emphasise each one's major features. All limitations that had an effect on or impacted the interpretation of the findings or the design of the model are also provided. Finally, the modeller specifies certain recommendations and proposals for future work that were not evaluated or considered when the problem was formulated but which have the potential to further enhance the system's performance.

2. Systems, Models and Simulation

According to the definition proposed by Schmidt & Taylor (1970) a system is a "collection of entities, such as people or machines, that act and interact together toward the accomplishment of some logical end.". For one research, collecting entities that constitute a system could only represent a portion of the whole system (Sinha et al., 2001). An example of that may be revealed through the development of a bank system that estimates the number of tellers required to give satisfactory service to customers who wish to cash a check or make a savings deposit. In this case, the system may be described as the portion of the bank, including the tellers and customers in line or being serviced. The concept of the system must be enlarged if, on the other hand, the loan officer and safe deposit boxes are to be included (Fishman, 1978).

As per the research objectives, the set of variables required to characterise a system at a specific time is what is referred to as the system's state. The number of busy tellers, the total number of clients, and the time each customer enters the bank are a few examples of state variables that may be used in bank research (Law, 2015). Following this example, the number of customers in the bank is constantly fluctuating, and the corresponding system state variable may vary across particular time samples. This number is explained by the flow of customers that enters or finishes being served and departs. This provides a precise definition of a discrete system. By the definition of Eriksson et al. (2022), discrete systems are characterised by state variables that change instantly at distinct points in time. Contrary, the state variables of a continuous system change continuously concerning time. Since state variables such as height and velocity may vary continuously for time, a continuous system is exemplified by an aeroplane travelling through the air. In actuality, few systems are entirely discrete or continuous; yet, because one form of change predominates for most systems, it is often viable to characterise a system as discrete or continuous.

Law (2015) gives a variety of methods for studying a system expressed in the *Figure 2*.

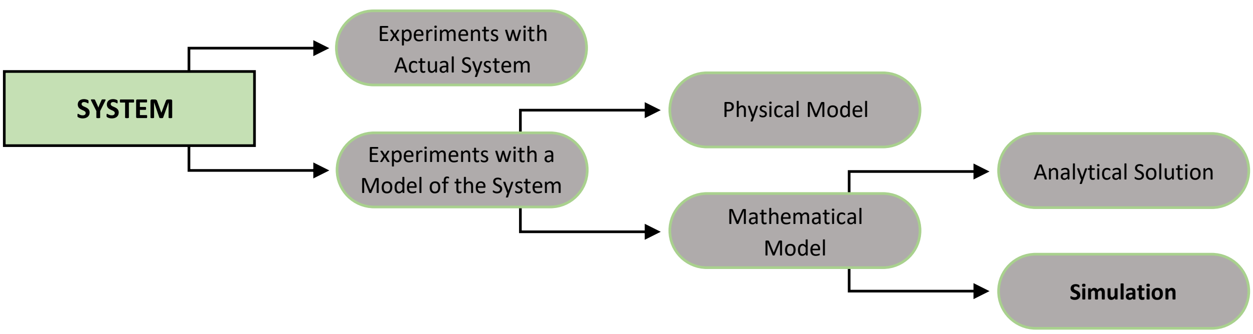


Figure 2 *– Ways of studying a system. Adapted from Law (2015).*

2.1. Experiments with the Actual System and Experiments with a Model of the System

In its majority, experiments with the actual system are rare possible since such an experiment would often be too expensive or disruptive to the system. Although, if it is feasible and cost-effective to physically modify the system and then allow it to run under the new circumstances (Morgan, 2005), it is likely preferable to do so since there will be no doubt about the validity of our research. For instance, a bank may consider lowering the number of teller positions to save expenses. Nevertheless, implementing this measure might result in lengthy client wait times and resentment.

Guala (2002) remarks that when the system is graphically designed into a model, it may be examined in its different suggested alternative configurations to determine how it should be constructed. Due to some factors (e.g., a planned communications network or a strategic nuclear weapons system), it is typically required to construct a model to represent the system and analyse it in place of the natural system. When using a model, there is always the question of whether it adequately represents the system for decision-making purposes (validation).

2.2. Physical Model and Mathematical Model

Physical models permit the display of information about the entity they represent. A model may be a physical thing on a big scale, such as a building's architectural model or a small scale, such as a molecule (Law, 2015). These models have been primarily associated with architectural and engineering purposes once it brings advantages to constructing them for studies.

According to Hidayat et al. (2020), mathematical models are represented through quantitative and logical relationships. They are then altered to see how the model responds and hence how the system would respond, considering that the mathematical model is valid. The relation between two variables, such as, represents a simple example of a mathematical model, where is the distance travelled, is the velocity and is the time travelled. This may be a viable model in one case (e.g., to discover the velocity a car travels, on average, to reach a specific destination in a certain time interval) but a flawed model for other applications.

2.3. Analytical Solution and Simulation

After the mathematical model has been constructed, it must be evaluated to see how it may be utilised to answer questions about the system it is meant to represent. If the model is simple enough, it may get an accurate, analytical solution by manipulating its relationships and quantities (Steuben et al., 2019). In the last example, if the distance to be travelled and the velocity are known, one can use the model to calculate the needed time $t = d/v$. This is a straightforward, closed-form solution, but some analytical solutions can become extraordinarily complex and require vast computing resources; inverting a large non-sparse matrix is a well-known example of a situation in which an analytical formula is known in principle, but obtaining it numerically in a given instance is not trivial. If an analytical solution to a mathematical model is available and computationally efficient, it is often preferable to examine the model in this manner rather than via simulation (Law, 2015). However, many systems are very complicated, such that accurate mathematical models are also complex, preventing an analytical solution. In this instance, the model must be investigated using simulation, i.e., numerically exercising the model for the relevant inputs to determine how they impact the output performance metrics (Schruben, 2008).

A simulation model may be integrated as part of continuing efforts to assess and enhance performance and boost productivity. Darema (2004) defends that, in this capacity, a simulation model is created to perform tests and integrate with the organisation's operational information systems. This is done to examine the longitudinal behaviour of a system in order to recommend adjustments and adaptations while the system works, and dynamic data is generated. When simulation models are entirely incorporated into the ordinary fabric of healthcare delivery, i.e., the current information system applications that support the daily operations of the healthcare provider, the actual value of simulation may be realised (Gaba, 2004). The objective is not to use simulation as a tool for doing a one-time set of tests when a significant change is planned but rather to have simulation models operate concurrently with other programmes as a common element of the daily work environment. Given that the simulation-study mathematical model requires searching for the appropriate tools, it is helpful to categorise simulation models along three dimensions for this purpose.

2.3.1. Static and Dynamic Simulation

A static simulation is a simulation model that represents a system at a certain moment or one that may be used to describe a system in which time is irrelevant (Law, 2015). It has no internal history of previously applied output and input variables.[4] In short, static simulation models are executed by specifying the parameters of the equations followed by adding the values of inputs necessary (Murray et al., 2012). This simulation is employed, for instance, when engineers compute the maximum weight a ship can carry. To determine the ship's maximum carrying capacity, a model based on the weight distribution will be constructed, disregarding any other factors (e.g., weather and strength of the tides). Since these elements that may impact the ship while carrying the load are not considered, it cannot offer correct findings for other situations that may arise while the ship is sailing.

In contrast, a dynamic simulation model is one method for modelling the behaviour of systems across time that accounts for multiple components of a phenomenon and focuses on how the system and its components change through time (Cole & Yount, 1994). A dynamic model is distinguished from a static model in that it maintains an internal memory of earlier inputs, variables, and outputs, unlike the static model. A study to optimize bed capacity in a healthcare facility may be represented through

[4] https://study.com/academy/lesson/static-vs-dynamic-simulation-in-quantitative-analysis.html

a dynamic model. It depends on several key factors, such as the number of daily admissions, service level and occupancy level (Kokangul, 2008).

2.3.2. Deterministic and Stochastic Simulation

Law (2015) assumes that a model is said to be deterministic if the variables that represent its state can be uniquely determined both by the parameters of the model and by the sets of states that these variables have held in the past. Deterministic models, as a result, always provide the same results when applied to the same set of parameters and beginning circumstances, and their solutions are always one and the same.[5]

In queueing models, a stochastic process is a collection of time-ordered random variables with their parameters described by random variables or distributions of those variables. The primary property of a random number is that the number that follows is independent of all preceding numbers (Law, 2015). Similarly, probability distributions may be used to represent the values of state variables. Therefore, a stochastic model will provide a multitude of equally plausible solutions, enabling the modeller to assess the inherent unpredictability of the underlying system being modelled (Renard et al., 2013). These models depict all conceivable states and their transitions, transition rates, and probabilities. Frequently, Markov models are used to represent the probability of distinct states and the rates of their changes. Typically, the approach is used to model systems. Markov models may also identify trends, generate predictions, and learn sequential data statistics. In this case, it is assumed that patients belong to homogenous groups and move from one state to the next. The benefit of using these models is that once designed can be evaluated and executed very fast. They may be used exploratively to determine the impacts of demand and supply on resource consumption and other output indicators (Davies & Davies, 1994).

2.3.3. Continuous and Discrete Simulation

The definitions of discrete and continuous simulation models are equivalent to those of discrete and continuous systems. Section *3* provides more explicit definitions of discrete-event simulation, which will be the base of this research. Notably, a discrete model is not always used to represent a discrete system, and vice versa (Law, 2015). The selection of a discrete or continuous model for a given system relies on the study's specific aims. For instance, if the features and movement of individual automobiles are significant, a traffic flow model on a motorway would be discrete. Alternatively, the traffic flow may be represented by differential equations in a continuous model if automobiles can be addressed aggregated.

[5] https://bookdown.org/manuele_leonelli/SimBook/types-of-simulations.html

3. Simulation Techniques

In addition to their effectiveness in designing, constructing, analysing, and developing physical systems, modelling and simulation methods are also frequently used in organisational and operational systems (Zhang et al., 2018). It entails designing a real-world or envisioned system model, such as a concept design, and then running tests with the model to understand the system's effectiveness under various operating conditions and evaluating possible management strategies and decision-making processes (Yin & McKay, 2018).

Numerous scientists have contributed to the development of technologies for modelling and simulation. Klingstam & Gullander (1999), for instance, presented the Discrete-Event Simulation (DES) approach, Macal and North suggested an Agent-Based Simulation (ABS) (Siebers et al., 2017) while Forester provided a System Dynamics approach (Forrester, 2017).

The following sections detail the functioning of each approach, with more emphasis on the DES as it will be the technique used to develop the work.

3.1. Discrete-Event Simulation (DES)

From the previously discussed models, Discrete-Event Simulation (DES) includes all dynamic, stochastic, and discrete properties combined in a system yet is an inexpensive, secure, and quick tool for analysing complicated systems, with the capacity to assess performance indicators (Mourtzis, 2019). DES is a computer-based modelling technique for decision-making that offers an easy and adaptable way (Karnon et al., 2012) of simulating the dynamic behaviours of complex systems and the interactions between people, communities, and their surroundings (Alejandro Huerta-Torruco et al., 2022). This model comprises the required properties and logic to reflect the system's actual behaviour (Urbani et al., 2020). Ideally, Law (2015) recommends DES to be conducted on a digital computer due to the number of data that must be saved and handled for the great majority of systems in the real world. However, depending on the volume of the data, it may also be performed by hand calculations.

Compared to aggregate models without interaction (Brennan et al., 2006), such as decision trees or Markov models, DES is a more beneficial operational research approach for modelling complex systems at the individual level than the cohort level (X. Zhang, 2018). It portrays the flow of individual entities that move through a succession of discrete events (activities) in a system that evolves over time (Law, 2015).

These events are defined at specific intervals, between which entities must wait in lines owing to the limited availability of resources. These situations are often seen as queuing networks (S. Brailsford & Hilton, 2000). For instance, in a hospital simulation, the entities are often patients who are transferred from one ward or institution to another. This simulation approach allows people to have traits that affect their progression in the system and resources to have limits (Alejandro Huerta-

Torruco et al., 2022). Thus, it gives an accurate illustration of how patient characteristics affect the functioning of a given health system. Furthermore, (Ahmad et al., 2020) defend those minimal restrictions on the factors that may be used to determine how and when patients transfer from one state to another.

Since each result from a DES is only a sample from a distribution, conducting several independent simulation runs may be necessary to get accurate measurements of the output parameters. This is especially problematic in unstable systems when the arrival rate is near the service rate or if activity time fluctuations are substantial (almost invariably accurate for activities representing treatment survival). An additional issue is a tendency to build systems more sophisticated than necessary, resulting in a rise in the need for data (Renard et al., 2013).

The planning of medical services for acute and chronic patients may benefit from patient flow models (Vázquez-Serrano et al., 2021). Some models assume that subgroups of patients are all the same and that events occur at regular intervals that are evenly spaced in time. When investigating patient movement in huge population groups, these approaches are practical. (Manzoor et al., 2019) state that DES models make it possible for patients to have unique characteristics and to interact with the supply of resources. However, testing and running these models takes much more time. They are handy for modelling healthcare delivery systems, mainly when the limitations of available resources are critical (Forbus & Berleant, 2022). In addition, they may be utilised on unrestricted population models that include many thousands of patients.

The ability to describe entities in such a way that they may participate in many activities simultaneously and interrupt one another is a huge step forward for simulation. The validity of any model depends on the availability of reliable data, which is only sometimes readily available (Davies & Davies, 1994).

3.1.1. Time Advances and Simulation Clock

Once the dynamic nature of DES models necessitates the continuous record of simulated time throughout the simulation process, it is essential to go forward in time from one value to another whenever desired (Law, 2015). The simulation clock is the variable responsible for determining the current simulation time, which is unrelated to the simulation's execution time on the computer.

Law (2015) identifies two approaches to make the simulation clock advance in time: next-event time advance and fixed-increment time advance. With the next-event time advance method, the simulation clock is set to zero, and the timings of future events' occurrences are calculated based on the input data. The simulation clock is subsequently advanced to the moment of the earliest (first) of these upcoming occurrences (state transaction). At this moment, the system's state is changed to reflect the fact that an event has happened, and the knowledge of when future events will occur is also updated. Since these state transitions may occur at any moment and all states remain constant from the present simulation time to the time of the earliest state transition, time can progress until the time of the subsequent earliest state change (Rowaei, 2011). At this point, the clock advances to the second occurrence (or first in the upcoming list of occurrences), the system's state is once again updated, and the future time event is determined. This procedure of advancing the simulation clock from one event time to the next is repeated until a predetermined stopping condition is met. This approach is far more prevalent than the fixed-increment time advance in DES models, although the latter is a particular instance of the former.[6] The primary difference between these two is that fixed-increment time advance allows periods of inactivity in the system. Contrary to the first approach, it does not skip over these inactive periods where no occurrences occur. The fixed-increment time advance is recognised for the equally spaced time intervals in which the simulation clock is advanced. After every clock update, the system determines whether any events should have occurred during the period.[7] This can result in a more time-consuming process for the developer (or analyst).

The next-event time advance method for a single-server queueing system is demonstrated through the following algorithm, adapted from Law (2015):

$t_i =$ time of arrival of the customer i $(t_0 = 0)$

$A_i = t_i - t_{i-1} =$ time elapsed between two arrivals of consecutive customers

$S_i =$ time that the customer i spends on server S

$D_i =$ time that customer i spends on queue (delay) before going to server S

$c_i = t_i + D_i + S_i =$ time at when customer i completes the service and leaves the system

$e_i =$ time value that the simulation clock takes to occur the event i

Assuming that the distributions of the interarrival and service times are explicit and have known cumulative distribution functions (obtained from the data available of the interest system and afterwards combined into distributions with these data using methodologies) all connected to the activities in the model, this simple simulation system is ready to run. The system begins at time $e_0 = 0$ (idle). Then, the system generates the value of t_1 (time of the first arrival) from the interarrival time distribution through a random number generation technique from a specified distribution and adds it to e_0. The simulation clock is now at instant $e_1 = e_0 + t_1$ where the first arrival occurs. The first customer finds the service inactive when they first enter the system. This prompts the customer to quickly occupy the server without waiting in the queue ($D_i = 0$) before the server's status changes to occupied. Next, the time at which the customer spends on the server, S_1, is randomly generated through a specific distribution from the service time distribution function and added to t_1. The

[6] https://web.mit.edu/urban_or_book/www/book/chapter7/7.3.html

[7] https://rossetti.github.io/RossettiArenaBook/HowDEDSClockWorks.html

simulation clock is now at instant $e_2 = e_1 + S_1$ or $e_2 = t_1 + S_1$, which marks the time that the first customer finalizes and leaves the server. The time at when the customer finalizes the service and leaves the system is given by $c_1 = t_1 + 0 + S_1$, which coincides with the instant e_2 of the simulation clock. The time of the second arrival is randomly generated through the same process as t_1 and added to it, t_2. If $t_2 < c_1$, means that the second customer arrived before the first customer has left the server. In this case, the simulation clock, which at the instant e_1, is advance to the time of the next arrival, $e_2 = t_2$ (otherwise, the simulation clock would be advanced to $e_2 = c_1$). When the second customer arrives, they are unable to instantly access the server since the server is currently serving the previous customer. The number of customers in the queue is increased to 1. The time at which this customer is going to spend on the server, S_2, is determined as soon as the server is available, at the time instant of e_3, which is also the instant at when the first customer finalizes the service, c_1. Therefore, at the instant e_3 of the simulation clock, the second customer enters the service, the queue is decreased to 0, and they finalize the service at instant e_4, considering that the third customer does not enter the system while the second one is being served. Then the time instant at when the second customer leaves the server at instant e_4 is given by $c_2 = t_2 + D_2 + S_2$ or $c_2 = c_1 + D_2 + S_2$.

3.1.2. Components of a DES

Despite the wide range of real-world systems to which simulation has been applied, DES models all share several standard components (Law, 2015). These components are also logically organised to facilitate the programming, debugging, and future modification of a simulation model's computer programme (Karnon et al., 2012). The majority of DES models that use the next-event time-advance method and are written in a general-purpose language will typically have the following elements:

- **System State:** the set of state variables required to depict the system at a certain moment.
- **Simulation Clock:** a variable displaying the current simulation time value.
- **Event List:** a list detailing the next occurrence of each event type.
- **Statistical Counters:** variables for keeping track of statistical data concerning system performance.
- **Initialization Routine:** an auxiliary programme to start the simulation model at time zero.
- **Timing Routine:** a subprogram that selects the subsequent event from the list of events and then sets the simulation clock to the time at which that event will take place.
- **Event Routine:** a subprogram that, whenever a specific type of event happens, changes the system state (there is one event routine for each event type).
- **Library Routines:** a collection of subprograms that employ probability distributions found as part of the simulation model to produce random observations.

- **Report Generator:** a subprogram that, upon simulation completion, calculates the necessary performance metrics (from statistical counters) and generates a report.
- **Main Program:** a subprogram that calls the timing routine to identify the next event and then passes control to the appropriate event routine to update the system state. When the simulation is complete, the main programme may check for termination and run the report generator.

The whole of the preceding section's explanation of how the next-time advance technique works is accomplished by executing each routine as mentioned above associated with each step.

According to Law (2015)'s description, the simulation begins at time zero when the primary programme calls the initialization routine. As a result, the system state, statistical counters, and event list are initialized. After returning control to the primary programme, the timing routine is invoked to decide which event is most imminent. For that event, the simulation clock is advanced to the time when that event will occur, and control is restored to the primary programme. Then, the primary programme invokes the event routine where typically three activities take place: (I) the system state is updated to reflect the fact that an event has occurred; (II) the statistics counters are updated to collect information about system performance and (III) the times at which future events will occur are produced and added to the event list. These future event timings are often determined by generating random observations from probability distributions.

After everything has been processed, either in the event routine or the primary programme, a check is usually conducted to see whether the simulation should now be ended (relative to some ending condition). To calculate estimates of the necessary performance metrics from statistical counters and create a report when it is time to end the simulation, the report generator is called from the primary programme. Control is returned to the primary programme if it is not yet time to terminate, and the cycle is repeated until the halting condition is finally fulfilled.

3.2. Agent Based Simulation (ABS)

An agent is an autonomous "entity" (a concept or software abstraction comparable to common programming standards such as objects, methods, and procedures) that can perceive its surroundings, including other agents, and make choices based on this data (Abar et al., 2017). Agents' behaviours are determined by their qualities and a set of simple if/then rules (Singh et al., 2016). They will need some memory to learn (get a better grasp of the state of other agents and their surroundings) and adjust their behaviour (alter their decision rules) over time. Humans, animals, automobiles, and organisations are examples of conceivable agents. An ABS is a DES in which the entities (agents) interact significantly with one another and their environment (Siebers et al., 2017).

ABS may be called a variation of DES because, in the vast majority of extant ABS, state changes occur at a countable number of time points (Law, 2015). Therefore, Agent-Based Modelling and

Simulation (ABMS) is a class of computer models that simulate the agents' dynamic actions, responses, and intercommunication protocols in a shared environment to assess their design, performance, and emergent behaviour and features (Singh et al., 2016). ABMS may be used to solve environmental, wildlife, healthcare, and financial challenges. These models have also been used to address epidemic concerns, benefiting humanity. It can mimic many features of actual disease outbreaks, and the forecasts are simple to understand, allowing disease control managers to intervene with pre-emptive vaccination measures. ABMS focuses on human diseases, acute inflammation, carcinogenic malignant tumours, wound healing, epidemiology and infection, and immunology (Robertson, 2005).

3.3. System Dynamics (SD)

SD, developed by Forrester (Forrester, 2017), is an effective method increasingly being used in a variety of domains, including the social, economic, and political sciences (Asif & Zeeshan, 2020; D. F. Andersen et al., 2009). SD models examine systems at a higher aggregate level and are used for more strategic decision-making than most DES models. The bulk of models of SD is deterministic. However, random components are conceivable (Law, 2015). It combines two unique features, qualitative and quantitative, to strengthen knowledge of a specific problem and enhance a grasp of the problem's structure and the connections between crucial variables (S. Brailsford & Hilton, 2000). Therefore, contrary to the ABS paradigm, this model focuses on simulating the system's behaviour as a whole instead of modelling the behaviours of system actors. Radzicki (2020) identifies three components for its structure: stocks, flows, and feedback loops. Stocks collect the data or material that flow into and out of them; they are sometimes referred to as "levels" or "states". The dynamics of a system are produced by the flows of information and material that enter and leave its stocks. The rate of change of a stock is determined by its net flow into or out of it. Finally, a feedback loop is created when information is sent from a stock back to its flow(s), either directly or indirectly. Feedback loops carry the amount of data (or material) gathered in a system's stocks that are sent and returned (Irwin & Wang, 2017).

Due to their ability to connect observable patterns of behaviour of a system to micro-level structures and decision-making processes, SD models are appealing in analysing energy policy issues. SD models are causal models (Qudrat-Ullah, 2012). Barlas (1989) considers the crucial step in the SD modelling process to identify the structures and decision policies contributing to a system's observable behaviour patterns.

4. The GE-DES Framework for Developing a Conceptual Model for Healthcare Systems

One of the most significant parts of a simulation project is conceptual modelling. It entails abstracting a model from a real-world system, determining what must be represented and how it will be depicted (Furian et al., 2018). Robinson (2008) provided a well-structured framework that incorporates the bulk of elements of conceptual models. It comprises the identification of objectives, input factors, responses or output measurements, and the content specification of a model (concerning its scope and the level of detail to be included). Therefore, the framework for Generalizable Discrete Event Simulation for Healthcare Systems consists of four sections: (I) understanding the issue scenario, (II) establishing the modelling objectives, (III) choosing the model content and (IV) developing the model. When applicable, the conceptual modelling components of *sections (I)* through *(III)* offer a connection to *section (IV)*'s preliminary and exploratory data analysis stages (Boyle et al., 2022). Understanding the issue scenario and identifying and determining the conceptual model's components are thoroughly crucial to the conceptual modelling process (Furian et al., 2018).

4.1. Understanding the Issue Scenario

Monks et al. (2017) identify four categories of situational knowledge: (I) study population, (II) current system performance; (III) process map of the status quo; and (IV) experimental decision factors. The information gathered for each of the four domains is intended to guide modelling operations about modelling objectives (*section 4.2.*), model content (*section 4.3.*) and model development (*section 4.4.*). For instance, investigating decision factors yields a list of possible model inputs. At the same time, the comprehensive process mapping provides a list of relevant components to help select model content. To construct an accurate model that adequately addresses the concerns posed, the modeller must thoroughly understand the topic (Robinson, 2008). This activity's method depends on how effectively consumers and subject matter experts comprehend and express the problem in determining whether DES is the best option (Boyle et al., 2022).

4.1.1. Study Population

An awareness of the types of patients treated by the emergency department (such as adults and children, as well as patients with varying degrees of urgency), as well as any particular services, should be formed (e.g., direct hospital admission for patients presenting with chest pain, resuscitation area, or long-stay observation area) (Boyle et al., 2022). All of this information should be captured in the data as a collection of characteristics, which may be used to educate patient flow paths and impact the duration of stay in the emergency department (LOS). According to Furian et al. (2018), nearly all EDs classify patients by acuity. This is often the responsibility of triage nurses, triage teams, or other medical personnel. The ESI, which consists of five stages, is the most used triage grading system in U.S.

hospitals.[8] ES1 denotes patients who require immediate life-saving interventions; ES2 marks high-risk patients who should be seen immediately; ES3 includes patients who need numerous resources or exhibit potentially life-threatening vital signs; ES4 patients require one resource; and ES5 patients require no resources. Unlike simulation research, ESI tools include testing and expert consultations. Additionally, the MTS (immediate, urgent, urgent, standard, and non-urgent) is widely the most used system for emergencies in Portugal, and the CTAS (resuscitation, emergent, emergent, urgent, less urgent, and non-urgent) is also five-grade triage systems. Various approaches use triage scales depending on the maximum time a patient should wait before seeing a physician (ATP) (Facchin et al., 2010; Duguay & Chetouane, 2007). First, there is never a miss-triage, i.e., a patient's triage category is always correct; and second, a patient's triage grade stays the same throughout their ED stay. Even though it is rare, mimicking a change in triage grade is essential for patients who destabilise and become critical. Modelling requires defining differentiated patient groups and treatment routes (Furian et al., 2018).

4.1.2. Current System Performance

Abo-Hamad & Arisha, 2013) state that the best method to determine how well things are currently doing is to consult with specialists in the area and examine the facts. For instance, domain specialists might demonstrate how well they comprehend challenges in the ED, such as when there are insufficient beds to accommodate (Boyle et al., 2022). KPIs, or Key Performance Indicators, consist of statistics derived from simulation experiment data and ought to be utilised to evaluate and interpret the current performance of the ED system. For instance, the LOS is the most popular KPI for EDs, which reflects the length of total patient stay in the ED once, including time spent in all stages (Furian et al., 2018). More specifically, it can be denoted independently for admitted patients (time elapsed between the arrival and conversion), discharged patients (arrival time to unleash time), and transferred patients (arrival time to transfer conversion time).

Furthermore, they compare and assess system performance and behaviour across various scenarios (Baesler et al., 2003). KPIs may be categorised as time-based, limit-based, state-based, financial, and combined (Welch et al., 2011). Time-based measures include summary statistics variables (e.g., mean, standard deviation (or variance) or risk measure) on durations that entities are in specific stages. This may include aggregate measurements that indicate the total durations in numerous phases and single time-based metrics that capture data only in a particular condition, such as the wait for triage and the previously discussed LOS. Other measurements commonly used are ATT which is the time that elapses until a patient is placed in a bed or room; ATP, which denotes the time between the arrival of the patient and the first time seen by a doctor; PD, which accounts for the whole

[8] https://epmonthly.com/article/on-your-mark-get-set-triage/

duration of the service provided by a physician and WT which reflects system-wide or individual waiting times (Furian et al., 2018).

Limit-based measurements track the number or percentage of patients who complete specific tasks within a specified time limit. An example could be stabilising a four-hour LOS time limit threshold for EDs, the analogous limit-based metric of the proportion of patients who leave the ED within this time frame. (Ashour & Okudan Kremer, 2013) identifies tardiness as an indicator of the ATS, i.e., if a patient's waiting to be allotted to a bed or room exceeds a specific predetermined limit. Also, LWBS (Leave Without Being Seen) represents a patient whose departure predates the triage or any uncompleted treatment procedure. Typically, this occurs while awaiting a resource, such as a bed, or being seen by a physician or a nurse. The departure may be triggered by a time restriction (from the patient's side), the model's status (such as many patients in the waiting room), or other variables.

On the other hand, state-based metrics are statistics based on the status of the model's variables across time, such as queue lengths or resource usage (Facchin et al., 2010). In addition, a typical issue often in EDs is ambulance diversion, a reasonably usual indicator indicating an overloaded model condition where no new patients are taken. Welch et al. (2011) describe it as a limit measure, i.e., the number of hours (threshold) the ambulance is in the diversion in ED. However, it may be calculated as the ratio of diversion hours to ED operational hours. Financial measures represent the cost or profit associated with a particular situation. In addition to being a meaningful KPI, it is not the primary objective of many EDs. (Paul et al., 2010) explore that several issues in EDs are based on resource utilisation and waiting times (as seen by the case studies described in the following sections). Hence performance metrics will be more focused on times and rates. Finally, as combined measures, it is possible to utilise multiple KPIs aggregated into a single KPI.

For instance, Bair et al. (2010) utilised The National Emergency Department Crowding Scale (NEDOCS), which, at any given point, combines any KPI measurement previously mentioned, such as bed occupancy; the time spent for patients to be seen by a physician upon arrival; The ratio of the number of patients in the waiting room to the total number of hospital beds for inpatients and the maximum boarding time of patients. Eskandari et al. (2011) integrated WT, resource usage, and expenditure into a KPI, which is then ranked using a technique known as the Technique for Order Preference by Similarity to the Ideal Solution (TOPSIS). Abo-Hamad & Arisha (2013) integrated WT, LOS, resource usage and layout efficiency using a balanced scorecard method. Azadeh et al. (2016) proposed the incidence of human mistakes as a KPI and examined scenarios using Stochastic Data Envelopment Analysis (SDEA). Lin et al. (2015) quantify ED performance using a crowdedness index (Furian et al., 2018).

4.1.3. Process Map of the Status Quo

This section examines a map of the healthcare processes applicable to model development. There are three phases of patient flow in healthcare: waiting room time, treatment time, and extended time (Kovalchuk et al., 2018). These data may be easily accessible from most medical information systems. They may be rendered with varying degrees of detail based on the study's aims (e.g., they can be combined as a set of particular times corresponding to different patients in a probabilistic distribution) (Abo-Hamad & Arisha, 2013). At this stage in the model's evolution, the LOS data could be divided into three portions. The first portion is the time measured between the patient's arrival at the hospital and the first time seen by a doctor. The period between the treatment commencement and the treatment completion time follows. It involves consultations with other physicians and nurses, x-rays, diagnostic testing, and further forms of treatment. The subsequent and last portion addresses the swift admissions decision, which can be dominated by extended time. The patient may then be discharged from the unit or admitted to another service (Boyle et al., 2022).

Beyond the three-phase patient flow, in the case of an ED, identifying separate functional areas, such as triage, treatment, and diagnostics, is a fundamental approach to organise an ED. With the emergence of functional areas, it is essential to establish structural and organisational areas, which are sometimes confused (Azadeh et al., 2016). Structural areas are defined as sub-units of the model that aggregate non-movable resources, such as fixed beds, cubicles, rooms, and static diagnostic equipment. Consequently, resource allocation does not alter over time. Conversely, organisational areas include temporarily aggregate resources, such as staff availability, temporary assignment to streams, and movable resources (Furian et al., 2018). This is not always the case, despite how linear and straightforward it may seem. Occasionally, structural and organisational areas may be connected with the same resources but from contrasting perspectives. For example, Saghafian et al. (2012) describe virtual streams for admitting and discharging patients who may belong to the same structural area (e.g., general ED) but are managed in separate organisational units with temporarily allotted resources.

A well-defined difference between organisational and structural areas is important to represent the complete variety of conceivable ED layouts and structures. First, published models were screened for devoted resources (beds, spaces, cubicles, registration desks, etc.) through textual descriptions, symbolic images, tabular structure, and floor layouts which changed greatly from published models. Each model's resource groupings were then assessed. The objective was to identify groupings of resources that fulfil a defined purpose or share of operations. In a subsequent phase, the authors analysed whether such resource groups constitute a particular area or be seen as diverse resources within an area (e.g., doctors with different specialities or multiple bed types within surgery) (Furian et al., 2018). Due to the modellers' degree of detail, this was not always achievable due lack of detail.

4.1.4. Experimental Decision Factors

Experimental decision factors (test analysis) involve domain expertise and primary data. Discussions could establish decision variables that can be handled (or controlled) within the system (e.g., number of physicians, nurses available, beds or room capacity). Examining resources, processes, and environmental circumstances as variables subject to change (Paul et al., 2010) is viable.

(Eskandari et al., 2011) comment that regarding the resource variables, an experiment may involve altering the number of beds, machinery or the rate at which patients are admitted. Process scenarios may be considered to introduce new approaches to enhance service quality, such as fast-track pathways or a new route for a specific patient type. The environmental changes may modify patient demand. As the studies progress, it may be convenient to acquire more information to complement those variables collected (Boyle et al., 2022). For example, to explore the improvement of a suggested novel route for fractured bones, it would be necessary to identify which patients in the historical data were classed as fractured bones.

4.2. Establishing the Modelling Objectives

For Monks et al. (2017), modelling goals specify how simulation research is intended to help client decision-making by analysing alternative system configurations based on a predetermined performance metric. Here, system configurations refer to alternate stroke route setups. Typically, configuration options are constrained by hospital money, physical space, and laws. Gunal (2012) states that modelling objectives, level of detail, and generality may be interrelated. Over a certain level of detail, further information may render a model less generic. According to Furian et al. (2018), objectives may be categorised as either general (run-time and visualisation needs, development effort and re-use flexibility) or modelling (answering the question: "what are the most important issues to be addressed by experimenting with the model?"). The general objectives are essentially associated with the model's scope, i.e., the generic and reusable content for the ED model. In contrast, the modelling objectives of chosen models may be categorised according to their primary function.

4.2.1. Identifying the Model Outputs (Responses)

Define the model outputs that are relevant to the modelling objectives. KPIs are often used as results in ED DES research (X. Zhang, 2018). Waiting time, LOS, patient throughput, bed occupancy, and the proportion of patients released without being seen are examples of KPIs. As previously mentioned, LOS is the most often used model output in ED DES models. It is employed in calculating a large number of time-based KPIs (Furian et al., 2018), such as the number of patients seen within a particular time frame.

4.2.2. Identifying the Model Inputs (Experimental Factors)

Input variables indicate many possibilities that the model may assess during experimentation. Four broad potential scenarios exist: resource-related, process-design-related, policy-related, and stochastic distribution or duration-related (Robinson, 2008). Specifically, input factor changes may be applied to the model's structure (e.g., number of resources), individual behaviour (e.g., process design), system behaviour or control (policy design), and model parameters (e.g., data-related scenarios). The succeeding terms follow a comprehensive explanation of the words and categories used, as Furian et al. (2018) define.

Human and physical resource scenarios may be further divided about resources. Physical resources include space-related resources (beds, rooms, and more) and diagnostic equipment, while human resources might vary according to employee rosters and skill levels (Bair et al., 2010).

In process-design scenarios, the introduction of new patient pathways is evaluated. Note that this does not contain alternative techniques for dispatching resources in response to activity requests if the route stays identical; policy scenarios handle these. Moreover, since they are considered data scenarios, they do not contain changes to process-step durations (e.g., shorter assessment periods).

Changes in resource allocation, such as dispatching and control rules, focus on policy-related scenarios. These include assigning personnel with varying levels of expertise and experience to jobs and patient selection procedures, such as dynamic priority (Hay et al., 2006 & Tan et al., 2012) or sophisticated triage approaches (Ashour & Okudan Kremer, 2013).

Data-related situations are those in which the input data varies throughout simulation runs, such as varying patient arrival rates and average treatment durations (Furian et al., 2018). In addition, they may be quantitative (e.g., arrival rate, number of resources or activity duration) (Mohiuddin et al., 2017) or qualitative (e.g., patient satisfaction). Following this definition, the experimental factors are a subset of the overall input data necessary for model realisation. As with the responses, the observed elements are supposed to be the mechanism by which the modelling goals will be attained. Alternatively, they may be gathered by asking the patients and domain experts how they propose to implement the required system enhancement a priori (Kuo et al., 2016). Occasionally, it is advantageous to experiment with parameters over which there is little or no control, such as the patient arrival rate. Such testing may benefit system comprehension and future event planning. In addition to identifying the experimental variables, it is helpful to specify the range within which those factors may be altered, for example, the minimum and the maximum number of staff on a shift (Robinson, 2008).

These metrics are used to represent the simulation model's output. Therefore, the simulation model will produce quantitative values of the specified performance metrics. Qualitative variables such as patient satisfaction may be connected to quantifiable indicators such as average waiting time and LOS (Abo-Hamad & Arisha, 2013).

4.3. Choosing the Model Content

In selecting model content, the framework separates a difference between (I) model scope, which identifies model boundaries by including or omitting a representation of parts of the system under investigation as model components, (II) model detail (or depth of the model) which focus on characteristics (Monks et al., 2017) and (III) identifying model assumptions and simplifications (Boyle et al., 2022). Before evaluating the extent and amount of complexity of the suggested simulation model, Robinson (2008) argues that its use should be questioned. The simulation selection is based on three key considerations: variability, interconnectedness, and complexity of the being-modelled system. Robinson (2008) also cites the use of queuing systems as the majority of operational systems may be seen as one, henceforth the primary rationale for the applicability of DES. In addition to an awareness of these reasons, the description of the issue scenario, the modelling objectives, experimental factors, and responses will guide the judgement as to whether simulation is the appropriate method. Most of the debate up to this point has yet to be particular to conceptual models for simulation. It is conceivable that a different modelling strategy will be used. The conceptual model becomes simulation-specific from this point forward.

4.3.1. Model Scope

Model scope consists of (i) stablishing the model boundaries, (ii) listing all relevant model components and determining whether each of the listed components should be included (Robinson, 2008).

Boundaries of the Model

The boundaries of the model may be experienced by changing the experimental factors of the model and analysing the responses. However, a better practice of stabilising the edges is acting with the same behaviour as the system does. As unit and facility-specific ED studies are frequently impacted by unscheduled patient arrivals and congestion in inpatient beds (Richardson & Mountain, 2009), it is convenient to model inpatient admission as a line that would not underestimate boarding times. The reason is that boarding times are affected by other attributes other than the availability of inpatient beds (Bair et al., 2010). This leads to significant uncertainty in ED operations. Levin et al. (2008) and S. Levin et al. (2011) designed an alternative approach that successfully modelled boarding times. They used the “Cox PH” model, which enabled them to work with covariances, including variables associated with competing bed demand from several sources. General ED-DES demonstrates how to model boarding times as a function of inpatient hospital characteristics using survival analysis approaches.

Different degrees of information may be employed as covariates based on the available data (Boyle et al., 2022).

Relevant Model Components

Adapting from Robinson's framework (Robinson, 2008), four categories of a component may be used to conceptualise simulation models: entities, activities (active states), queues (dead states), and resources.

Entities are presented by objects with attributes and consume resources in an activity. They are considered passive entities because they do not have the autonomy to make decisions independently. The entity has the role of moving through the system, waiting for the next activity to take place. In this case, a patient is considered an entity in the system.

Activities (or events) affect the system's state, more precisely, the entities and resources. It provokes the movement of entities along the flow and resources to cooperate. It has a finite duration at happens at a specific time. The check-in at a healthcare facility, the triage or the treatment are examples of activities in the system.

Usually, queues are passive activities that are put between two active activities. A queuing system generally exists whenever patients demand services from a facility. The duration of the queue cannot be determined a priori.

Resources are the objects in the system that provide a service to the entities. It includes staff or equipment. Instead of being depicted separately, resources are countable items. Some replacement is possible between resources and other specific components. For example, a machine might be modelled as an activity or as a piece of equipment (resource) needed to support another activity.

With more detail, human resources can be separated into four fundamental categories: doctors (or physicians); mid-level care providers (such as nurse practitioners or physician extenders); nurses and supplementary employees (such as clerks, technicians, or laboratory staff). Medical staff employees may be categorised based on their fields of specialisation and/or responsibilities, as well as their degrees of experience and proficiency (Bair et al., 2010). Typically, the range of available domains of specialisation corresponds to defining structural areas and/or patient categories, such as surgical/medical and paediatric/general. Indeed, areas such as internal medicine/surgical medicine and paediatrics need personnel with specialised training. However, they seldom are portrayed as their abilities but rather as domains of responsibility (Tan et al., 2012). Possible fields of responsibility include virtual streams for admitting and discharging patients in distinct medical and surgical areas. The model's control mechanism assigns workers to duties (temporarily or permanently). Physicians are the most prevalent sort of human resource represented in models. Most models consist of a single physician (Furian et al., 2018). However, the names of these resources may vary (e.g., emergency physician, general practitioner, or simply physician, these models are characterised by a lack of distinction between the experience and skill levels of the respective group. Mid-level care providers include physician extenders, assistants, and nurse practitioners. Special training enables them to execute minor treatments or operations that transcend nurses' competence but need doctors' supervision (not necessarily concurrently) (Robinson, 2008).

Most publications only investigated one kind of nurse in general and triage departments, although some models have floating or emergency nurses. Abo-Hamad & Arisha (2013) describes many types of nurses. However, it is still being determined whether the model differentiates them.

Administrative employees consist of clerks and receptionists (they may theoretically be nurses but do not have medical tasks in the models), paramedics, laboratory staff, technicians, and diagnostic staff are often not the heaviest burden on emergency services. Internal/surgical medicine and paediatrics demand qualified professionals (Furian et al., 2018).

The most frequent physical resource discovered models are accommodations for patients throughout their stay. Except for waiting rooms, they consist of rooms, cubicles, and hallway beds (sometimes referred to as trolley beds, stretchers or gurneys). Gurneys indicate a resource that may be shared since they relate to beds that are not allocated to rooms (even if temporarily removed from the rooms). Types of beds often correspond to distinct sections indicated in the model, such as standard and FT beds. To reduce duplication, the models only evaluated different bed types if they occurred in the same organisational area and summarised all other kinds as general or per-area beds. Triage facilities (rooms, cubicles, spaces), recliners, inpatient beds, and diagnostic equipment were also designed. Due to the absence of specific information, however, the difference between regions and resources within a region was only sometimes crystal clear. Takakuwa & Shiozaki (2004) give a floor layout that depicts a multipurpose room that is shared by several areas, while Connelly & Bair (2004) specifies the presence of ordinary beds and trauma bays within multiple sites, i.e., high and medium-acuity regions include both.

Robinson (2008) remarks that each component's influence on the model as validity, credibility, utility, and feasibility must be evaluated. If they are not required to fulfil these requirements, they should be discarded from the model without compromising its accuracy and damaging its credibility. (Robinson, 2008) advises retaining components in the model whose influence on the model's validity is uncertain.

4.3.2. Model Level of Detail

This pertains to the simulation model's depth, especially the needed degree of information for each component (Boyle et al., 2022). Determining the level of detail, i.e., the level of detail for each entity, activity, queue, and resource to be included in the model (Robinson, 2008), demands judgments on the level of detail for each component within the model scope. Unless there are trustworthy facts to guide the model parameters, it is inappropriate to incorporate specific detail (Monks et al., 2017). In terms of its properties, each component may be modelled at varying degrees of specificity. Keeping this in mind, one example is treatment time, which is an amalgamation of diagnostic tests, evaluations, therapy, and the time necessary to organise this work (i.e., waiting) (Boyle et al., 2022). This demonstrates that in certain instances, a specific activity may be subdivided into sub-activities, each of which may use resources that the others do not. Or they may use distinct experimental factors. The modeller, clients, and domain experts may go through the specifics for each component in the model scope, selecting whether the detail should be included or removed and how each detail should be modelled.

Similarly to the model scope, the choice to include or exclude a particular feature should be based on its perceived impact on the model's validity, credibility, utility, and feasibility (Robinson, 2008). Nevertheless, the possibility of having adequate data to produce reliable estimates of such parameters is minimal, necessitating an additional effort to quantify model performance uncertainty. Including many patient characteristics also presents dependency concerns (Monks et al., 2017).

4.3.3. Model Assumptions and Simplifications

According to Ranjkesh et al. (2019), assumptions are made when there are uncertainties or beliefs about the actual world being modelled, while simplifications are used to simplify the analysis as much as possible. The assumptions and simplifications may be detected mainly by referencing the model's previously excluded details and particulars. In addition to eliminating specifics, aggregating model components and substituting model components with random variables are other effective simplification techniques (Robinson, 2008). Darema (2004) & Furian et al. (2018) remark that once all assumptions and simplifications have been identified, the modeller, customers, and subject matter experts must assess their impact, based on judgement, on the model's answers (whether they are high, medium, or low) and the degree of trust that can be placed in them. Those assumptions and simplifications deemed to have a significant effect but for which confidence is low should get special consideration. The conceptual model might be modified if required to alleviate concerns over any underlying assumptions and simplifications (Vázquez-Serrano et al., 2021). The discovery of simplification options is primarily dependent on the expertise of the modeller; however, discussions between the modeller, customers, and subject matter experts may also provide simplification suggestions. Too many simplifying assumptions have been made when a simplified model can no longer accurately predict the behaviour of the actual object. In addition, it is helpful to refer to a standard set of simplifications (Robinson, 2008).

4.4. Model Development

Complex social interactions are prevalent in healthcare systems, especially at decision points. Therefore, healthcare service delivery and patient flow management issues are often difficult to characterise. Abo-Hamad & Arisha (2013)d & Arisha (2013) suggest that a deeper comprehension of the healthcare process is crucial for making proper, defensible choices and delivering successful solutions. Therefore, it is vital to develop the highlighted issue from the perspective of those directly engaged in the service delivery process.

The process of developing the model may start after formulating the problem, identifying inputs, outputs, assumptions, and entities, and after reviewing the preceding parts. (Steuben et al., 2019) defends that the quantitative data (observations) are either saved in databases or recorded on any form of storage media (in the form of records), whilst the qualitative data may be gathered by direct observation of the system and expert interviews. In healthcare, experts are hospital employees such as physicians, nurses, consultants, administrators, and managers (Baesler et al., 2003).

The hospital information system collects patient records, including information on the patient's treatment route, method of arrival, referral type, and discharge or admission time (Abo-Hamad & Arisha, 2013). Different personnel enter patient data (e.g., administrators, doctors, and nurses through the stages of patient care). Due to the stringent limits imposed on healthcare systems, hospital records often lack precision and uniformity. Therefore, data mining procedures are required to extract the most dedicated group of documents from them before extracting any data.

In the data-gathering phase, expert and clinician observations and interviews are combined to produce the inputs for the model. This offered a comprehensive understanding of numerous system problems and facets.

4.4.1. Conceptualization

After collecting data using one of the well-established modelling languages, the identified business processes are mapped into a conceptual model in which subprocesses and activities are described (Khanna et al., 2016). The control flow specification is created by defining the elements that move through the system (such as patients, personnel, and medical resources) and detailing the links between the different process phases. Finally, the resources are identified and, if necessary, assigned to the activities.

Abo-Hamad & Arisha (2013) state that the conceptual model serves two purposes in the simulation model: first, it serves as guidance for the actual simulation model, which contains and accounts for a greater level of detail, and second, it serves as a communication platform with the experts working in the existing system to validate the model.

4.4.2. Verification and Validation

Throughout healthcare simulation mode's development stages, verification and validation are part of extensive research. They should be performed to build trust and credibility in the simulation model outputs. Without them, it would be dangerous, if not catastrophic, to rely on choices or projections (Kleijnen, 1995). After each stage of model development, Abo-Hamad & Arisha (2013) argue that the model should be verified and validated in connection with the completed phases.

The model logic is verified during the verification phase to guarantee that patients follow the predicted correct treatment route. This was accomplished by using animation and monitoring intermediate output data, such as queue lengths, average processing times, number of entities that have completed the activity, number of resources available at a specific time and others. The verification methodology is much more straightforward than validation (S. C. Brailsford et al., 2019). In the verification, the analyst only requires knowing the developer's conceptual description and specifications and then confronts them with model implementation, associated data and simulation results. In validation, the analyst needs to assess the degree to which a simulation model and its related data accurately replicate the actual world from the viewpoint of the model's intended applications.[9]

4.4.3. Decision-Making

After model validation and verification, decision-makers can use the model to study the consequences of alternatives and prospective outcomes (also known as "what-if scenarios") (Abo-Hamad & Arisha, 2013). Experts and decision-makers may then evaluate and comprehend the results, which guide implementing recommended options and strategies and asset standards that represent the most outstanding performance feasible with the available resources and personnel levels (Facchin et al., 2010). Consequently, more realistic ideas and methods may be proposed and assessed using the simulation model. In addition, the simulation's capabilities may provide fascinating information on cause-and-effect relationships in performance (Paul et al., 2010). However, the sheer amount of performance indicators impedes evaluating and interpreting the simulation's results. This is because some of these needs are incompatible and violate natural laws. (Jain et al., 2011) identify that by using multicriteria decision analysis, one may evaluate the trade-off between many goals.

[9] https://www.mitre.org/publications/systems-engineering-guide/se-lifecycle-building-blocks/other-se-lifecycle-building-blocks-articles/verification-and-validation-of-simulation-models

5. The use of DES in the ED of Central Baptist Hospital

This section describes a case study developed by Wang et al. (2012) in which DES was used to enhance the performance of resources whose primary goal was to minimise the LOS of patients' stays in the unit. This was accomplished by decreasing queue wait times, increasing the number of resources, and exploring the system in a manner that could study the impact of improving these. Although the research was conducted in an ED, therefore at a different department of a healthcare unit, the whole process of preparation and analysis was comparable to what would be executed in the Imaging Department. In addition, the aims of interest are identical in minimising queue waiting times to boost patient satisfaction and demand for this research. This case study uses SIMUL8 for analysis and model development, the same software used in this research.

5.1. Problem Context

Lexington's Central Baptist Hospital (CBH) has served Bluegrass communities for over 50 years. CBH aims to reduce ED LOS to improve treatment. As one of the primary activities, a DES model is built to analyse patient outcomes, identify crucial resources and processes, perform "what-if" analysis for alternative staffing and operational situations, and provide suggestions to hospital management. Wang et al. (2012) provide a model and explain its results and insights. It is shown that adding a floating nurse, allocating varied nursing workload, combining registration and triage, concentrating on certain essential operations (e.g., IV/Med, triage, nurse access, and disposition), and requiring a physician visit in 30 minutes should drastically decrease ED LOS. The system patient flow model is presented below, in *Figure 3*.

5.2. Simulation Model

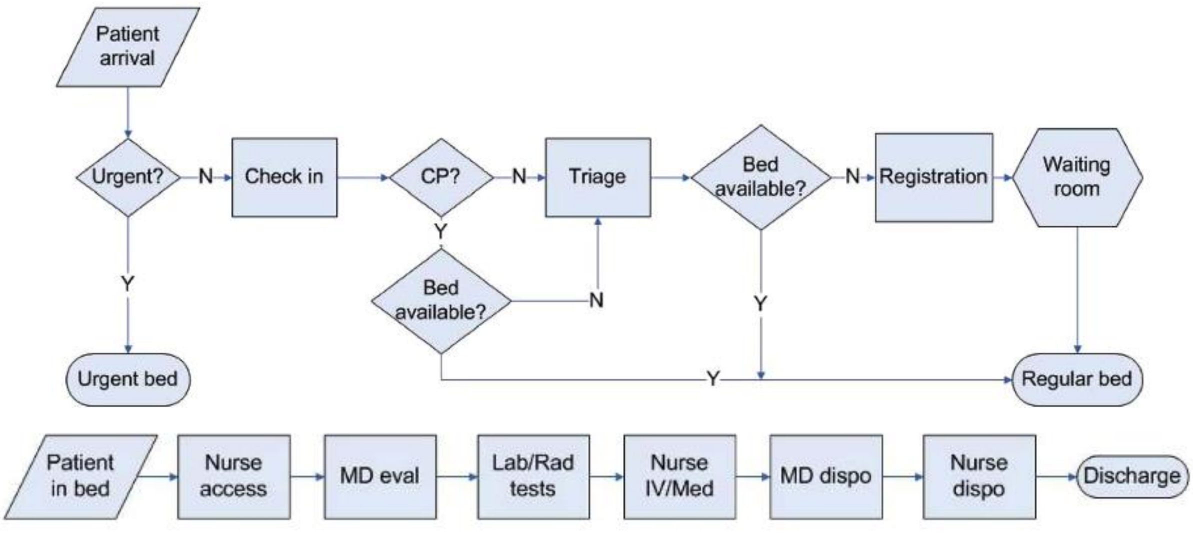


Figure 3 *– Patient movement flows in pre-bed and in-room services. Taken from* Wang et al. (2012).

This case study simulates CBH ED patient flow and suggests LOS reductions. The model should analyse resource constraints (e.g., staff and diagnostic equipment), critical operations, and management choices to reduce LOS. This requires a thorough patient flow process analysis.

A simulation model is constructed using SIMUL8 and the patient flow sketch designed above. Each procedure or operation is handled as a "machine" in this paradigm, complete with processing time, necessary resources, and routing in/out logics. The resources include patient rooms/beds, nurses, doctors, and radiological diagnostic equipment. Each part of the model was introduced, as shown below.

Patient Arrival

There are two sorts of patient arrivals: walk-ins and those brought in by ambulance. Among all patients, 0.44% of the walk-in patients and 1.8% of the ambulance arrivals represent urgent patients, all of whom have CP symptoms. The ratios relative to the acuity level of the patients are given in *Table 1A* and *2A.*

Resources

The current schedule for work shifts for both physicians and nurses are presented in *Table 3A* and *4A*, respectively. The nurses are responsible for managing a total of 21 rooms/beds. The nurses are split into a charge nurse responsible for the urgent bed, five primary nurses who supervise four regular beds, and a float nurse accountable for caring for patients in any bed when the primary nurse is occupied. A rotating schedule among the nurses ensures that such an assignment is available at any moment (i.e., the nurse number in *Table 4A* indicates job function as opposed to a particular individual). Tables 3 and 4 include information on doctors and diagnostic equipment, respectively. The information about the quantity of diagnostic equipment is presented in *Table 5A*.

Operation Times

The operation timings of service procedures are gathered by ED personnel by direct measurement or estimate. *Table 6A* provides the typical operation times associated with ED personnel. For operation times in SIMUL8, the average distribution (normal with a coefficient of variance equal to one-fourth) is employed. According to Reynolds et al. (2010), the LOS and other performance metrics are influenced mainly by the coefficient of variance rather than the distribution type.

Control Logistics

The acquired patient information determines the patient routing logic. Each patient is allocated a label that describes the patient type, acuity level, potential laboratory and radiation testing, and more, which governs the patient’s course through the ED. Visual control logic is implemented to determine whether or not criteria are met and to regulate patient flow. For instance, in inpatient testing processes, loops confirm that all relevant tests have been performed until they are done. In addition, visual logic is used to gather all required data and analyse it.

Complete Model

Using the previously stated facts, an exhaustive simulation model is constructed. The pre-bed service process is modelled to generate patient arrivals (walk-in or ambulance and arrival times), assign labels to define patient identification (acuity, CP or non-CP, urgent or not, and more), and route patients to triage and bed assignment. After pre-bed service, the patients move to the in-room service, which is composed of a set of activities, such as the nurse's initial service, physician's first visit, lab and radiological tests (as determined by patient labelling), waiting for the results, IV/Med, physician and nurse depositions, discharge, and room cleaning.

5.3. Model Validation

This section contrasts the simulation results with the accurate (practical) data. The key performance measures used are the period at bed assignment, the physician's first visit, decision, and deposition. All these activities' time added gives the patient's length of stay in this block of activities. Checking these time frames will be beneficial for future efforts to enhance the model and invalidate it. *Table 7A* displays the accuracy of these estimations. As one can see, all other results result in reasonably good accuracy, except for the duration from bed assignment to physician appointment (Bed to Physician). The physician may explain why they may not record the time right away after entering the patient's room. This model used simulations to simulate patient flow in the ED and conduct sensitivity and what-if analyses. Although the simulation model can provide information for single patients with varying acuity and variability, the study focuses primarily on the mean performance of operations for all patients, treating all patients the same way.

5.4. Sensitivity Analysis

In addition to simulating the patient flow in the ED, the simulation model will allow analysis and forecasting of possible opportunities for improvement. Using the previously mentioned model, we do sensitivity analysis, which enables to find the most crucial resource or method whose improvement may result in the most significant increase in system performance, i.e., a decrease in LOS.

5.4.1. Sensitivity to Operation Times

Wang et al. (2012) assess the possibility for improvement whether a procedure's operating time may be lowered by 10%. As indicated in *Table 8A*, Nurse IV/Med improvement might lead to the most significant decrease in LOS. Other processes, such as Triage, Nurse Access, and Nurse Discharge to Home, might also reduce LOS. Therefore, these procedures should be the focal point of the endeavour to enhance. The model assumes that, if available, the physician will visit the patient promptly following the first nurse access.

The study above addresses the sensitivity to the mean operation times for all patients. In ED, not only the mean but also the variance is significant. Due to the variety of patient types, syndromes, ages, medication records and histories, and other factors, it is normal for operation timeframes to vary significantly. Consequently, it is essential to investigate the sensitivity to various operation times. Therefore, the paper examines the effect of more significant variances (i.e., standard deviations raised by 100%, 200%, and 300%). As seen in *Table 9A*, the LOS will grow significantly when the variability is high. Consequently, regulating the variation in operating times is crucial.

5.4.2. Sensitivity to Resources

In addition to examining the sensitivity of operating times, the number of available resources is also significant. Distributing the investment to the essential resource is essential for hospital administration. Should extra physicians and nurses be hired? Should additional laboratory/radiology testing equipment be acquired? To address these concerns, sensitivity tests involving human and equipment resources are conducted to examine the effect of increasing each resource. As demonstrated in *Table 10A*, adding one extra float nurse may shorten the length of stay by more than 17%. However, more physicians or radiological test equipment do not give a good advantage.

5.4.3. Sensitivity to Patient Volume

In recent years, the number of patients attending the ED has increased. In addition, disruptive events may contribute to a rise in the number of patients. Consequently, it is essential to consider how an ED can operate with a high patient load. *Table 11A* illustrates that even with a 5% increase in patient volume, the length of stay will increase by 11%. The length of stay will increase by more than 60% for every 10% increase in patient volume. If more patients attend the ED, the number of patients will increase exponentially, making the ED untenable. It is essential to plan for enough ED capacity and personnel to address this issue.

5.5. What-If Analysis

In addition to determining the target processes or the essential resource for improvement, what-if analysis is conducted to analyse the proposed new operation management rules to provide predictable outcomes before deployment. Some possibilities are discussed below, and all of them could decrease the LOS in the ED significantly.

Wang et al. (2012) reorganise nursing positions such that the float nurse may function as a primary nurse with a more extended shift to oversee three regular beds. This makes one primary nurse continue to care for four beds while the other nurses handle three regular beds, and the charge nurse cares for one regular bed and one urgent bed. Implementing the modified timetable for nurses might lower the LOS by 24.03%. Additionally, dividing the number of beds equally among all nurses (other than the charge nurse) may reduce the usage of each nurse, which gives greater responsiveness to patients.

The study also conducts an experiment combining registration and triage, given that registration and triage always need overlapping paperwork. During triage, patients are asked to complete registration under the supervision of the triage nurse so that repetitive inquiries and paperwork may be prevented. The simulation findings indicate that such an activity might reduce the LOS by over 5%, making it a realistic means of improvement.

The delay between bed assignment and the first physician visit feels too lengthy. The model anticipates that the physician will only visit the patient after the nurses have completed their initial service. However, such a delay may be optional. It is suggested that the physician see the patient within thirty minutes (even if the nurse still needs to finish her first service). According to the simulation model, enacting such a strategy might reduce LOS by as much as 7.43%, making a significant impact.

From *Table 8A*, the four most sensitive nursing operations to improve LOS are Nurse IV/Med, Triage, Nurse Access, and Nurse Disposition to Home, showing that every 10% reduction in operation time will result in a 3% to 4% reduction in LOS. Consequently, a 10% decrease in all four procedures simultaneously indicates that it will result in a 17.37% reduction in total LOS. Wang et al. (2012) also note that combining the addition of a float nurse with the simultaneous reduction of the operating periods of the top four most sensitive operations results in a substantial 32.82% drop in LOS.

Lastly, Wang et al. (2012) find that the existing ED capacity is insufficient to react to a larger patient volume, as shown in *Table 9A*, which can be explained through the variability of patients in the ED. Therefore, the article investigates the possibility of employing a float nurse. Observations suggest that LOS is reduced by 28.25% when one float nurse (working from 08:00 PM to 00:00 AM) is added for every 10% increase in patient load. An additional 9.35% may lower LOS if the float nurse's shift is prolonged from 9:00 AM – 12:00 PM, for a total reduction of 37.6%. As a result, adding a float nurse to respond to the surge may be possible.

6. Problem Formulation, System Analysis and Conceptual Model Presentation

This chapter explains all of the information provided by the Imaging Department management at Hospital da Luz regarding the unit under study for developing the plan and subsequent conceptual model.

6.1. Imaging and Ultrasound Services Functioning

The Imaging Department of Hospital da Luz now offers approximately ten types of examinations. Mammography, X-Ray, Magnetic Resonance Imaging (MRI), Computerized Axial Tomography (CAT), and Ultrasound were cited by the decision-maker as five exams that could improve the quality of service in terms of waiting times. Currently, the department has two pieces of mammography equipment, allowing up to two patients to get an examination simultaneously. Additionally, the department has four pieces of X-ray equipment, allowing for the same number of patients, six pieces of MRI equipment, three pieces of CAT equipment, and eight pieces of Ultrasound equipment. It was stated that each piece of examination equipment is designated for a single office. Therefore the number of offices equals the number of examination equipment available. Regarding the health personnel assigned to supervise and conduct each examination, it is necessary to have one technician in Mammography, X-Ray, MRI and CAT; one assistant in X-Ray, MRI, CAT and Ultrasound; and a physician in Ultrasound.

The journey begins when the patient retrieves a ticket from one of the check-in locations. In the context of the problem, the unit has two entrances, with a check-in counter at each one. Ideally, zones A (Imaging) and B (Ultrasound) are stipulated by the examinations. The first zone, A, is characterised by Mammography, X-Ray, MRI, and CAT examinations, while the second zone, B, is determined by an Ultrasound examination. Patients are instructed to enter the area where they have exams scheduled. For example, suppose a patient is scheduled for a Mammography examination. In that case, they will enter through the entrance of A, whereas for an Ultrasound examination, they will enter through the entrance of B. A patient who has both zone A and zone B examinations scheduled will enter through one of the zones being directed to one of the waiting rooms immediately following the check-in service. Due to the particularity of the patient entering through zone A but having an examination scheduled in zone B, or vice versa, the check-in service is usually performed in the location where the patient entered. However, in this instance, the patient will move to the waiting room in the opposite area after the service.

The patient must then wait until they are summoned to the office for the examination. The duration of each examination is predetermined and corresponds to the time slot provided for each patient. This period will include completing the examination and all associated activities (preparation of equipment, instructing the patient, and filling out the medical report, if necessary, among others).

By the end of this period, the patient will have left the office to make way for the next one. For all exams besides MRI, the prescribed duration is maintained regardless of the number of examinations performed by the patient. For the second case, the specified period is relative to a single examination, and the time will be duplicated if, for instance, the patient undergoes two MRI examinations.

When the patient leaves the office, they have the option of moving to a different waiting area for another examination or leaving the unit. In the first scenario, they will proceed to the waiting area associated with the other planned examination, and the process will be repeated. As all scheduled exams of the same type are done inside the same time slot, the patient does not return to the queue they were in previously.

6.2. Objectives

Understanding the problem's objectives is crucial for the formulation phase and the simulation's direction since they will drive the whole procedure.

In this context, it is desired to identify decision-supporting solutions that reduce the average wait time of patients. This may be accomplished by increasing the number facilities under consideration while considering each resource's occupancy rate.

As stated before, the decision-maker's primary objective is to decrease patients' general average waiting time in the examination queue. However, it is possible to specify additional sub-goals that serve as a method to achieve the primary purpose and function as a kind of intermediary stage. The sub-objectives may consist in reducing the average queueing and maximum time for both check-in and examination activities separately; increasing the number of facilities inside the unit to have a more significant patient flow (fewer patients in queues) so that they may be treated more promptly to boost patient satisfaction and enhance the occupancy rate of each resource to increase their profitability.

6.3. Variables

Variables may be divided into controlled (decision) and non-controlled (exogenous). The variables that can be adjusted by the decision-maker whenever they wish are known as controllable variables and are permanent in the system.

The variables recognised as controllable in the context of the issue are the number of technicians, assistants, physicians, secretaries and facilities.

As the term indicates, non-controllable variables reflect the external environment's effect on the examined system — the healthcare unit — and are outside the decision-maker's control.

In this situation, the exogenous variables under investigation are the number of patients arriving hourly, following an exponential distribution depending on the time of the day as explicit in *Table 10* from *section 8.1.5.*; the number of patients waiting in the queues to enter a specific service at a particular time; the number of technicians, assistants, physicians and secretaries on duty at a particular time; duration at the check-in service which follows a log-normal distribution according to the *Table 11* from *section 8.1.5.* and the average and maximum time a patient stays in queue for each service.

6.4. Key Performance Indicators

After defining the problem and comprehending the intended objectives, it is necessary to establish performance metrics. These will play a crucial role in the whole duration of the project since they will permit the evaluation of the degree to which the predetermined objectives have been attained, allowing the conclusion of the system's performance. In this way, these meters must be well-defined from an early stage to evaluate the initial solution's performance.

Since the presented problem's primary objective is reducing waiting times in service queues, the defined KPIs will need to measure the quality of service of the same, similar to the pre-objectives already established. The following measures defined for this purpose are the average and maximum queueing time in the queue of each check-in service or examination and the occupancy rate of each resource. These metres will be reviewed after executing a number of the system runs to assess their performance across subsequent runs to determine if the system is trending towards deterioration or improvement.

6.5. Relevant Entities

Regarding the entities, which are the essential components of the system, it is feasible to distinguish between two sorts: temporary and permanent.

Temporary entities are the system's objects of interest, such as a patient waiting in a queue for an examination in one of the waiting rooms. Because they are considered temporary entities, they have a short life span, only countable for the period they traverse the system from the entrance to the exit point, going through the activities designed for them. Once they exit the system, they will not re-enter through the entry point during the execution time. Upon joining the system, each patient receives a unique identifier that enables the distinguishment from all other patients. Consequently, the modeller may track the course of a specific patient throughout the system.

Permanent entities refer to the resources utilised to serve temporary entities. This study identifies four entities of this type: technicians, assistants, physicians and secretaries. Each of them will be responsible for activities that need the requisite resources to be performed. The equipment necessary to complete each activity might also be considered a resource. However, since the machinery is unique to each office serving the activity, it must contain the essential equipment without sharing equipment across activities. It explains why it is not immediately applicable as a resource.

6.6. Activities

Activities represent a period with a given duration. Refers to a group of operations that modify the state of an object. Although activities and events are often used interchangeably, they have distinct meanings in the context of simulating. An event is a change in the system's state caused by an activity, with the event serving as the consequence and the activity as the cause. For example, a patient served at the ultrasound activity changed the physician's state to busy.

The system's activities are comparable to building blocks. Each structural element is accountable for the movement of the work item. In a simulation model, they may be broken down into four primary building blocks: Work Entry Point, Work Centre, Queue, and Work Exit Point.

Work Entry Point: This is the point at which each patient enters the system. The arrival pattern of work items may adhere to a predetermined schedule (deterministic behaviour) or a certain probability distribution (stochastic behaviour).

Work Centres: This is the location where staff-related activities take place. The duration of the work may be modified to conform to a specific probability distribution. The output may be routed in several ways to other objects.

Queues: This is where patients are kept while awaiting processing. The simulation analyst may manage the queue's capacity, shelf life (the maximum time a patient can be in the queue), and queue discipline.

Work Exit Point: This is the point where patients leave the system. There may be several Work Exit Points in the model. For instance, it may monitor patients who left a particular examination and went straight to the exit.

6.7. Assumptions

To design the model, it was required to define some basic assumptions that would guide the whole procedure, beginning with its building and continuing through analysis and the search for solutions. These assumptions helped to assist the creation of the model and to assume the proper conditions required for modelling although there were those which were not mentioned when the issue was formulated. The following assumptions were made:

- A patient only goes through the check-in procedure only once each day for an unlimited number of examinations.
- A patient who has records on several days will be considered as one new patient each day they attended the unit.
- Every patient exits the department having had at least one examination.
- The patient waits in the queue for an examination for as long as is required.

- The patient who enters the unit on a particular day is required to be seen on the same day and, therefore, must depart the unit on the same day, never being permitted to remain there from day to day.
- A patient is only permitted to take one examination at a time.
- The duration of each examination activity corresponds to the slot provided for the patient (and its associated resources).
- A patient leaving an examination may proceed to the exit or to another waiting queue for another examination, without ever having to return to the check-in area.
- If a patient has several examinations of the same sort, they will be done inside the same time slot with no change in duration. In the case of MRI, the activity duration is multiplied by the number of times the patient has more than one examination in it.

6.8. Sketch of Conceptual Model

From a macro perspective, the model to be developed may be divided into two sections. The first section corresponds to the time between the patients' entrance (the period corresponding to the

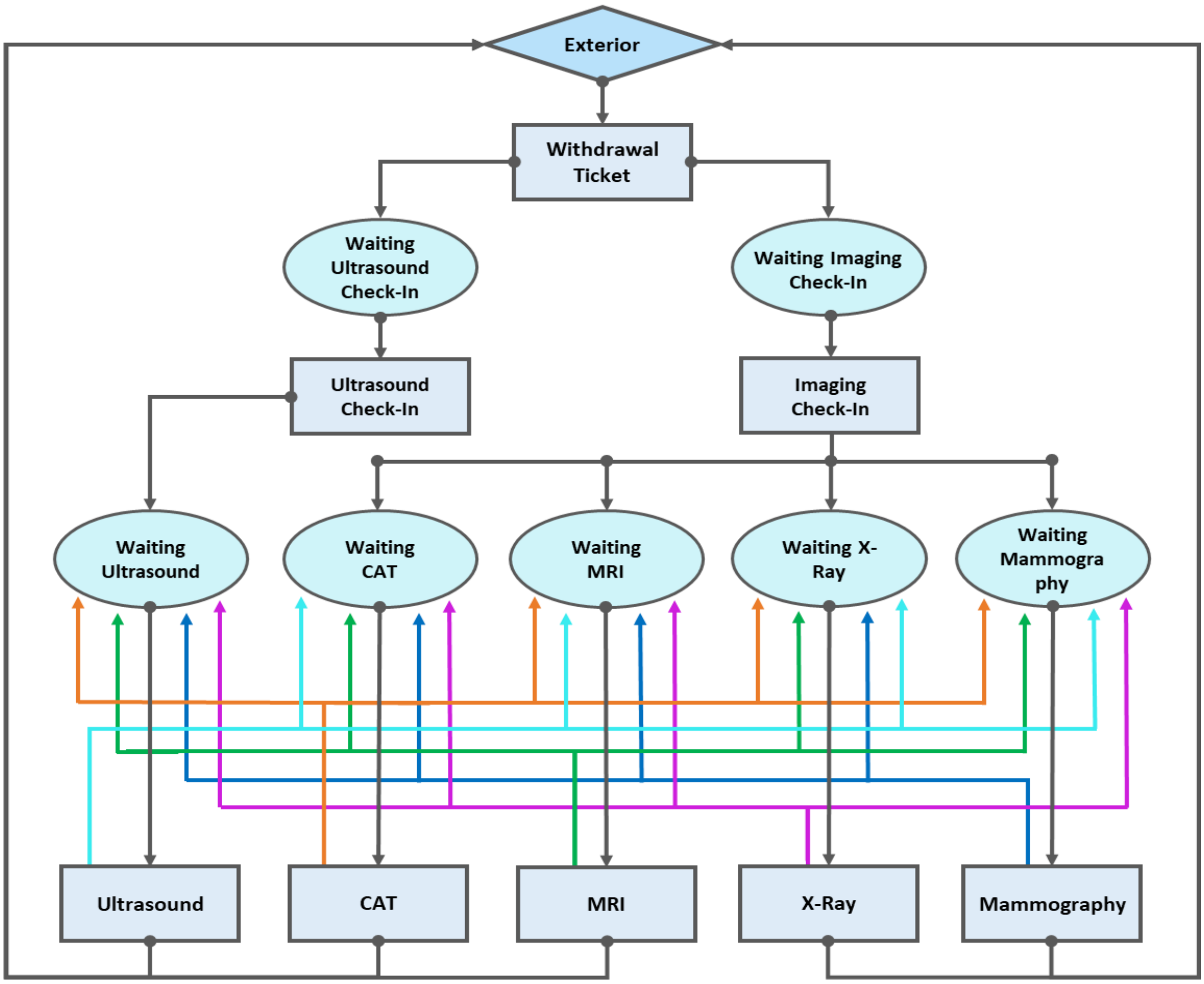


Figure 4 *– Base schema of the model that will be implemented in SIMUL8.*

withdrawal of the ticket from the kiosk) and their departure from the check-in service to join one of the queues heading to an examination. Consequently, the second section of the model addresses the remainder of the journey, from the patients' departure from the check-in until they depart from the system, with the transit of patients through various examinations as its primary focus.

In the first section of the model, when patients arrive at the unit, they are directed to either the IS or the US. In any case, there is a check-in activity service linked to each zone and a queue before each of these activities.

After abandoning the check-in activities, patients are sent to one of the queues for one of the five facilities. There are four facilities within IS, corresponding to Mammography, X-Ray, MRI, and CAT examinations, although only one facility for US, which is the ultrasound itself. After the patient has left the examination activity, they will have the option of proceeding to another queue of an examination or exiting the system. The diagram represented in the *Figure 4*, above, depicts the life cycle diagram of the model, which also contains an overview of the model to be designed in SIMUL8.

7. Data Collection and Treatment

This phase focuses on collecting data and how it was treated to implement in the conceptual model.

Before beginning the data treatment procedure, it is essential to be aware of the data privacy, and security law created and approved by the European Union, which became effective on May 25, 2018, known as General Data Protection Regulation (GDPR). The legislation imposes requirements on companies worldwide as long as they target or collect data on European Union citizens. If this law's privacy and security criteria are broken, penalties of up to tens of millions of euros may be applied. The GDPR establishes the following data protection principles[10]:

- **Legality, fairness and transparency –** Processing must be legal, fair and transparent for the data subject.
- **Purpose limitation –** Data must be processed for the legitimate purposes explicitly specified to the data subject when it was collected.
- **Data minimization –** Only data absolutely necessary for the purposes specified must be collected and processed.
- **Accuracy –** Personal data must be accurate and always up to date.
- **Storage limitation –** Personally identifiable data may only be stored for as long as necessary for the specified purpose.
- **Integrity and Confidentiality –** Processing must be done in a way that ensures adequate security, integrity and confidentiality (e.g., using cryptography).
- **Responsibility –** The data controller is responsible for demonstrating GDPR compliance with all principles.

Considering the mentioned regulation and the topic of this dissertation, it is essential to note that every clinical study must adhere to the regulation. Since the provided data does not include any meaningful information about the user, his anonymity is likewise protected, and it would not be able to directly tell them that the information about his movement inside the unit is being utilised. Therefore, it is only necessary to maintain the confidentiality and anonymity of data concerning the provider, which is the hospital board of directors in this situation.

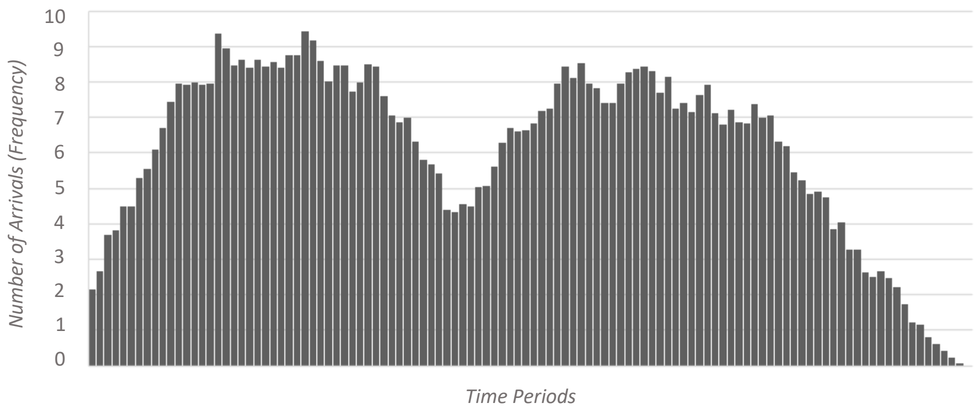
7.1. Patients Entries

The entries of the patients concern the patient's arrival at the unit depending on the time of the day. The given data included the moment each patient arrives at the unit, more specifically, the timestamp (composed by date and time) the patient receives the ticket for check-in. Since obtaining

[10] https://gdpr.eu/what-is-gdpr/

the check-in ticket is a zero-duration action, this seemed to be a viable method for defining the instant a patient enters the system. A few more data procedures enabled the arrival date to be matched with the day of the week, revealing a greater influx of patients on weekdays than on weekends.

It was revealed that the department started accepting patients at 07:30 AM each day (Monday through Friday) and continued until 8:30 PM. Therefore, because the system operates for 13 hours, it was necessary to divide it into several time intervals to analyse better the events that act in each of them and a way to maintain tracking control for when validation occurs later on. Graphic 1 presents the histogram relatively to the number of arrivals in the function of the extracted timestamps divided into 112 bins (classes), i.e., the 13 hours of labour split into 112 time periods, which gives approximately seven minutes for each time slot.



Graphic 1 *– Histogram of the average number of arrivals per time interval per day (112 bins).*

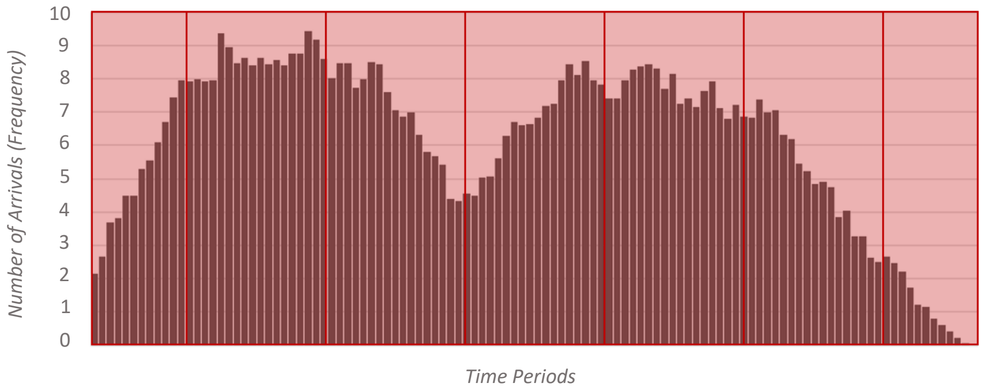
Upon the first examination of the graph, it is apparent that the number of arrivals begins to rise gradually over time, i.e., the rate of patients per time interval rises until it stabilises for some time and then drops slowly until half day of labour. In other words, the time gap between two successive arrivals decreases progressively until it stabilises, at which point it begins to climb. The rate of patients at each time interval is inversely proportional to the period between successive arrivals, as the longer this delay, the fewer patients come per interval (Ranjkesh et al., 2019). During the second half of the labouring day, the procedure is similar to that of the first half. However, the stabilisation is less pronounced since it starts to drop very slowly and then gradually intensifies, although always to a lower extent than in the first half.

Visually, the number of classes in a histogram is of tremendous importance. The greater this quantity, the greater the ability to examine how the system operates at different times of the day and the more evenly the data is distributed around the histogram, increasing the proximity to reality.[11] Additionally, with the reduction of bins, there is a growing amount of information that is lost because

[11] https://www.statisticshowto.com/choose-bin-sizes-statistics/

it gets more condensed. The procedure was executed using 50 and 10 bins to facilitate comprehension and comparison. These histogram plots are shown in *Graphic 1A* and *2A*, respectively.

On the other hand, working with several bins may be laborious and useless since it would consume a great deal of time. Therefore, balancing the retrieved data and the number of classes is essential. This might be accomplished by grouping the areas of the graphic that have typical behaviour and studying them separately. In this study, it was easy to do this separation by eye, as shown in *Graphic 2* below.



Graphic 2 – *Histogram of the average number of arrivals per time interval per day (112 bins) with the areas with a similar behaviour marked to be analysed.*

Therefore, there are seven bins, two smaller than the remaining. By further examining the data, it was possible to set proper intervals for the samples limited by each area in red. In a simplified way of doing those calculations, one can check that the first region of the graphic contains twelve bars. As indicated earlier, each bar is equivalent to seven minutes. Then, for the first region, the interval has a length of 84 minutes which might be extended to 90 minutes to equal one hour and a half. The procedure is repeated for the last region of the graphic. This way, three of the 13 hours were allotted to the intervals, leaving ten hours for five intervals or an average of two hours for each interval (or 120 minutes). This way, seven intervals were established, as shown in *Table 1*.

Each block's arrivals are segregated from the others so that each block is independent. The purpose of the following section is to assign a probability distribution to each block containing arrivals. This is perhaps the part that requires the most attention, given that the entire system will depend on the number of patients who arrive, and because it is pointless to complete any of the work if the patients do not come following reality or as close to reality as possible since this conclusion will not serve as an example. Before doing so, a thorough analysis was required. First, the given data revealed several instances in which a single patient was admitted multiple times, almost always in sequence. This may be explained, for example, by the fact that the patient collected many tickets from the ticket machine because they were not sure which one corresponded to their service the best. It was initially reported that a patient receives the check-in service once during his whole hospital stay, regardless of the number of examinations they must take. To circumvent the issue of the same patient accessing the system several times, all of the patient's records with a check-in time of fewer than two minutes were deleted. This will be discussed in *section 7.2.* that follows, concerning this activity in specific, although this criterion has been established in conjunction with department managers.

For each workday, the number of arrivals within each time slot was counted and stored in a vector comprising these numbers. Each vector element represents the number of arrivals from that time slot on a given day. For example, the vector corresponding to the first time slot (07:30 AM – 09:00 AM) could be represented as $AV_1 = \{39, 67, 4, 49, 43, 53, 60, \dots, 55, 54\}$. This means that on the first day, there were precisely 39 patients that had arrived within that time slot, 67 on the second day, and from then on. Another factor that was seen and had to be maintained is that working hours vary from day to day. There are days at which patients are not recorded at a certain hourly period for unclear reasons. Either the department was temporarily closed, or the registration was missing. And that is quite clear in the vector, AV_1, above. An entry in the vector only registered four patients from 07:30 AM – 09:00 AM, which is partially true. Even though it seems to be considerably far from the others observed, looking into that day, in particular, the department only operated for four minutes within that time slot, resuming operations at 12:00 PM, five hours later. Therefore, along with the vector containing the arrivals for each time slot, it is essential to keep the effective hours that the department had worked for each day. With this, a second vector, relative to the first time slot, containing this information (converted to decimals) could be represented as $TV_1 = \{1.47, 1.43, 0.09, 1.43, 1.33, 1.42, 1.43, \dots, 1.38, 1.40\}$. With the two vectors, it was possible to find the rate at which the patients came to the unit (expressed in patients/hour) by simply dividing the number of arrivals by the hours the department had been working during the specific time slot. Or dividing vector AV_1 by TV_1. A third vector was then emerged storing those values, $RV_1 = \{27, 47, 44, 34, 32, 37, 42, \dots, 40, 39\}$.

The circumstances to apply the Poisson process[12] to the set of arrival rates for all time slots, $RV_1, \dots, RV_7$, looked favourable. Nonetheless, each vector was subjected to an examination to determine whether or not its data were consistent with a Poisson distribution. Using SPSS Statistics software, the Chi-Squared Test with a significance level of 0.05 was then performed on each vector corresponding to each time slot, holding the block of arrival rates of that period as well as the mean.[13] It demonstrated that the test satisfied the conditions for applying the Poisson process to the dataset indicating that the arrival rates may be modelled as a Poisson process.

[12] https://brilliant.org/wiki/poisson-distribution/
[13] https://www.ibm.com/docs/en/spss-statistics/24.0.0?topic=tests-chi-square-test

At this stage, questions arose over the manner in which patients were shaped at the entryway. There are currently two methods for representing arrivals.[14] The first method relates to applying the previously estimated patient arrival rate as being the number of arrivals per unit of time. The second method considers the time interval between two consecutive arrivals, i.e., interarrival times, which corresponds to technique that SIMUL8 employs in the entry point by default. These methods are inversely related to one another. For example, if the interarrival time of patients is equal to 10 minutes, on average, one could expect a rate of $(1/10) \times 60 = 6$ patients per hour to arrive at the unit. Similarly, saying that the average patient arrival rate is equal to 20 patients per hour is equivalent to say that the average interarrival time is $(1/20) \times 60 = 3$ minutes. Therefore, one could express the formula (1) that relates this two methods as follows:

$$Interarrival\ Time = \frac{1}{Arrival\ Rate} \cdot Interval\ Length \tag{1}$$

The interarrival times of a process in which the arrival rate follows a Poisson distribution may be proven to have an exponential distribution. In order to swap between these two distributions, the technique is similar to the formula above. The inverse of the mean of a set of values that follows a Poisson process corresponds to the parameter λ of an exponential distribution. For instance, the first time slot saw an average of 38.082 patients per hour. Since the time is measured in minutes, there will be, on average, about (1/38.082) x 60 = 1.576 minutes between successive arrivals.

The following table displays the hourly patient arrival rate and conversion to the average time between arrivals for each time slot.

Table 1 – *Hourly patient arrival rate and corresponding interarrival time, in minutes, per time slot.*

# Time Slot	Interval	Arrival Rate (Patients/h)	Interarrival Time (min)
1	07:30 AM – 09:00 AM	38.082	1.576
2	09:00 AM – 11:00 AM	59.935	1.001
3	11:00 AM – 01:00 PM	50.202	1.195
4	01:00 PM – 03:00 PM	44.161	1.359
5	03:00 PM – 05:00 PM	54.244	1.106
6	05:00 PM – 07:00 PM	41.718	1.438
7	07:00 PM – 08:30 PM	16.169	3.711

The beginning of *section 7.3.* explains the services of interest for the department managers. Examinations such as Densitometry, Interventional Radiology, Orthopantomography, and others were neglected, and as a result, patients who were incorrectly classified as arrivals had to be removed. It should be highlighted that a patient is only deleted from the system if they only have rejected records associated with one of these examinations. If the patient shares entries with another examination of

[14] https://blog.simul8.com/simul8-tip-whats-the-difference-between-arrival-rates-and-inter-arrival-times/

interest, only that patient's entry for that examination will be considered. Therefore, the rate of patients per hour in each time slot had to be readjusted. Using Excel, it was possible to determine that 4,378 patients did not meet the criteria mentioned above and would consequently be deleted from the system. Knowing that the data extended for 124 days, it was easy to calculate the average number of patients removed each day, i.e., 35.306 patients/day. As arrivals occur for 13 hours a day, it was feasible to determine the hourly average number of dismissed patients. i.e., 2.716 patients/hour. With this information, each rate of patients per hour had to be removed in 2.716 units. Therefore, the new rates are according to the following table.

Table 2 – *New rate of patient per hour and interarrival time, in minutes.*

# Time Slot	Interval	Arrival Rate (Patients/h) (new)	Interarrival Time (min) (new)
1	07:30 AM – 09:00 AM	35.366	1.697
2	09:00 AM – 11:00 AM	57.219	1.049
3	11:00 AM – 01:00 PM	47.486	1.264
4	01:00 PM – 03:00 PM	41.445	1.448
5	03:00 PM – 05:00 PM	51.528	1.164
6	05:00 PM – 07:00 PM	39.002	1.538
7	07:00 PM – 08:30 PM	13.453	4.460

Table 10 from *section 8.1.5.*, contains information on the exponential distribution of each slot.

After defining the time slots and calculating the average interarrival time for each, the proportion of patients who proceeded to IS or US was determined within each time slot. The results are presented in the table below, in *Table 3*.

Table 3 – *Proportion of patients that proceeded to one of the services per time slot in the moment of arrival.*

		Imaging Service		Ultrasound Service		
# Time Slot	Interval	Number of Patients	%	Number of Patients	%	TOTAL
1	07:30 AM – 09:00 AM	5,202	81.192	1,205	18.808	6,407
2	09:00 AM – 11:00 AM	11,167	77.792	3,188	22.208	14,355
3	11:00 AM – 01:00 PM	9,377	77.856	2,667	22.144	12,044
4	01:00 PM – 03:00 PM	8,872	83.572	1,744	16.428	10,616
5	03:00 PM – 05:00 PM	10,072	77.009	3,007	22.991	13,079
6	05:00 PM – 07:00 PM	7,808	77.963	2,207	22.037	10,015
7	07:00 PM – 08:30 PM	2,456	92.714	193	7.286	2,649

In *section 8.1.5.*, it is described how these percentages were implemented at the *(F) Shift* activity.

7.2. Check-In Service

In addition to the patients' entrance timestamps in the unit, each patient's admission duration at check-in was also accessible.

As indicated in the preceding section, all check-in admission durations of less than two minutes were deemed outliers and were eliminated with the agreement of department managers once they most likely referred to a patient who had previously been registered into the system.

When making the first analysis of these durations, one might verify that the bulk of these durations falls between five and 10 minutes. Moreover, when looking into more detail, some durations between 20 and 50 minutes and even higher were observed. These values are a little out of place. They might be related to the fact that the system has considered the period during which an employee was absent from the workplace after finishing his shift with the previous patient and before returning to the workplace to resume their job. Other possibilities would have been overlapping durations of different patients or the count of the dead time between when a worker finalizes admitting one patient and when another patient comes considerably later. This second possibility is possible during periods with a lower overall number of patients.

Consequently, it was necessary to take precautions to ensure that these values were not taken into consideration to avoid producing disruptive outcomes. Subsequently, it was necessary to detect and eliminate outliers from the dataset. In this research, statistical principles such as the computation of quartiles and the upper limit calculation were used (the lower boundary was not calculated since there are only acceptable values above two minutes).[15] The dataset's first quartile (25th percentile) was computed along with the third quartile (75th percentile). The values of these two quartiles were used to calculate the inner quartile range, which is the difference between the two quartiles, i.e., $IQR = Q3 - Q1$. The upper bound limit specifies the maximum value that can be accepted before it is deemed an outlier. Its calculation is provided by the formula (2):

$$UB = Q3 + (1.5 \times IQR) \tag{2}$$

This resulted in about 19 minutes, the maximum allowable length for the check-in service. Nonetheless, this value was presented to department managers who believed it to be an acceptable threshold.

Now that the check-in durations range between two and 19 minutes, the last stage was to identify a probability distribution that could accurately represent the admission durations for this service. To accomplish this, the dataset was submitted to the fitting distribution feature inside the SPC for Excel software to determine the optimal distribution for the data.[16] Consequently, the time set

---

[15] https://www.absentdata.com/how-to-find-outliers-in-excel/
[16] https://www.spcforexcel.com/knowledge/basic-statistics/deciding-which-distribution-fits-your-data-best

followed a log-normal distribution with a mean of 7.527 and a standard deviation of 3.648. The corresponding distribution was saved and used later, as explained in *section 8.1.5.*, related to this activity.

7.3. Imaging and Ultrasound Services

The provided data also contains information about the specific services each patient went through throughout their time in the unit. There was a total of five services that department managers are interested in the study, which consists of Mammography, X-Ray, MRI, CAT and Ultrasound. This step aims to determine the probability that a specific patient will use each service. These calculations are crucial because they will enable the simulation to determine whether the patient will choose one route or another depending on the actual use of each service. For instance, if the X-Ray service got a more significant number of patients who used the service, then the probability that a particular patient would follow this route is higher than the other services.

The first step was to eliminate the data of patients who had received services not relevant to the research, as previously explained at the end of *section 7.1.*.

Hereupon, using Excel tools, the total number of individuals that entered the hospital for examinations and the number of patients who specifically used each service was counted. It is important to note that there are records in which the same patient went through the same service more than once. This patient is counted just once for the service in question to preserve consistency. If this were not the case, there would be much more patients in circulation than recorded as arrivals. For example, if a patient has done three X-Rays, they will be considered as one patient who visited the X-ray service for the examination. The scenario of counting two or more exams at the same service on the same patient is discussed further in this section.

Another observation that was taken into account was that the same patient underwent several services during their stay. This is only applicable for the same day, as the visit does not extend from one day to the next. The objective is also to determine the proportion of patients who move from one service to another without leaving the unit. Otherwise, it will be counted as two arrivals. With this in mind, a table was arranged in ascending day and month order, with each row representing a patient and each column representing one of four services. Each cell in the table described the number of times each patient used that particular service during their stay.

The total number of patients that entered the hospital for any examination was 56,045. In terms of the overall number of days that were examined, a total of 32,254 patients entered the IS, which is about 58% of the general patients. The IS is divided into four facilities, which received 17,885 X-Ray patients, 9,627 MRI patients, 11,119 CAT patients and 6,102 Mammography patients. Consequently, the percentage of patients who underwent each of these facilities might be estimated by dividing the number of patients that entered each facility by the total patients that entered the IS, being approximate 55%, 34%, 30% and 19% for X-Ray, CAT, MRI and Mammography, respectively.

The same thing might be done for the US. During the period under review, 23,991 patients used this service, which is about 42% of the overall patients. The table below shows the percentage of patients who attended each one of the five possible examinations.

***Table 4** – Percentage of patients in each examination.*

	Mammography	X-Ray	MRI	CAT	Ultrasound
% Of Patients	19	55	30	34	42

Since the percentage of usage of each Imaging facility differs from the one in the US, it was required to adjust them to the same denominator, i.e., the overall patients. This resulted in 11%, 32%, 17%, 20% and 42% of usage for Mammography, X-Ray, MRI, CAT and Ultrasound, respectively. These percentages are summarised in *Table 5*.

Two different scenarios were studied. The patient may have several examinations within the same service in the first scenario. The second scenario concerns the slight chance that one patient will have examinations in two different services. Note that these scenarios may be combined. For example, one patient could have two examinations in the MRI service and one in the CAT service, i.e., more than one examination within the same service and another in a different service. In the latter scenario, it was necessary to aggregate all two-by-two combinations amongst all services regarding the proportion of patients who used these two services.

For the first scenario, in the US, 13,077 of the total number of patients (23,791) sent to this service had just one examination, while the remaining 10,714 had two or more. This corresponded to 55% and 45%, respectively.

For the X-Ray service, 13,171 out of the total patients (17,885) sent to this service had just one examination, while the remaining 4,714 had two or more. This corresponded to 74% and 26%, respectively.

For the CAT service, 7,508 of the total patients (11,119) sent to this service had just one examination, while the remaining 3,611 had two or more. This corresponded to 68% and 32%, respectively.

For the MRI service, 6,506 of the total patients (9,627) sent to this service had just one examination, while the remaining 3,121 had two or more. This corresponded to 68% and 32%, respectively.

For the Mammography service, 5,753 patients (6,102) sent to this service had just one examination, while the remaining 349 had two or more. This corresponded to 94% and 6%, respectively.

With this information, it was feasible to construct *Table 5*, which would compile all future-useful data and the percentages of a patient making more than one examination at the same service.

Table 5 *– Number of patients per service and its percentage of usage in the system.*

Number of Patients	Mammography	X-Ray	MRI	CAT	Ultrasound
Examinations = 1	5,753	13,171	6,506	7,508	13,077
Examinations > 1	349	4,714	3,121	3,611	10,714
TOTAL	6,102	17,885	9,627	11,119	23,791
% Of Patients (Examinations = 1)	94	74	68	68	55
% Of Patients (Examinations > 1)	6	26	32	32	45
Total of Patients	56,045				
% Of Usage	11	32	17	20	42

For the second scenario, all combinations of two-by-two services were included with the exact number of patients who underwent these two services. Even though a significant proportion of patients have had examinations in more than two services, e.g., X-Ray, CAT, and Ultrasound, this probability is still considered. In this instance, the patient was tallied in X-Ray & CAT, CAT & US, and X-Ray & Ultrasound pairs. The results are shown in *Table 6*.

Table 6 *– Number of patients who used two different services according to the combined pairs.*

# Combination	Services	Number of Patients	Total Patients	Intersection Probability
1	**X-Ray & MRI**	1,366	56,045	0.024
2	**X-Ray & CAT**	728		0.012
3	**X-Ray & Mammography**	531		0.009
4	**X-Ray & Ultrasound**	3,078		0.055
5	**MRI & CAT**	332		0.006
6	**MRI & Mammography**	63		0.001
7	**MRI & Ultrasound**	311		0.006
8	**CAT & Mammography**	185		0.003
9	**CAT & Ultrasound**	1,029		0.018
10	**Mammography & Ultrasound**	5,917		0.106

The information in the table above made it possible to calculate the probability of the intersection of patients who underwent each pair of services. Those probabilities are expressed in the last column of the *Table 6*.

Similarly, it was conceivable to determine the chance of a patient who had just had one examination at a service leaving the unit permanently. These probabilities are expressed in the *Table 7.*

Table 7 *– Number of patients who had just examination(s) at one facility.*

# Combination	Services	Number of Patients	Total Patients	Intersection Probability
1	**X-Ray & Exit**	12,985	56,045	0.232
2	**MRI & Exit**	7,765		0.139
3	**CAT & Exit**	9,238		0.165
4	**Mammography & Exit**	161		0.003
5	**Ultrasound & Exit**	14,432		0.258

By making use of the individual probabilities of the usage of each service as well as the probabilities of the intersection of the usage in two different services, one was able to arrive at the calculation of all conditioned probabilities that will serve as input data for the simulator. What is important to know is the likelihood that a certain patient may switch services. Mathematically speaking, one wants to determine the chance of a patient, for instance, going to a Mammography after having had an X-Ray. Since these probabilities cannot be determined just by their intersection, since the likelihood of one event may vary depending on the occurrence of another event first, it is reasonable to evaluate the chance of occurrence of event A given the occurrence of event B. This indicates that the chance of a patient leaving a Mammography for an X-Ray may vary from the probability of a patient leaving an X-Ray for a Mammography, for example.

The definition of the conditional probability, given two events A and B from a sample space S, is denoted by the following formula (3):

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \tag{3}$$

in which $P(A \mid B)$ is the probability of occurrence of event A given event B, $P(A \cap B)$ is the probability of the intersection of those two events and $P(B)$ is the probability of occurrence of event B.

Analogously to how the probability of intersection between all conceivable pairs of services were computed, *Table 8* displays the conditional probabilities of each of these pairings in a more intuitive form that have previously been obtained using the formula shown above. It should be emphasized that, except for services intersecting with the exit, two conditional probabilities will exist between each pair of services.

__Table 8 –__ Conditional probabilities of a patient leaving one service and entering another/leave the unit.

1st Examination	2nd Examination	Intersection Probability	Conditional Probability	% Of Patients
X-Ray	MRI	0.024	0.075	8
	CAT	0.012	0.038	4
	Mammography	0.009	0.028	3
	Ultrasound	0.055	0.172	17
	Exit	0.232	0.719	72
MRI	X-Ray	0.024	0.142	14
	CAT	0.006	0.035	4
	Mammography	0.001	0.006	1
	Ultrasound	0.006	0.035	4
	Exit	0.139	0.808	81
CAT	X-Ray	0.012	0.061	6
	MRI	0.006	0.030	3
	Mammography	0.003	0.015	2
	Ultrasound	0.018	0.091	9
	Exit	0.165	0.833	83
Mammography	X-Ray	0.009	0.082	8
	MRI	0.001	0.009	1
	CAT	0.003	0.028	3
	Ultrasound	0.106	0.972	97
	Exit	0.003	0.028	3
Ultrasound	X-Ray	0.055	0.130	13
	MRI	0.006	0.014	1
	CAT	0.018	0.042	4
	Mammography	0.106	0.250	25
	Exit	0.258	0.608	61

8. Simulation Model Development

Once gathered, data (from observations and interviews with experts) were combined to conceive the conceptual model using SIMUL8, where processes and activities could be defined. The conceptual model guides the actual simulation model, which contains more system detail. Besides that, it serves as a communication mechanism for validating the model. The implementation of the model in SIMUL8 is shown in *Appendix 4*.

8.1. Explanation of System Modelling

This section will describe the full process of modelling the system in the software by going over the many entities and components that act in the design and how they all interrelate. At this stage, it is intended to use all information gathered and processed in the past as inputs for each component.

8.1.1. Simulation Clock

Before beginning the simulation, it was important to establish the system's unit of measurement. Given that most of the information provided is presented in minutes and that the pre-established performance metrics also use this time unit as a reference, it was determined that the minute would be the most appropriate unit of measurement for evaluating the acquired outcomes.

Concerning the simulation's running hours on working days, according to the information provided, it was specified that the simulation period begins each day at 07:30 AM and ends at 09:00 PM. The duration of each working day is 13 hours and 30 minutes.

8.1.2. Arrival Shifts

A total of seven shifts were created with the weekly pattern option enabled. The shifts created follow the same logic as the time intervals of the arrival time slots, so a specific shift covers each time slot, as presented in the *Table 9*, below.

Table 9 *– Start time and end time of each shift.*

# Time Slot	Interval	Shift
1	**07:30 AM – 09:00 AM**	*S_07:30h_09h*
2	**09:00 AM – 11:00 AM**	*S_09_11h*
3	**11:00 AM – 01:00 PM**	*S_11h_13h*
4	**01:00 PM – 03:00 PM**	*S_13h_15h*
5	**03:00 PM – 05:00 PM**	*S_15h_17h*
6	**05:00 PM – 07:00 PM**	*S_17h_19h*
7	**07:00 PM – 08:30 PM**	*S_19h_20:30h*

These changes were only used to split the patients by the moment they arrive at the system.

8.1.3. Resources

Throughout the simulation, the resources serve as methods that the system retains and employs for numerous actions. An activity that relies on them necessitates using personnel (or equipment not exclusive) to this activity alone. In general, a resource is shared by several activities, with the others on active standby until it becomes available for use by another activity. Technicians, assistants, physicians and secretaries will be able to comprehend the four distinct and potential resources described in this research for application in examination-related tasks.

Therefore, four resources were created: *R_Technician*, *R_Assistant*, *R_Physician* and *R_Secretary*, each with a separate function. However, knowing their roles in the system is irrelevant to the problem. The only essential information is to understand how many resources of each type are required for each activity that depends on them and when they are no longer needed.

8.1.4. Labels

Labels allow for adding more control to the system. Labels can be attached to work items when they enter the simulation or at any point. The label's value can be tested and/or changed at any workstation.

In this context, the labels will serve as a trail for each patient, indicating the latest examination they passed as well as the other tests where they have gone, with the latter incrementing the value of the exam-specific label each time the patient enters this activity. Since the patient should not re-enter the queue for an examination in which they have previously participated, this will also function as a debug to determine whether the patient has repeated an examination. Using labels, conditions will be set that prevent these incidents from occurring. The following labels were created to achieve this: *L_Exam_Mammo*, *L_Last_Exam_Mammo*, *L_Exam_XRay*, *L_Last_Exam_XRay*, *L_Exam_MRI*, *L_L ast_Exam_MRI*, *L_Exam_CAT*, *L_Last_Exam_CAT*, *L_Exam_Ultrasound* and *L_Last_Exam_Ultrasound*.

8.1.5. Activities

Listed below is a description of how each activity was implemented in the software following the order designed in the model presented at the beginning of *section 7.2.*.

Additionally, fictitious activities do not fall inside any of the categories introduced in *section 6.7.*. They helped identify certain situations with them, such as sorting patients by arrival time or examination. These activities are regarded as having a fixed duration equal to zero. In the designed model, all these fictitious activities are denoted by an *(F)* at the label's start. At the same time, activities (or entry points) that have some fixed or variable duration associated with a time distribution are denoted by an *(A)*; the queues and exit points are represented by a *(Q)* and *(E)*, respectively. To avoid work items from staying in the system during the day-to-day transition, it is essential to establish a clean-up of the work items from the system when this transition happens using visual logic.

- ***(A) Arrivals***: This activity coincides with the moment patients take their tickets from the kiosk at the checkpoint. This is where all information on patient arrivals must be stored, notably the temporal distributions, since the programme expects a distribution that specifies the time between patient arrivals, as previously stated. This activity is time-dependent, once composed of several smaller distributions that vary according to the time of the day. Therefore, the modeller must specify the slots each distribution wants to be related to. These small distributions are regulated to adhere to an exponential probability distribution (stochastic behaviour). *Table 10* below shows the distribution relative to each time slot. The variables $\chi_1 + \chi_2 + \cdots + \chi_n$ represent the intervals between each consecutive arrival within the time slot.

***Table 10** – Exponential distributions that comprise the overall distribution for the (A) Arrival activity.*

Activity	Interval	Distribution
(A) Arrivals	**07:30 AM – 09:00 AM**	$\chi_1;\ \chi_2;\ \dots\ \chi_n \sim Exp(1.697)$
	09:00 AM – 11:00 AM	$\chi_1;\ \chi_2;\ \dots\ \chi_n \sim Exp(1.049)$
	11:00 AM – 01:00 PM	$\chi_1;\ \chi_2;\ \dots\ \chi_n \sim Exp(1.264)$
	01:00 PM – 03:00 PM	$\chi_1;\ \chi_2;\ \dots\ \chi_n \sim Exp(1.448)$
	03:00 PM – 05:00 PM	$\chi_1;\ \chi_2;\ \dots\ \chi_n \sim Exp(1.164)$
	05:00 PM – 07:00 PM	$\chi_1;\ \chi_2;\ \dots\ \chi_n \sim Exp(1.538)$
	07:00 PM – 08:30 PM	$\chi_1;\ \chi_2;\ \dots\ \chi_n \sim Exp(4.460)$

Excluding arrival distributions, all labels previously defined in *section 8.2.4., i.e., L_Exam* and *L_LastExam*, were set to zero to indicate that the patient has not yet entered or left any examination.
Since there are two checkpoints (and therefore two entries), each representing the queried service, only a single entry for all patients was considered when designing the system. The duty of dividing patients according to the entry zone will be detailed in the next activity.

- ***(F) Shift* and *(F) Sep***: Depending on the arrival time, patients will be transferred immediately to one of seven fictitious activities that reflect the seven shifts (slots) the system uses to operate arrivals. Consequently, each shift created in *section 8.1.2.* had to be assigned with each of these activities of the *(F) Shift* type, implying that each activity will only take patients during its assigned shift. This was necessary after accommodating the percentage of patients that take either the Imaging or the Ultrasound path since percentages vary from slot to slot. For this reason, each *(F) Shift* activity is linked to two fictional activities, i.e., *(F) Sep_I* or *(F) Sep_US*, with the percent mode enabled in the routing out of the first. The input values to each *(F) Shift* activity are presented in *section 7.1.*, *Table 3*. Each *(F) Shift* activity is labelled with an *(F)* followed by *Shift* and the start and end timings of the shift, e.g., *(F) Shift 09h_11h*. In the latter case, *(F) Sep* can be either *(F) Sep_I* or *(F) Sep_US*, forcing all patients who engage in one of these activities to undergo Imaging or Ultrasound, respectively, according to the routing out percentages.

- ***(Q) Wait_Check_In*:** The first queue the patient encounters after obtaining the ticket at the checkpoint is the wait to be summoned to the service to begin the check-in procedure. There are a total of two queues of this kind, *(Q) Wait_Check_In_I* and *(Q) Wait_Check_In_US*, one for each zone, each of which is connected to one of the immediately preceding activities, *(F) Sep_I* and *(F) Sep_US*, respectively, and therefore, only pass patients designated for the respective services. The capacity of the queue is infinite, i.e., there is no limit to the number of patients in the line waiting to be seen for the next activity, and no shelf life has been specified, i.e., patients are required to remain in the queue for as long as it takes before they may proceed to the next activity.

- ***(A) Check_In*:** This procedure involves the staff confirming the patient's data and notifying the system that the patient will conduct the day's planned examinations. The log-normal distribution is employed for the duration of both activities, as determined in *section 7.1.2.* and reported below. The variables $\chi_1 + \chi_2 + \cdots + \chi_n$ represent the set of durations for the activity.

Table 11 – *Log-normal distributions for the (A) Check_In activities.*

Activity	Distribution
(A) Check_In_I	$\chi_1 + \chi_2 + \cdots + \chi_n \sim Log\ Normal(7.527, 3.648)$
(A) Check_In_US	

Similar to the preceding activities, each queue, *(Q) Wait_Check_In_I* and *(Q) Wait_Check_In_US*, is linked to an activity, *(A) Check_In_I* and *(A) Check_In_US*, respectively.
The resources and their rules related to these activities are detailed in the *Table 12*.

Table 12 – *Resources and their specific rules for the (A) Check_In activities.*

Activity	Resources	Rules
(A) Check_In_I	*R_Secretary*	1. Require recourses before collecting any patients. 2. Release resources as soon as the task complete. 3. Require and release resource here.
(A) Check_In_US		

- ***(F) Go_Wait_Exam:*** This fictitious activity portrays the patient's movement from leaving the check-in desk until they enter one of the examination waiting rooms. In reality, a patient may be served at the check-in relating to a service and wait in an opposite waiting room of the other service. This scenario is estimated to have a 15% chance (as determined by department managers). This explains why two pathways lead from each check-in activity to this sort of activity. Two activities, *(F) Go_Wait_Exam_I* and *(F) Go_Wait_Exam_US*, relate to separate services as usual. According to the scenario described above, for instance, 85% of the patients who travel to *(F) Go_Wait_Exam_I* activity are derived from the *(A) Check_In_I* activity, 15% from the *(A) Check_In_US* activity. Therefore, it was defined in the routing out dialogue of the *(A) Check_In_I* activity as the percent option with an 85% probability of a patient following the path to *(F) Go_Wait_Exam_I and 15%* to follow the other route.

Additionally, the patients who leave the *Go_Wait_Exam_I* activity have four alternatives for distinct pathways, which correlate to the four waiting rooms for the relevant examinations. From *Table 4* from *section 7.3.*, it is possible to directly extract the probability of a patient going to one of the four facilities within the IS. It should be noted that the values of these percentages use the number of IS patients as the common denominator and not the overall number of patients since the chance of a patient travelling to the ultrasound waiting room is null in this activity in particular. In this instance, the values of the percentages provided in the routing out dialogue for this activity correspond to the ones presented in *Table 4* from *section 7.3.*.

Conversely, patients who pass through the *Go_Wait_Exam_US* activity go straight to the next queue.

- ***(Q) Wait_Exam*:** This is the second and last relevant queue for the study. There are a total of five queues of this kind, each preceding the appropriate examination-related activity. Hence, the queues *(Q) Wait_Exam_Mammo*, *(Q) Wait_Exam_XRay*, *(Q) Wait_Exam_MRI*, and *(Q) Wait_Exam_CAT* precede the activity *(F) Go_Wait_Exam_I* activity, based on the percentages mentioned in its description above, and the queue *(Q) Wait_Exam_US* that precedes the activity *(F) Go_Wait_Exam_US.* For the same reason as the *(Q) Wait_Check_In* queues, these are also specified as having an unlimited capacity, and shelf life equal to zero.
- ***(A) Exam*:** There are five activities of this type, *(A) Exam_Mammo*, *(A) Exam_XRay*, *(A) Exam_MRI*, *(A) Exam_CAT* and *(A) Exam_Ultrasound*, each one succeeding the corresponding queue. This activity represents each examination and has a defined duration in minutes, i.e., it follows a fixed distribution equal to the examination's duration. The distribution utilised for each activity is shown in *Table 13*, in which the variable corresponds to the duration of that examination.

Table 13 – *Distribution of the duration for the (A) Exam activities.*

Activity	Distribution
(A) Exam_Mammo	$\chi_1 \sim Fixed(15)$
(A) Exam_XRay	$\chi_1 \sim Fixed(10)$
(A) Exam_MRI	$\chi_1 \sim Fixed(40)$
(A) Exam_CAT	$\chi_1 \sim Fixed(20)$
(A) Exam_Ultrasound	$\chi_1 \sim Fixed(15)$

The duration of each examination corresponds to the time slot provided for a patient, independent of the number of examinations the patient will conduct, except for RMI, for which a patient would need twice as much time if, for example, completing two exams. However, this situation will be addressed later in this section.

Table 14 – *Quantity of resources and their specific rules for the (A) Exam activities.*

Activity	Resources	Rules
(A) Exam_Mammo	1 *R_ Technician*	1. Require recourses before collecting any patients. 2. Release resources as soon as task complete. 3. Require and release resource here.
(A) Exam_XRay	1 *R_Technician* 1 *R_Assistant*	1. Require recourses before collecting any patients. 2. Release resources as soon as task complete. 3. Require and release resource here.
(A) Exam_MRI	1 *R_Technician* 1 *R_Assistant*	1. Require recourses before collecting any patients. 2. Require here, but do not release the resource.
(A) Exam_CAT	1 *R_Technician* 1 *R_Assistant*	1. Require recourses before collecting any patients. 2. Release resources as soon as task complete. 3. Require and release resource here.
(A) Exam_Ultrasound	1 *R_Physician* 1 *R_Assistant*	1. Require recourses before collecting any patients. 2. Release resources as soon as task complete. 3. Require and release resource here.

Each of these activities also redefines the value of the labels according to the following rule. For the activity under analysis, the value of the *L_Exam* label is incremented by one unit to indicate that the patient took an examination of that type. In contrast, the value of the *L_Last_Exam* label is set to one to tell that this was the patient's most recent examination, along with other labels of this type set to zero.

The resources are essential to all activities of this type since all these activities depend on the number of available workers to be performed. Each activity utilises the stated resources, although only some adhere to the same discipline. *Table 14* highlights the resources necessary and application guidelines for each activity.

(F) Dispatch_MRI: This fictitious activity is intended to direct the patient that finishes the *(A) Exam_MRI* activity to either repeat it or move to the subsequent one it is linked to, i.e., *(F) MRI_Other_Exam*. As it is the only activity in which the time is dependent on the number of examinations performed, there is a probability, which may be found in *Table 5*, *section 7.3.*, that a patient will be able to undertake a second examination after completing the first. Therefore, in the routing out dialogue of this activity, it is defined as a probability of 32% for a patient to return to the (A) Exam_MRI activity and a 68% probability to proceed to *(F) Other_Exam_MRI*. For a patient who attends this activity and desires to repeat the examination, it is crucial to set his precedence over patients in the queue *(Q) Wait_Exam_MRI*. To avoid competition between these two patients travelling to *(A) Exam_MRI* activity, the routing in dialogue for this activity includes the option to wait until the exit is clear. Thus, a patient who is in the waiting queue *(Q) Wait_Exam_MRI* is only allowed to enter the *(A) Exam_MRI* activity when the *(F) Dispatch_RMI* activity is empty. As a result, a patient in this activity never loses the resources linked with them in the previous activity, as there is a chance that they will repeat the examination, and consequently, the resources will remain the same.

- ***(F) Other_Exam***: Essentially, this fictitious activity aims to refer patients who have finished the preceding *(A) Exam* to another examination or out of the system, ending their stay in the unit. Consequently, each activity of the type *(A) Exam* is linked to an activity of the type *(A) Other_Exam*, having, therefore, a total of five activities of this sort, such as *(F) Mammo_Other_Exam*, *(F) Xray_Other_Exam*, *(F) MRI_Other_Exam*, *(F) CAT_Other_Exam* and *(F) US_Other_Exam*. Each of these fictitious activities has defined, in the routing out dialogue, the probabilities (in the form of conditional probability) of a patient leaving that examination and pursuing each path leading to another examination facility (or eventually leaving the system). These probabilities can be found in *section 7.3.*, *Table 8*.

  In the diagram, all possible paths out of each activity are highlighted with the same colour to help visualize the possible paths the patient might follow. It is essential to underline that there is no straight route back to an examination after a patient has left it. However, if the patient undergoes a second examination, there is always a way back to the initial one. In reality, a

patient does not return to an examination where they have already been. Thus, it is necessary to account for such situations and find a way to prevent the patient from returning to that examination.
Due to the patient's departure from the MRI facility, the resources associated with the two preceding activities are released as soon as the patient enters the *(F) MRI_Other_Exam* activity (*Table 15*).

***Table 15** – Resources and their specific rules for the (F) MRI_Other_Exam activity.*

Activity	Resources	Rules
(F) MRI_Other_Exam	1 R_Technician 1 R_Assistant	1. Only release the resource here.

- ***(Q) Mammo*, *(Q) XRay*, *(Q) MRI*, *(Q) CAT* and *(Q) US*:** These queues have no significance for the system since they exist to impose conditions for the passage of individual patients, operating as fictitious queues with no capacity limit or shelf life. These queues reflect the remaining routes to a patient's additional examinations after completing one. Therefore, except for the queue associated with that test, since the patient has departed, all of these queues are linked to the last activity, i.e., *(F) Other_Exam*, each with the probabilities specified in the routing out the dialogue as indicated above. It is important to note that up to this point, a patient is not prohibited from retaking an examination they have previously passed. For instance, a patient who has had an X-Ray and then a Mammography may return to the X-Ray since there is always a chance that they will follow that route after leaving Mammography. In order to prevent this situation, the system must verify whether this patient has previously been in the X-Ray facility before sending them. One way to do this is by checking, in this fictitious queue, whether the field of the label *L_Exam_ XRay* has a value other than zero. If so, it indicates that the patient has already been through the X-Ray facility and, as a result, cannot proceed to the activity that would lead them back to *(Q) Wait_XRay*. If the label field has zero value, it indicates that the patient is exempt from this examination and may go down this route. So far, patients who were previously excluded from moving out of this fictitious queue would accumulate and remain there indefinitely because their condition had not been validated. However, it is vital to maintain a steady flow of these patients.
For this reason, the label of type *L_Last_Exam* should be used so that it is possible to determine where the patient came from before entering that fictitious queue, i.e. which examination they passed most recently. Therefore, if the patient is held in the fictitious queue, all labels of this type are examined to determine which one has the value so that the patient is returned to the

corresponding preceding activity based on the label's value and obeying that activity's set priorities. As it is not always possible for a modeller to construct the system's profound logic using only the dialogues already offered by SIMUL8, it is required to follow the Visual Logic path to manage the system's behaviour precisely. Consequently, a code snippet was added in each fictitious queue that is examined immediately after a patient enters the queue, testing all the reasoning illustrated above. *Table 12A*, there is an illustration of the code used for each queue.

- ***(E) Exit_Mammo, (E) Exit_XRay, (E) Exit_MRI, (E) Exit_CAT*** **and** ***(E) Exit_US*****:** As noted previously, there is a chance that a patient who leaves an examination leaves the unit, finishing their stay. Therefore, for each activity of type *(F) Other_Exam*, it is linked to the corresponding exit point that catches the patients who leave the unit by that facility.

- ***(F) Go_Wait*****:** Each fictitious activity of this type, i.e., *(F) Go_Wait_Mammo*, *(F) Go_Wait_XRay*, *(F) Go_Wait_MRI*, *(F) Go_Wait_TAC*, and *(F) Go_Wait_US*, is succeeding the corresponding fictitious queue and its purpose is to receive those patients who validated the condition to proceed to the examination facility. Each activity forwards the patient to the corresponding type *(F) Wait_Exam* queue. This activity does not have any enforced conditions or defined routing out. However, it only exists because there must be an activity between two queues since it is the force responsible for extracting work items from queues.

8.2. Model Validation

To validate the model, it was necessary to run the simulation for a sufficient time and pause to confirm that the system's behaviour adheres to the previously specified requirements during various times of the day on different days. Thus, a simulation monitoring was conducted with a running period of a week, beginning at 07:30 AM on Monday and finishing at 08:30 PM on Friday, with the primary purpose of detecting any implementation faults for different operational conditions.

During the monitoring procedure, the pauses made at various times and days helped to validate some of the requirements needed by the decision-maker; therefore, the implemented model was subjected to four verifications. The first verification had to do with the average hourly arrival of patients at the unit. The second verification with the portion of patients who went to the check-in of each zone (IS or US), depending on the shifts in which these arrivals occur. The third with the proportion of patients who go to each of the queues for the respective exams after leaving the Imaging check-in service. The fourth and final verification ensures that patients do not return to the queue where they were before to complete a test if they have taken other routes than departing immediately after completing the examinations.

Verification 1: The rate of patient arrival to the service between 07:30 AM and 09:00 AM is 35.366 per hour, as stated in *Table 2* of *section 7.1.*. Around 1.5 x 35.366 should come this time, resulting in an average of 53.049 patients. The expected arrival value for the hour between 3:00 PM and 5:00 PM is 2 x 51.528, which equals an average of 103.056 patients. From 7:00 PM to 8:30 PM, there is an arrival rate of 13.453 patients per hour, averaging 1.5 x 13.453 = 20.180 patients. Using the average predicted values, the system was executed throughout the above-described execution time, and the values on the number of patients who arrived within each shift were retrieved. These values were introduced in *Table 16* for comparison with the predicted ones.

Table 16 – *Simulation results on the number of patients that entered each shift below for the whole running period.*

	Number of Patients					
Arrivals	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Expected**
07:30 AM – 09:00 AM	52	65	52	59	63	**53.049**
03:00 PM – 05:00 PM	102	100	89	91	86	**103.056**
07:00 PM – 08:30 AM	16	31	21	12	11	**20.180**

Daily values seem to swing somewhat around the predicted value. As they reflect an average number of patients each period, these deviations from the expected value are not indicative of a system that needs to be better adapted to reality. These deviations translate to a very crucial issue, namely uncertainty since it is not feasible to determine in advance the precise number of patients who arrive at the system during a specific time of the day; only its average is known, with the potential of days when these numbers are lower or more significant.

Verification 2: In the scenario described, a patient's likelihood of visiting one of the check-in stations depends on the arrival shift. For instance, in the 1:00 PM – 3:00 PM shift, the probability that a patient would visit the Imaging check-in service or the ultrasound check-in service is 83.6% and 16.4%, respectively. Given that the arrival rate for that shift is 41.445 patients/h, a total of 2 x 41.445 = 82.89 patients is expected to arrive at that time. The chances that resulted from the simulation were then examined to determine whether they agreed with the assumptions believed to have been made. To compute the corresponding probabilities, *Table 17* indicates the number of patients that arrived at the unit throughout the analysis period within that shift and how many of them went to each check-in.

Table 17 *– Simulation results on the number of patients who entered each check-in station, during the 01:00 PM – 03:00 PM shift, for the whole running period.*

Day	**Mon**		**Tue**		**Wed**		**Thu**		**Fri**		**Expected**	
TOTAL	73		113		73		81		80		**82.89**	
Check-In Station	**I**	**US**	**I**	**US**	**I**	**US**	**I**	**US**	**I**	**US**	**I**	**US**
Number of Patients	64	9	91	22	60	13	67	14	70	10	**69.3**	**13.6**
% Of Patients	87.7	12.3	80.5	19.5	82.2	17.8	82.7	17.3	87.5	12.5	**83.6**	**16.4**

It should be emphasised, however, that we are not dealing with fixed values. Thus the percentages may change but will remain near the predicted values. The number of patients who entered each check-in activity during that shift may be determined by conducting a run with a simulation time one minute before the shift began and deducting it from the patient number when the shift ended. The verification procedure for the other shifts operates similarly. However, as it is a straightforward process that cannot deceive the system, it sufficed to verify the prescribed percentage in which each shift activity points for each check-in station in the model.

Verification 3: Similar to when patients are distributed to one of the check-in stations after entering the system, the percentage of patients who go to each of the queues after leaving the check-in follows the same logic. For this reason, the premise for the third verification is similar to what was done with the second one, along with the explanation provided regarding the allowed values being approximately in the predicted range. In this particular scenario, the number of patients in queues was tallied at the end of each day. To eliminate possible patients who go to the queue for each examination by other paths, not directly from the check-in service, it was necessary to perform a simulation with these paths cut, thus leaving only the values of interest. The following table shows, by day, the total number of patients who attended each of the queues. It was feasible to determine the chance that a patient had gone to each of the four waiting queues, given the total number of patients who entered the system.

Table 18 *– Simulation results on the number of patients who entered each queue after leaving the Imaging check-in service, for the whole running period.*

	Mammo		**X-Ray**		**MRI**		**CAT**		**TOTAL**
Day	**# Of Patients**	**%**	**# Of Patients**	**%**	**# Of Patients**	**%**	**# Of Patients**	**%**	
Mon	41	11.7	145	41.3	66	18.8	99	28.2	351
Tue	66	16.2	171	41.9	74	18.1	97	23.8	408
Wed	51	13.8	144	38.8	91	24.5	85	22.9	371
Thu	58	15.1	147	38.5	89	23.2	89	23.2	383
Fri	67	17.8	149	39.5	70	18.6	91	24.1	377
Expected	-	**13.8**	-	**39.9**	-	**21.7**	-	**24.6**	-

It is essential to remember that this verification is not valid for the Ultrasound check-in since patients have no choice but to follow the route that leads them to this queue, which has a 100% probability.

At the beginning of *section 7.3.*, the percentage of patients who underwent each facilities within the IS was computed and presented in *Table 4* from *section 7.3.*. These percentages correspond to the chance that a patient would follow one of the pathways leading to the waiting room for that examination. Notably, the sum of all percentages exceeds 100%, which is expected given that not all patients were limited to a single examination type throughout their course. Considering the potential of patients having more than one examination of the same kind in the registry, if the number of examinations done of each type were divided by the total number of tests completed, the percentages of usage would be different and add up to 100%. It should be highlighted that the purpose is to compute the proportion of patients who had an examination at that facility. In addition, the above percentages were entered into the model; however, the programme attempted to standardise them to add up to 100% without requiring manual computation, as they are explicit in *Table 18*.

Verification 4:

With the apparent possibility of the patient moving to a queue where they have been previously, through other exams, it must be verified whether or not this situation occurs. As previously described in *section 8.2.5.*, code snippets were written in the software's Visual Logic so that the passage to one of the queues where the patient has previously been denied, making their return to the last activity to choose a different path. Thus, the system was run within the allocated time, with periodic execution pauses to consult the labels of patients either in the queue or engaged in an activity connected to the examination shortly after. In all the pauses to which the system was subjected, it was checked that the labels assigned to the activities where the patients were all contained the zero value, indicating that the patient had not yet completed these activities while inside the system having the system been validated.

9. Results Discussion and Presentation

In order to conclude, it will be essential to calculate the initial and final circumstances while simultaneously evaluating the system. After proposing a formal solution search technique and presenting this methodology, the results are further examined and interpreted.

9.1. Initial Conditions Definition

When examining the system's long-term behaviour, the results produced can be affected by the starting circumstances specified, hence modifying the system's behaviour. Define a strategy to prevent the scenario and assure the quality and dependability of the outcomes. In order to tackle this issue, a warm-up period will be considered, during which no data will be taken, guaranteeing that the results will only pertain to the model's stable states. To define this period, it was necessary to deduct the minimum number of required resources.

9.1.1. Minimum Number of Resources

This number reflects the minimum resources the system can minimally work and serves as the basis for future resource alterations. The minimum number of resources needed to conduct each of them in each activity will be measured. When determining this number, one of the variables must be considered is the amount of time devoted by each resource to a particular task. As soon as a patient begins an activity, resources start to be used and are released when the patient departs after the activity's length has passed. Thus, for each activity, each resource spends the same amount of time as the activity's duration.

Additionally, it is vital to consider the average number of patients entering the system at any given time. For this computation, many simulations were carried out, which lasted days or weeks, extracting the total number of arrivals and then dividing it by the total number of hours the system was operating based on the 13 hours per day in which the system works.

For example, *Table 13A* was generated containing the total number of arrivals in a working day, which were distributed among the several time slots. This value was then divided by the 13 hours that the system was admitting new patients, yielding an average rate of 38.385 patients/h. It is predicted, however, that the number of simulations performed is independent of this average arrival rate since the ratio between the total number of arrivals and the number of hours the system works will remain constant.

Although the acquired average arrival rate value is universal, the number of minimal resources for each activity cannot be computed using this value since there are patients who are split among the current activities and should not be considered for this reason. Therefore, the proportion of total patients who travel through each activity must be determined through the simulation results and multiplied by the average arrival rate. This final value is multiplied by the time each resource spent on the relevant activity and then divided by 60 to get the minimal number of resources required per activity each hour.

The parameters to compute the minimal number of resources required for each examination activity are summarized in *Table 19*.

***Table 19** – Parameters to determine the minimum number of resources for each (A) Exam activity.*

Activity	Resources	Average Arrival Rate (patients/h) – Simulated	% Of Patients (%) – Simulated	Duration (min)
(A) Exam_Mammo	**1 Technician**	38.385	19	15
(A) Exam_XRay	**1 Technician 1 Assistant**		42	10
(A) Exam_MRI	**1 Technician 1 Assistant**		22	40
(A) Exam_CAT	**1 Technician 1 Assistant**		22	20
(A) Exam_Ultrasound	**1 Physician 1 Assistant**		45	15

As an activity needs just a single resource of a given kind at any moment, it may be assumed that the minimal number of resources required for this activity is equal to the number of activities of the same type that can occur concurrently in the system. This is because if an activity requires a certain resource twice, adding one additional resource would decrease its execution time in half. In reality, this would entail a second activity occurring concurrently so that two patients might be despatched at the same time as previously one could. As a result, as there would be two facilities working simultaneously, the occupancy rate would be cut in half, i.e., personnel would exert half the effort to provide for the same number of patients.

For Mammography, the calculated minimum number of resources (technicians) is 1.823. As a result, a technician would no longer be an option since the system would be in deficit. The occupancy rate would reach its maximum level, and patients would wait a long time for this activity. This value should thus be rounded up to the closest integer. Therefore, the minimum for this activity to work decently would be to have two technicians (which is equivalent to two activities of this type taking place simultaneously, as explained above).

The same reasoning used in this activity is also applied to the others. For the X-Ray, the minimum number of resources is three, i.e., three technicians and three assistants, meaning three activities occurring in parallel. For the MRI are six of each, i.e., six technicians and six assistants, i.e., six activities in parallel. For the CAT, three of each, i.e., three technicians and three assistants, i.e., three activities in parallel. Finally, for Ultrasound, are five of each, i.e., five physicians and five assistants, i.e., five activities in parallel.

The minimum number of resources of a certain kind in the system is equal to the sum of the minimum resources present in each activity. In total, there will need to be at least 14 technicians, 17 assistants and six physicians per hour, which translates into having simultaneously two activities for Mammography, three for X-Ray, six for MRI, three for CAT and five for Ultrasound. *Table 14A* helps to consolidate these numbers by providing a table summary.

The same procedure needed to be done for the check-in activities since these activities also required resources to be carried out. Unlike the other resources previously listed for the examinations, these resources are exclusive for the check-in activities, with no resource sharing across activities of different types. Since there is only one type of resource, and only one person performing this activity per patient, it makes the most sense to examine the minimum number of simultaneous replicas of this activity in parallel since it has already been shown that it correlates with the number accessible of resources. Therefore, the parameters to compute the minimum number of resources required for each check-in activity are summarized in *Table 20*.

***Table 20** – Parameters to determine the minimum number of resources for each (A) Check-In activity.*

Activity	Average Arrival Rate (patients/h) – Simulated	% Of Patients (%) – Simulated	Duration (min)
(A) Check_In_I	38.385	80	7.527
(A) Check_In_US		20	

For the Imaging check-in activity, the minimum number of activities required for the system to work appropriately is four.

On the other hand, for the check-in activity in the ultrasound facility, the minimum number of activities required for the system to work correctly is one.

In conclusion, stating that the minimum number of activities of this type to co-occur is four for the IS and one for the US is equivalent to saying that there will be four employees at the check-in counter in the first zone and one employee at the counter in the second zone. This means the system will require five employees to meet the minimum conditions for working on these two activities. These tabulated data may be found in *Table 15A.*

9.1.2. Warm-Up Period

After determining the minimal number of resources required for each activity (*Tables 14A* and *15A*), the next step was to set these number of resources and replicas of activities within the system and determine when the occupancy rate of each resource along with the average queueing time in the corresponding activity began to stabilize. Therefore, multiple runs were conducted weekly to determine that period of stabilization.

According to *Tables 16A* to *26A* (and the complementary information from *Graphic 3A* to *13A*), the results of the simulation after 13 weeks of execution seem to be entirely stable after a period of small oscillations relating to the average waiting times. On the other hand, the occupancy rates of the resources have stabilized since the first week.

Due to the need to account for an extra safety margin to guarantee those beginning circumstances do not impact the results as much as possible, it was determined that the warm-up period would be 26 weeks (twice the 13 weeks verified), which corresponds to 105,300 minutes, the time unit of the system. After obtaining this value, entering it in the clock properties was sufficient.

9.2. Termination Conditions Definition

Once the warm-up period has been set, the timeframe for collecting results must be determined. Variability in these results depends on the number and duration of runs done. Thus, it is supposed to examine, across different periods, the variation related to all KPIs specified during the problem formulation phase. For this analysis, it was determined once again that the resources are at their minimum.

The number of trials for each experiment was first determined to ensure the precision of the confidence intervals surrounding the estimated mean of the simulation results. The number of trials tells the number of times the system is executed within the specified time period. Consequently, an accuracy of up to 5% of the mean value was established. This indicates that there is a 95% probability that, in the future, a genuine result for a specific KPI will fall within the range limited by its minimum and maximum values. It turned out that three trials were sufficient, but five were added to ensure a safety margin. This means that each data collecting period will consist of five runs or samples from the same experiment, each exhibiting a degree of variability due to the random numbers regulating its behaviour. In turn, the values to be accepted within the stated confidence interval established for each KPI will be presented in the results manager of SIMUL8, along with the mean.

With this in mind, the results collecting periods were arbitrarily determined. This duration was altered successively after each trial's set of five runs. The purpose is to compare the indicated confidence intervals for the same KPI across different results collection periods. Consequently, the results collecting periods 5, 10 and 20 weeks were selected. This implies that after the initial conditions no longer affect the system (the warm-up period concludes), the system will run five times and gather the results for each period.

After running the system for each of these periods, a table was constructed containing, for each KPI, the mean value and confidence interval, with an accuracy of 5%, along with the variance between the extreme values. *Tables 27A*, *28A* and *29A* reflect the results following each collecting period of 5, 10 and 20 weeks, respectively. *Table 30A* was constructed to reunite the variances for each interval's extremes for the same KPI throughout the various collecting periods to further analyse the variances in each confidence interval. A quick study reveals that the variances of each confidence interval tend to decrease with the length of the results collection period, indicating that the system gets more stable with time. Consequently, analysing the system over more extended collection periods makes sense. However, one must also consider the intensive computing work of the machine's CPU while operating the system for extended periods. Therefore, the results collection period was fixed at 20 weeks.

9.3. Solution Searching Technique

Considering the minimum resources and replicas of activities identified above, many software-based experiments were conducted, based on the data collection period previously set, to examine all possibilities regarding the number of resources to be compared and assessed. In the initial phase of the analysis, the number of replicas and resources will be maintained at their minimum values, according to the *Tables 14A* and *15A*, except for the activity and its associated resources that are being studied, whose quantity will vary between executions.

As the number of minimum resources is defined based on the minimum number of activities working in parallel so that the system can minimally work, each type of activity has defined a collection of values, each one corresponding to the exact number of replicas of that activity that will be served as the basis for each experiment run. Along with the number of replicas, the number of resources is defined accordingly. So, for each execution, a pair is determined by the number of existing replicas of that specific activity and the number of resources. Henceforth, these values will be referred to as the resource's "quantity."

Because the number of resources is defined through the number of replicas, it can be proved, taking as an example the check-in-related activities, that when the number of resources increases upon the unchanged number of replicas, the resource's utilisation rate will suffer a decrease (there are more resources for the same number of replicas, which makes the load decrease between them – *Table 31A*. In contrast, the resources' utilisation rate will increase if there are more replicas compared to the number of resources (the load increase between them). Therefore, it becomes necessary to proportionally alter the number of resources alongside the number of replicas to guarantee sufficient resources so that all activities may be executed (although the aspect of decreasing resources and maintaining the same number of replicas will be evaluated later). This is essential to notice improvements (or losses) for the same increase from run to run, precisely to determine how the resource occupancy rate fluctuates as the number of resources and replicas rise proportionally.

The following sections will demonstrate how increasing replicas and resources in each activity affects its three KPIs. The number of replicas and resources will be increased by one unit throughout the subsequent three runs, beginning at the start point, with the number of resources being reduced to their absolute minimum. After each iteration, the values of all three KPIs will be retrieved. An analysis will determine at what levels the system will most likely provide the maximum benefit.

Because the model is separated into two parts, the procedure described above must be performed independently since one part may congest the other, or the first's congestion may impact the second. In a system with a single-entry point and two activities linked by a queue, situations such as this may occur owing to the activities' varying durations. For instance, if the duration of the first activity is greater than the duration of the second activity, it indicates that the departure rate of the first activity is lower than the entrance rate of the second activity, leaving the first activity to condition the rest of the system (bottleneck). Due to this, the waiting queue between these two activities will be less congested than the queue before the first activity. However, duration is not necessarily the most important criterion for an activity to influence the system; the number of replicas (and, ultimately, resources) is also an important parameter controlling the system's flow. To examine the impact of adding replicas and executing resources, it is crucial that the immediately preceding activities have a natural flow to accurately interpret the KPI values. In other words, if check-in activities are thought to be "bottlenecks" in the system, the waiting queues for examinations would not behave similarly to what occurs, and it would be bad practise to believe the KPIs results to be the real ones and then make system adjustments based on them. To circumvent this issue, it was proposed that modifications to activities delivered to examinations be evaluated and assessed while keeping an optimal number of replicas and resources for check-in activities so that they are not affected by their conditioning.

9.3.1. Changes in the Imaging Check-In Service

Initially, four activities of this type were designated as minimum resources and four secretary resources, one for each replica. According to *Table 32A*, under the starting conditions of limited resources and a 20-week execution period, the average and maximum queueing times for this activity were observed to be 63.05 minutes and 219.72 minutes, respectively.

With the addition of one more replica, and, consequently, one more secretary, i.e., increasing from four to five replicas, in contrast to the previous results, the average queueing time dropped to 10.60 minutes and the maximum queueing time to 82.03 minutes, demonstrating 83% and 63% reductions, respectively.

With the increase from five to six replicas, the average waiting time decreased by 83% to 1.77 minutes, while the maximum waiting time decreased by 64% to 29.60 minutes.

At a seventh replica addition, the average queueing time was reduced by 69% to 0.55 minutes, while the maximum queueing time was reduced by 41% to 17.33 minutes.

For this activity, increasing the number of replicas from four to five had the same impact as increasing it from five to six, i.e., the benefit rose correspondingly. The effect is diminished when the number of replicas increases from six to seven. For this reason, it is preferable to abuse the system with six replicas.

The initial occupancy rate of the resources present in this activity was 96.10% and, depending on the increase in the number of replicas and resources, decreased to 82.53%, 70.74% and 61.90%, respectively.

9.3.2. Changes in the Ultrasound Check-In Service

At first, one replica and one secretary resource were deemed the bare minimum resources. According to *Table 33A*, under the initial constraints of minimum resources and a 20-week execution period, the average and maximum queueing times for this activity were determined to be 101.65 minutes and 343.35 minutes, respectively.

In contrast to the previous results, with the addition of one more replica and one additional secretary, i.e., increasing the number of replicas from one to two, the average queueing time decreased to 3.90 minutes and the maximum queueing time to 57.58 minutes, showing 96% and 83% reductions, respectively.

The average waiting time fell by 85% to 0.58 minutes and the maximum waiting time by 59% to 23.64 minutes with the expansion from two to three replicas.

The average queueing time decreased by 83% to 0.10 minutes, while the maximum queueing time decreased by 50% to 11.85 minutes upon installing the fourth replica.

For this exercise, raising the number of replicas and associated resources from one to two proved to have the highest impact on the KPIs.

The initial resource occupancy rate for this activity was 96.10%, and when the number of resources and replicas increased, it declined to 81.87%, 70.74%, and 61.40%, respectively.

9.3.3. Changes in the Mammography Facility

Initially, two mammography facilities were designated to operate concurrently, which makes a total of two technicians working simultaneously, one in each room. In addition, it was previously determined that *(A) Check_In_I* and *(A) Check_In_US* would be considered optimised with six and two replicas, respectively, with eight resources of that type.

According to the *Table 34A*, under starting conditions of minimal resources and a period of 20 weeks, after execution, the KPIs' average queueing time and maximum queueing time for this activity were determined to be 61.49 minutes and 244.22 minutes, respectively.

With the increase of one more facility, i.e., from two to three replicas of this activity, there is a massive improvement in the KPIs results, with the average queuing time decreasing to 4.51 minutes and the maximum queuing time decreasing to 52.60 minutes, which represents a 93% and 78%, decrease in comparison to the previous values, respectively.

It was observed that for a new increase from three to four facilities, the average queueing time was 0.83 minutes, representing a decrease of 82%. In contrast, the maximum queueing time was 24.90 minutes, representing a decrease of 53%.

The average queueing time at a fifth facility was reduced by 78% to 0.18 minutes, while the maximum queueing time was reduced by 47% to 13.30 minutes.

With this, it was verified that the percentage of improvement had the highest impact from the transition from two to three replicas. As this number increased, the benefit percentages decreased.

Regarding occupancy rate, the baseline value was 91.92%, lowered to 87.42%, 82.03%, and 77.17% throughout the three executions as additional facilities and resources were added.

9.3.4. Changes in the X-Ray Facility

Initially, three X-Ray facilities were authorised to function concurrently, requiring three technicians and three assistants in each room. Furthermore, with eight resources of such sort, it was previously found that *(A) Check_In_I* and *(A) Check_In_US* would be optimised with six and two replicates, respectively.

According to the *Table 35A*, after the first execution under initial conditions of minimum resources and a period of 20 weeks, the KPIs' average queueing time and maximum queueing time for this activity were found to be 33.72 minutes and 174.97 minutes, respectively.

With the addition of one additional facility, i.e., from three to four replicas of this activity, there is a significant improvement in the KPIs, with the average queueing time reducing by 91% to 3.05 minutes and the maximum queueing time decreasing by 75% to 43.31 minutes, compared to the initial values.

It was discovered that with a new increase from four to five facilities, the average queueing time decreased by 78% to 0.66 minutes, while the maximum queueing time was reduced by 57% to 18.80 minutes.

At a sixth facility, the average queueing time was decreased by 67% to 0.22 minutes, while the maximum queueing time was cut by 17% to 15.60 minutes.

There was a more significant percentage difference when increasing from three to four replicas, which appears to translate into a better gain for the same increase of one resource of each kind and one replica in each of the executions.

The baseline occupancy rate of the resources was 91.92% for technicians and 91.50% for assistants. When more facilities and resources were added throughout the three executions, the values decreased to 86.88%, 81.62%, and 77.08% for technicians and 87.40%, 82.92%, and 79.01% for assistants.

9.3.5. Changes in the MRI Facility

At first, six MRI facilities were proposed to operate simultaneously, necessitating the presence of six technicians and six assistants, one of each in each room. In addition, it was previously determined that *(A) Check_In_I* and (*A) Check_In_US* would be considered optimised with six and two replicas, respectively, with eight resources of that type.

According to the *Table 36A*, the average queueing time and the maximum queueing time for this activity, as measured by the KPIs, were determined to be 34.95 minutes and 209.58 minutes, respectively, after the first execution under the starting conditions of minimum resources and a period of 20 weeks.

There is a significant improvement in the KPIs with the addition of one additional facility, i.e., from six to seven replicas of this activity, with the average waiting time decreasing by 63% to 12.84 minutes and the maximum waiting time decreasing by 53% to 97.49 minutes, compared to the initial values.

The average queueing time was reduced by 59% to 5.25 minutes, with a new increase from seven to eight facilities, while the maximum queueing time decreased by 32% to 66.04 minutes.

The average waiting time at a ninth facility was reduced by 54% to 2.42 minutes, while the maximum waiting time decreased by 48% to 34.50 minutes.

It should be noted that the system is much improved by the rise from seven to eight facilities compared to the other increases. In terms of the maximum queueing time, it was preferable to expansion the replicas from eight to nine, rather than from seven to eight, making this increase the second-highest benefit.

The occupancy rates of technical and assistant resources began at 91.92% and 91.50%, respectively, then fell as additional resources of the same kind were added in each execution, being the occupancy rate of 89.96%, 87.73%, and 85.68% for technicians and 89.87%, 88.08%, and 86.26% for assistants.

9.3.6. Changes in the CAT Facility

Initially, three CAT facilities were assigned to run concurrently, resulting in three technicians and three assistants operating simultaneously in each room. In addition, with eight such resources, it was previously determined that *(A) Check_In_I* and *(A) Check_In_US* would be considered optimal with six and two replicas, respectively.

According to the *Table 37A*, the results of the KPIs' average queueing time and maximum queueing time for this activity were determined to be 78.71 minutes and 291.57 minutes, respectively, under the beginning conditions of minimum resources and a period of 20 weeks of execution.

With the addition of one additional facility, i.e., increasing the number of replicas from three to four, the average queueing time decreased to 15.32 minutes and the maximum queuing time decreased to 127.88 minutes, which represented an 81% and 56% drop in comparison to the initial values, respectively.

For the increase from four to five facilities, the average queueing time was 2.90 minutes, representing a drop of 81%; meanwhile, the maximum queueing time was 50.95 minutes reflecting a decrease of 60%.

At a sixth facility, the average waiting time was decreased by 69% to 0.91 minutes, while the maximum waiting time was cut by 48% to 26.25 minutes.

This analysis reveals that the KPIs perform more optimally for the transition from four to five replicas.

The baseline occupancy rate for the resources was 91.92% for technicians and 91.50% for the assistants. Nevertheless, when more replicas and resources were added throughout the three executions, the values decreased to 86.13%, 83.70% and 79,19% for technicians and 89.24%, 84.74% and 80.83% for assistants.

9.3.7. Changes in the Ultrasound Facility

At first, five ultrasound facilities were assigned to operate simultaneously, resulting in five assistants and five physicians working simultaneously, one in each room. Additionally, it was previously found that *(A) Check_In_I* and *(A) Check_In_US* would be optimal with six and two replicas, respectively, with eight such resources.

According to *Table 38A*, under the initial circumstances of minimal resources and a 20-week execution period, the results of the KPIs for average queueing time and maximum queueing time for this activity were determined to be 31.24 minutes and 142.62 minutes, respectively.

With the addition of one more facility, i.e., expanding the number of replicas from five to six, the average queueing time reduced to 5.01 minutes, and the maximum queueing time decreased to 52.74 minutes, representing an 84% and 63% reduction, respectively, in contrast to the initial results.

With the expansion from six to seven facilities, the average queueing time was 1.29 minutes, representing a 74% reduction, while the maximum queueing time was 27.41 minutes, representing a 48% decrease.

At an eighth facility, the average waiting time was decreased by 71% to 0.37 minutes, while the maximum waiting time was cut by 55% to 12.25 minutes.

According to what has been seen, adding a sixth replica, with the respective physician and assistant, has a more substantial influence on the system than increasing from six to seven or even from seven to eight replicas. However, the second greatest benefit regarding the maximum queueing time is raising the number of facilities from seven to eight (55% against 48%), which makes this transition considered the second-highest improvement.

As more replicas and resources were added in each execution, the occupancy rates for assistant and physician resources, which had initially started at 91.50% and 90.79%, respectively, dropped to 86.78%, 82.26% and 78.12% for assistants and 76.06%, 65.16% and 57.06%.

9.4. Results Interpretation and Analysis

As was noted, the most significant improvement in the KPIs occurred almost immediately after the first run, where the values corresponding to the minimum number of replicas and resources were raised by one. After some time, this advantage began to diminish due to the increased number of replicas and resources included in each posterior execution. Nevertheless, these improvements were still significant, so it was impossible to entirely rule out the potential of climbing to a higher level. The person in charge of making a choice will determine how much of an increase in costs they are ready to accept to make this change in the system.

In general, the percentage of the extra benefit tends to stabilise as the number of replicas and resources increases. However, after a few runs, it is possible to see that for certain KPIs, the value between the last two runs changes by just a tiny amount. However, as the proportion of benefit decreases until it stabilises, there will come a time when increasing the number of replicas and resources is no longer advantageous, either in this context or in any other where a comparable simulation experience is conducted. From this point on, raising the number of replicas or resources is not advised since the utilisation rate would fall considerably. In other words, the proportion of replicas and resources to work items would be 1:1, queueing times would go toward zero, and the system would be undercrowded, which would be a fantastic benefit for patients.

Considering the number of replicas and resources where the benefit is more significant in each activity conducted in *section 9.3.*, these values (*Table 39A*) were included in the model to examine the effect of all modifications in concert. The table below compares the results of the KPIs when each activity was analysed individually and in combination with the other activities, always employing the number of replicas and resources for which there was a more significant benefit.

Looking at the *1st more significant benefit* column of the *Appendix 1*, the performance of the KPIs has deteriorated, as predicted, when the number of replicas and resources are optimised for all activities concurrently. When reviewing the KPI results for a particular activity while limiting the others to the bare minimum of resources, this conditioning in the waiting queues of these activities obstructed the regular flow. As a result of fewer patients circulating from these activities, fewer patients were added to the waiting queue for the "optimal" activity (via other examinations). On the other hand, when all activities are optimised simultaneously, the number of patients circulating between examinations increases. More patients are sent for examinations, causing more patients to move to other waiting rooms. In *Table 40A*, the number of patients waiting in queue for examinations before the individual and collective analysis is provided. Therefore, it is verified that the number of patients in all examination queues is greater when the system is implemented, considering an optimal solution for all activities simultaneously.

Consequently, it may be required to investigate, for each activity, the number of replicas and resources for which the second-greatest benefit will be realised. According to *Tables 32A* to *38A*, the second highest benefit occurs most frequently immediately after increasing the number of existing replicas from the first benefit by one. For MRI and Ultrasound examinations, however, *Tables 37A* and *38A* reveal that the total benefit (relative to the two KPIs in the queues) is more significant when the number of replicas is increased by two rather than one. Recall that this extra added benefit per unit stabilises at a specific value. Increasing the number of replicas is advantageous if the added benefit is equal to or greater than the immediately preceding value.

Then, the number of replicas will grow for each activity to reach the second-highest benefit, and the resource allocation will be readjusted. Consequently, the system was rerun for the same period (20 weeks), with the parameters presented in *Table 41A*. The results of the KPIs are shown in the last column of *Appendix 1*. As expected, the results revealed decreased queueing times for all queues; thus, the KPIs performed better with increased replicas and their resources in a single unit for the second-highest benefit. This provides a far more convenient solution to the system resulting from shorter times in queues. Notably, the difference between the individual and collective analyses for the second-highest benefit is lower than those for the first more substantial benefit, anticipating that the difference would decrease as replicas and resources rise.

The exact needs to be updated for check-in activities. The *Appendix 1* shows that during the two most considerable benefits analyses, there was a null difference between the individual and collective analyses. It is anticipated that similar behaviour will continue for the following benefits. One of the reasons this occurs is that the flow of both check-in processes is independent of one another, i.e., the performance of one is unaffected by the entrance rate of the other, apart from the fact that the same does not happen for activities related to examinations, as already explained. Since the entry rate of these activities depends on the exit rate of the check-in activities, the effect of a bottleneck on these activities acts as a limitation on the entire system.

When analysing resource efficiency, one may directly notice a decline when the queues' KPIs for a single activity improve. Lower occupancy rates are a direct result of lower efficiency and vice versa. This indicates that more resources must be wasted in a specific period for operations to be carried out efficiently. The fact that resources are less efficient throughout the tests, as a consequence of their increase based upon the number of replicas, is because there are more of them for the same number of patients on the waiting queue. The workload of each resource is decreased because, as predicted, the same number of patients is more evenly distributed among them. With the assumption that the activity time is constant, there is, thus, more free time for the resource. More free time denotes increased resource waste or resources awaiting allocation to an activity.

Even though the number of resources is proportionate to the number of activities, the workload is reduced even further when all activities are examined simultaneously because more resources are available. The fact that the activities frequently share a resource is the only reason this occurs. There is a propensity for stabilisation between individual and collective analyses when a physician is involved in the Ultrasound activity because it is the only exclusive resource of this activity. Since more patients have this examination simultaneously due to the increased flow, there is still a chance to detect a slight, nearly imperceptible rise, which was expected given the increased demand for the resource.

It is anticipated that the difference between the two collective analyses of the first and second-highest benefits will reflect the same logic given what has already been stated about efficiency significantly declining as the number of resources increases. *Appendix 1* shows a noticeable decrease in the utilisation rate between the two strategies, proving what was predicted.

To promote resource efficiency, the system was put into place to study the impact of reducing resources by resource type each while keeping the number of activities. The resource utilisation rate is expected to grow based on what has already been stated and demonstrated in *Table 31A*, for the opposite case. Because it is not possible to simultaneously improve resource efficiency and service delivery, there is a trade-off between the two. In order to do this, it is important to strike a balance between the two and determine a level of comfort where there are neither benefits nor losses concerning one another. Thus, the value of the performance of the KPIs of the second-highest benefit was used as a starting point, and three experiments were conducted, each with one less resource of each type compared to the previous experience, to determine how the efficiency of resources improved concerning the decrease in performance of the remaining KPIs. The results are presented in *Appendix 2*.

The analysis of the results reveals that there are still some improvements in the average queueing times for the Mammography and MRI queues. However, the values of the KPIs associated with other examination queues (except for the Ultrasound) indicate weak variances compared to the base experience. One of the reasons this occurs is because the check-in activities, which condition the system, have fewer resources. As a result, the system is more crowded in these queues, which lowers the flow in the second part of the system. It can be seen from that table for these queues that performance, namely the maximum queueing time, actually improves dramatically. Although the number of patients who wait longer than the average value is relatively small, it is still a case that merits consideration because the average waiting time is still low (*Graphics 14A* and *15A*). Although the performance of the KPIs associated with the Ultrasound queue deteriorated significantly due to the reduction in the number of physicians, it became evident that this resource is very dependent on the system's functioning as a whole. To circumvent this issue, a second execution was carried out, taking as the basis the experiment's values with less than three resources (specified in *Appendix 3*), increasing only the number of physicians by two.

According to *Appendix 3*, a reduction from 24 to 21 technicians increased their efficiency from 68.22% to 75.90% following a final general analysis based on the values of the second-highest benefit. With the reduction from 28 to 25 assistants, this resource's efficiency increased from 67.97% to 74.33%. The change in efficiency caused by the drop from 8 to 7 physicians was from 58.89% to 67.30%. A reduction from 10 to 7 secretaries increased productivity from 50.59% to 72.22%. In addition, it was demonstrated that the increase in resource efficiency had a minor effect on the performance of the queues, although some have varied more than others.

10. Conclusion and Future Work

This study addressed the need to enhance the quality of service provided to patients by the Imaging Department of Hospital da Luz. The objective was to reduce the amount of time customers must wait for services in order to satisfy patients' needs for a higher quality of service and increase demand. In order to provide a set of solutions, the metrics for minimising these wait times were explored in a variety of methods. These solutions will strengthen the decision-making ability of the individual responsible for making the call following the presentation of these results, as only the decision-maker will be able to determine which option is ideal for keeping the system operational.

Sometimes the decision-making process can be more difficult than expected. In situations where a clear trade-off is apparent, where a gain in performance in one factor leads to a decrease in another, it is crucial for the decision-maker to reflect and consider, for instance, which of the two factors has the greater impact for them. During the analysis, it was always observed the effect of the trade-off between keeping patients pleased and making better use of resources, given that increasing the quantity of one requires decreasing the amount of the other. If more resources are available, there will be a better flow of patients through the waiting queues, lowering the queueing times and, therefore, their length of stay in the system, so boosting patient satisfaction in this regard. Satisfied patients are more likely to return, leading to an increase in demand and higher revenue. On the other hand, the quantity of money supplied to employees in the form of wages, for example, will grow. This will lead to an increase in variable costs for each extra resource or activity. In addition, the fixed costs may be adjusted, but only to a small degree than variable costs. The fixed costs include equipment maintenance and electricity bills (more equipment running). Before providing a variety of alternatives that are most suitable for the person making the decision, in this case, as in others, it is vital to control the budget to make the most of the available resources while meeting the needs of the patients without significantly surpassing the opportunity cost that the decision-maker is facing.

During the analysis, it was possible to observe that the bottleneck effect on check-in activities is not necessarily detrimental to the system, as it frequently prevents examination waiting rooms from being overcrowded, which may be preferable given that patients may remain in more than one queue for an examination during their stay. As witnessed, the results of the KPIs were only able to fluctuate slightly between resource removal trials due to the conditioning of the check-in activities with the flow. A limitation on check-in activities might also be a successful strategy for reducing the number of health personnel (technicians, assistants and physicians). After taking this into consideration, as observed in the last experiment, reducing the number of resources did not result in a substantial difference in the performance of the examination queues other than enhancing the efficiency between the resources, so it may be advantageous to go this way.

Furthermore, when the aim of the modeller is to improve the performance of the queues, as was done in the initial analyses of the preceding chapter, it is crucial to underline the importance of reducing the bottleneck effect. It is important to reiterate the point that system enhancements will need to be made in phases. To put it another way, increasing the performance of examination-related tasks would be pointless if check-in procedures delayed patients. Because of this, it is crucial to evaluate and test the other activities from a position where patients are not backed up in the check-in lines, i.e., where the flow of these activities is already judged acceptable, and the exit rate is, in the best scenarios, close to the entrance rate. Therefore, it is argued that congestion will always occur in a system where an activity's exit rate is larger than the entry rate of the activity that precedes it.

Appendix 2 reveals that the resource utilisation is approximately 70%, for the maximum reduction the model was subjected, indicating that 30% of the time is spent waiting to be allocated to activities owing to a lack of patients or an excess of resources. Studying the periods when these resources are most in demand and the contrary, i.e., when they are less needed, is one strategy to improve even more the efficiency of these resources. Since this research deals with a limited number of patients whose admission follows particular temporal distributions, the number of patients fluctuates with the time of day, and the flow of patients through the system will not be constant throughout the day. However, the number of patients is similar from time to time. In other words, there is a propensity for patterns to repeat themselves based on the time of day. When keeping the same activities and resources throughout a changeable patient flow, it is only reasonable that the utilisation rate would oscillate. There may be certain times when the utilisation rate reaches close to 100% as well as other periods when it reaches a value below 50%, indicating that an adjustment could be made in the number of resources according to the time of day when this occurs so that there is a more efficient use of resources. In light of this, one could divide the system into many shifts and propose the necessary number of resources for each shift to battle the patient flow. Thus, it was guaranteed that less resources would be wasted, thereby boosting their efficiency without impacting the performance of the queues.

References

Abar, S., Theodoropoulos, G. K., Lemarinier, P., & O'Hare, G. M. P. (2017). Agent Based Modelling and Simulation tools: A review of the state-of-art software. *Computer Science Review*, *24*, 13–33. https://doi.org/10.1016/J.COSREV.2017.03.001

Abo-Hamad, W., & Arisha, A. (2013). Simulation-based framework to improve patient experience in an emergency department. *European Journal of Operational Research*, *224*(1), 154–166. https://doi.org/10.1016/J.EJOR.2012.07.028

Ahmad, J., Iqbal, J., Ahmad, I., Khan, Z. A., Tiwana, M. I., & Khan, K. (2020). A Simulation Based Study for Managing Hospital Resources by Reducing Patient Waiting Time. *IEEE Access*, *8*, 193523–193531. https://doi.org/10.1109/ACCESS.2020.3032760

Ahmed, M. A., & Alkhamis, T. M. (2009). Simulation optimization for an emergency department healthcare unit in Kuwait. *European Journal of Operational Research*, *198*(3), 936–942. https://doi.org/10.1016/J.EJOR.2008.10.025

Alejandro Huerta-Torruco, V., Hernández-Uribe, Ó., Adriana Cárdenas-Robledo, L., & Amir Rodríguez-Olivares, N. (2022). Effectiveness of virtual reality in discrete event simulation models for manufacturing systems. *Computers and Industrial Engineering*, *168*. https://doi.org/10.1016/J.CIE.2022.108079

Andersen, A. R., & Plesner, A. L. (2022). Optimization of the organizational structure in hospitals to account for patients with multiple diseases. *Artificial Intelligence in Medicine*, *130*, 102327. https://doi.org/10.1016/J.ARTMED.2022.102327

Andersen, D. F., Rich, E., & MacDonald, R. (2009). Public Policy, System Dynamics Applications to. *Complex Systems in Finance and Econometrics*, 655–670. https://doi.org/10.1007/978-1-4419-7701-4_36

Ashour, O. M., & Okudan Kremer, G. E. (2013). A simulation analysis of the impact of FAHP–MAUT triage algorithm on the Emergency Department performance measures. *Expert Systems with Applications*, *40*(1), 177–187. https://doi.org/10.1016/J.ESWA.2012.07.024

Asif, A., & Zeeshan, M. (2020). Indoor temperature, relative humidity and CO2 monitoring and air exchange rates simulation utilizing system dynamics tools for naturally ventilated classrooms. *Building and Environment*, *180*, 106980. https://doi.org/10.1016/J.BUILDENV.2020.106980

Azadeh, A., Pourebrahim Ahvazi, M., Motevali Haghighii, S., & Keramati, A. (2016). Simulation optimization of an emergency department by modeling human errors. *Simulation Modelling Practice and Theory*, *67*, 117–136. https://doi.org/10.1016/J.SIMPAT.2016.04.005

Baesler, F. F., Jahnsen, H. E., & DaCosta, M. (2003). Emergency departments I | Proceedings of the 35th conference on Winter simulation: driving innovation. *Emergency Departments I: The Use of Simulation and Design of Experiments for Estimating Maximum Capacity in an Emergency Room*, 0–190. https://dl.acm.org/doi/10.5555/1030818.1031081

Baines, P. G. (2008). Lorenz, E.N. 1963: Deterministic nonperiodic flow. Journal of the Atmospheric Sciences 20, 130–41.1: *Http://Dx.Doi.Org/10.1177/0309133308091948*, *32*(4), 475–480. https://doi.org/10.1177/0309133308091948

Bair, A. E., Song, W. T., Chen, Y. C., & Morris, B. A. (2010). The impact of inpatient boarding on ED efficiency: A discrete-event simulation study. *Journal of Medical Systems*, *34*(5), 919–929. https://doi.org/10.1007/S10916-009-9307-4/TABLES/7

Barjis, J. (2011). *Healthcare Simulation and its Potential Areas and Future Trends*.

Barlas, Y. (1989). Multiple tests for validation of system dynamics type of simulation models. *European Journal of Operational Research*, *42*(1), 59–87. https://doi.org/10.1016/0377-2217(89)90059-3

Boyle, L. M., Marshall, A. H., & Mackay, M. (2022). A framework for developing generalisable discrete event simulation models of hospital emergency departments. *European Journal of Operational Research*, *302*, 337–347. https://doi.org/10.1016/j.ejor.2021.12.033

Brailsford, S. C., Eldabi, T., Kunc, M., Mustafee, N., & Osorio, A. F. (2019). Hybrid simulation modelling in operational research: A state-of-the-art review. *European Journal of Operational Research*, *278*(3), 721–737. https://doi.org/10.1016/J.EJOR.2018.10.025

Brailsford, S., & Hilton, N. (2000). *A Comparison of Discrete Event Simulation and System Dynamics for Modelling Healthcare Systems*.

Brennan, A., Chick, S. E., & Davies, R. (2006). A taxonomy of model structures for economic evaluation of health technologies. *Health Economics*, *15*(12), 1295–1310. https://doi.org/10.1002/HEC.1148

Clifford Bleustein, David B. Rothschild, Andrew Valen, Eduardas Valaitis, Laura Schweitzer, & Raleigh Jones. (2014). Wait Times, Patient Satisfaction Scores, and the Perception of Care. *The American Journal of Managed Care*, *20*(5). https://doi.org/10.37765/AJMC

Cole, J. D., & Yount, K. B. (1994). Applications of dynamic simulation to industrial control problems. *ISA Transactions*, *33*(1), 11–18. https://doi.org/10.1016/0019-0578(94)90031-0

Connelly, L. G., & Bair, A. E. (2004). Discrete Event Simulation of Emergency Department Activity: A Platform for System-level Operations Research. *Academic Emergency Medicine*, *11*(11), 1177–1185. https://doi.org/10.1197/J.AEM.2004.08.021

Darema, F. (2004). Dynamic data driven applications systems: A new paradigm for application simulations and measurements. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *3038*, 662–669. https://doi.org/10.1007/978-3-540-24688-6_86/COVER

Davies, R., & Davies, H. (1994). Modelling Patient Flows and Resource Provision in Health Systems. In *Int. J. Mgmt Sci* (Vol. 22, Issue 2).

Doğan, N. Ö., & Unutulmaz, O. (2014). Lean production in healthcare: a simulation-based value stream mapping in the physical therapy and rehabilitation department of a public hospital. *Https://Doi.Org/10.1080/14783363.2014.945312*, *27*(1–2), 64–80. https://doi.org/10.1080/14783363.2014.945312

Duguay, C., & Chetouane, F. (2007). Modeling and Improving Emergency Department Systems using Discrete Event Simulation. *Simulation*, *83*(4), 311–320. https://doi.org/10.1177/0037549707083111

Eriksson, K., Ramasamy, S., Zhang, X., Wang, Z., & Danielsson, F. (2022). Conceptual framework of scheduling applying discrete event simulation as an environment for deep reinforcement learning. *Procedia CIRP*, *107*, 955–960. https://doi.org/10.1016/J.PROCIR.2022.05.091

Eskandari, H., Riyahifard, M., Khosravi, S., & Geiger, C. D. (2011). Improving the emergency department performance using simulation and MCDM methods. *Proceedings - Winter Simulation Conference*, 1211–1222. https://doi.org/10.1109/WSC.2011.6147843

Facchin, P., Rizzato, E., & Romanin-Jacur, G. (2010). Emergency department generalized flexible simulation model. In *2010 IEEE Workshop on Health Care Management (WHCM)*. https://doi.org/10.1109/WHCM.2010.5441240

Forbus, J. J., & Berleant, D. (2022). Discrete-Event Simulation in Healthcare Settings: A Review. *Modelling 2022, Vol. 3, Pages 417-433*, *3*(4), 417–433. https://doi.org/10.3390/MODELLING3040027

Forrester, J. W. (2017). Industrial Dynamics. *Https://Doi.Org/10.1057/Palgrave.Jors.2600946*, *48*(10), 1037–1041. https://doi.org/10.1057/PALGRAVE.JORS.2600946

Furian, N., Neubacher, D., O'Sullivan, M., Walker, C., & Pizzera, C. (2018). GEDMod – Towards a generic toolkit for emergency department modeling. *Simulation Modelling Practice and Theory*, *87*, 239–273. https://doi.org/10.1016/J.SIMPAT.2018.07.010

Gaba, D. M. (2004). The future vision of simulation in health care. *Qual Saf Health Care*, *13*(1), 2–10. https://doi.org/10.1136/qshc.2004.009878

Guala, F. (2002). Models, Simulations, and Experiments. *Model-Based Reasoning*, 59–74. https://doi.org/10.1007/978-1-4615-0605-8_4

Gunal, M. M. (2012). A guide for building hospital simulation models. *Health Systems*, *1*(1), 17–25. https://doi.org/10.1057/HS.2012.8

Harper, P. R., & Shahani, A. K. (2002). Modelling for the planning and management of bed capacities in hospitals. *Journal of the Operational Research Society*, *53*(1), 11–18. https://doi.org/10.1057/PALGRAVE.JORS.2601278

Hay, A. M., Valentin, E. C., & Bijlsma, R. A. (2006). Modeling emergency care in hospitals: A paradox - The patient should not drive the process. *Proceedings - Winter Simulation Conference*, 439–445. https://doi.org/10.1109/WSC.2006.323113

Henderson, S. G., Biller, B., Hsieh, M.-H., Shortle, J., Tew, J. D., Barton, R. R., & Sargent, R. G. (2007). *Proceedings of the 2007 Winter Simulation Conference VERIFICATION AND VALIDATION OF SIMULATION MODELS*.

Hershey, J. C., Weiss, E. N., & Cohen, M. A. (1981). STOCHASTIC SERVICE NETWORK MODEL WITH APPLICATION TO HOSPITAL FACILITIES. *Operations Research*, *29*(1), 1–22. https://doi.org/10.1287/OPRE.29.1.1

Hidayat, R., Syed Zamri, S. N. A., Zulnaidi, H., & Yuanita, P. (2020). Meta-cognitive behaviour and mathematical modelling competency: mediating effect of performance goals. *Heliyon*, *6*(4), e03800. https://doi.org/10.1016/J.HELIYON.2020.E03800

Irwin, M., & Wang, Z. (2017). Dynamic Systems Modeling. *The International Encyclopedia of Communication Research Methods*, 1–12. https://doi.org/10.1002/9781118901731.IECRM0074

Jain, S., Creasey, R. R., Himmelspach, J., White, K. P., Fu, M., & Institute of Electrical and Electronics Engineers. (2011). *Proceedings of the 2011 Winter Simulation Conference*. IEEE.

Joosten, M., de Blaauw, I., & Botden, S. M. (2022). Validated simulation models in pediatric surgery: A review. *Journal of Pediatric Surgery*. https://doi.org/10.1016/J.JPEDSURG.2022.06.015

Kao, E. P. C. (1974). MODELING THE MOVEMENT OF CORONARY PATIENTS WITHIN A HOSPITAL BY SEMI-MARKOV PROCESSES. *Operations Research*, *22*(4), 683–699. https://doi.org/10.1287/OPRE.22.4.683

Karnon, J., Stahl, J., Brennan, A., Caro, J. J., Mar, J., & Möller, J. (2012). Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. *Medical Decision Making : An International Journal of the Society for Medical Decision Making*, *32*(5), 701–711. https://doi.org/10.1177/0272989X12455462

Khanna, S., Sier, D., Boyle, J., & Zeitz, K. (2016). Discharge timeliness and its impact on hospital crowding and emergency department flow performance. *EMA - Emergency Medicine Australasia*, *28*(2), 164–170. https://doi.org/10.1111/1742-6723.12543

Kleijnen, J. P. C. (1995). Verification and validation of simulation models. *European Journal of Operational Research*, *82*(1), 145–162. https://doi.org/10.1016/0377-2217(94)00016-6

Klingstam, P., & Gullander, P. (1999). Overview of simulation tools for computer-aided production engineering. *Computers in Industry*, *38*(2), 173–186. https://doi.org/10.1016/S0166-3615(98)00117-1

Kokangul, A. (2008). *A combination of deterministic and stochastic approaches to optimize bed capacity in a hospital unit*. https://doi.org/10.1016/j.cmpb.2008.01.001

Kovalchuk, S. v., Funkner, A. A., Metsker, O. G., & Yakovlev, A. N. (2018). Simulation of patient flow in multiple healthcare units using process and data mining techniques for model identification. *Journal of Biomedical Informatics*, *82*, 128–142. https://doi.org/10.1016/J.JBI.2018.05.004

Kuo, Y.-H., Rado, O., Lupia, B., Janny, •, Leung, M. Y., Graham, C. A., Kuo, Y.-H., Rado, O., Lupia, B., Leung, J. M. Y., & Graham, C. A. (2016). Improving the efficiency of a hospital emergency department: a simulation study with indirectly imputed service-time distributions. *Flex Serv Manuf J*, *28*, 120–147. https://doi.org/10.1007/s10696-014-9198-7

Law, A. M. (2015). *Simulation Modeling and Analysis, FIFTH EDITION*. www.averill-law.com

Levin, S., Dittus, R., Aronsky, D., Weinger, M., & France, D. (2011). Evaluating the effects of increasing surgical volume on emergency department patient access. *BMJ Quality & Safety*, *20*(2), 146–152. https://doi.org/10.1136/BMJQS.2008.030007

Levin, S. R., Dittus, R., Aronsky, D., Weinger, M. B., Han, J., Boord, J., & France, D. (2008). Optimizing cardiology capacity to reduce emergency department boarding: A systems engineering approach. *American Heart Journal*, *156*(6), 1202–1209. https://doi.org/10.1016/J.AHJ.2008.07.007

Li, M., Lowrie, D. B., Huang, C. Y., Lu, X. C., Zhu, Y. C., Wu, X. H., Shayiti, M., Tan, Q. Z., Yang, H. L., Chen, S. Y., Zhao, P., He, S. H., Wang, X. R., & Lu, H. Z. (2015). Evaluating patients' perception of service quality at hospitals in nine Chinese cities by use of the ServQual scale. *Asian Pacific Journal of Tropical Biomedicine*, *5*(6), 497–504. https://doi.org/10.1016/J.APJTB.2015.02.003

Lin, C. H., Kao, C. Y., & Huang, C. Y. (2015). Managing emergency department overcrowding via ambulance diversion: A discrete event simulation model. *Journal of the Formosan Medical Association*, *114*(1), 64–71. https://doi.org/10.1016/J.JFMA.2012.09.007

Manzoor, F., Wei, L., Hussain, A., Asif, M., & Shah, S. I. A. (2019). Patient Satisfaction with Health Care Services; An Application of Physician's Behavior as a Moderator. *International Journal of Environmental Research and Public Health*, *16*(18). https://doi.org/10.3390/IJERPH16183318

Marshall, A., Vasilakis, C., & El-Darzi, E. (2005). Length of Stay-Based Patient Flow Models: Recent Developments and Future Directions. In *Health Care Management Science* (Vol. 8). Springer Science + Business Media, Inc. Manufactured in The Netherlands.

Mohiuddin, S., Busby, J., Savović, J., Richards, A., Northstone, K., Hollingworth, W., Donovan, J. L., & Vasilakis, C. (2017). Patient flow within UK emergency departments: a systematic review of the use of computer simulation modelling methods. *BMJ Open*, *7*(5), e015007. https://doi.org/10.1136/BMJOPEN-2016-015007

Monks, T., van der Zee, D. J., Lahr, M. M. H., Allen, M., Pearn, K., James, M. A., Buskens, E., & Luijckx, G. J. (2017). A framework to accelerate simulation studies of hyperacute stroke systems. *Operations Research for Health Care*, *15*, 57–67. https://doi.org/10.1016/J.ORHC.2017.09.002

Morgan, M. S. (2005). *Experiments versus models: New phenomena, inference and surprise*. https://doi.org/10.1080/13501780500086313

Mourtzis, D. (2019). Simulation in the design and operation of manufacturing systems: state of the art and new trends. *Https://Doi.Org/10.1080/00207543.2019.1636321*, *58*(7), 1927–1949. https://doi.org/10.1080/00207543.2019.1636321

Murray, S. N., Rocher, B., & O'Sullivan, D. T. J. (2012). Static Simulation: A sufficient modelling technique for retrofit analysis. *Energy and Buildings*, *47*, 113–121. https://doi.org/10.1016/J.ENBUILD.2011.11.034

Paul, S. A., Reddy, M. C., & DeFlitch, C. J. (2010). A Systematic Review of Simulation Studies Investigating Emergency Department Overcrowding. *SIMULATION*, *86*(8–9), 559–571. https://doi.org/10.1177/0037549709360912

Qudrat-Ullah, H. (2012). *On the validation of system dynamics type simulation models*. *51*, 159–166. https://doi.org/10.1007/s11235-011-9425-4

Radzicki, M. J. (2020). System Dynamics and Its Contribution to Economics and Economic Modeling. *System Dynamics*, 401–415. https://doi.org/10.1007/978-1-4939-8790-0_539

Ranjkesh, S. H., Hamadani, A. Z., & Mahmoodi, S. (2019). A new cumulative shock model with damage and inter-arrival time dependency. *Reliability Engineering and System Safety*, *192*, 106047. https://doi.org/10.1016/j.ress.2018.01.006

Renard, P., Alcolea, A., & Ginsbourger, D. (2013). Stochastic versus Deterministic Approaches. *Environmental Modelling: Finding Simplicity in Complexity: Second Edition*, 133–149. https://doi.org/10.1002/9781118351475.CH8

Reynolds, J., Zeng, Z., Li, J., & Chiang, S. Y. (2010). Design and analysis of a health care clinic for homeless people using simulations. *International Journal of Health Care Quality Assurance*, *23*(6), 607–620. https://doi.org/10.1108/09526861011060960/FULL/PDF

Richardson, D. B., & Mountain, D. (2009). Myths versus facts in emergency department overcrowding and hospital access block. *Medical Journal of Australia*, *190*(7), 369–374. https://doi.org/10.5694/J.1326-5377.2009.TB02451.X

Robertson, D. A. (2005). *Article in Academy of Management Learning and Education, The ·*. https://doi.org/10.5465/AMLE.2005.19086798

Robinson, S. (2008a). Conceptual modelling for simulation Part II: a framework for conceptual modelling Conceptual modelling for simulation Part II: a framework for conceptual modelling Conceptual modelling for simulation Part II: a framework for conceptual modelling. *Journal of the Operational Research Society*, *59*(3), 291–304. https://doi.org/10.1057/palgrave.jors.2602369

Rowaei, A. (2011). *Calhoun: The NPS Institutional Archive The Effects of Time Advance Mechanism on Simple Agent Behaviors in Combat Simulations*. http://hdl.handle.net/10945/38629

Saghafian, S., Hopp, W. J., van Oyen, M. P., Desmond, J. S., & Kronick, S. L. (2012). Patient streaming as a mechanism for improving responsiveness in emergency departments. *Operations Research*, *60*(5), 1080–1097. https://doi.org/10.1287/OPRE.1120.1096

Schruben, L. (2008). *ANALYTICAL SIMULATION MODELING*. www.insead.edu/issrw/

Siebers, P. O., MacAl, C. M., Garnett, J., Buxton, D., & Pidd, M. (2017). Discrete-event simulation is dead, long live agent-based simulation! *Https://Doi.Org/10.1057/Jos.2010.14*, *4*(3), 204–210. https://doi.org/10.1057/JOS.2010.14

Singh, D., Padgham, L., & Logan, B. (2016). Integrating BDI Agents with Agent-Based Simulation Platforms. *Autonomous Agents and Multi-Agent Systems*, *30*(6), 1050–1071. https://doi.org/10.1007/S10458-016-9332-X/FIGURES/5

Sinha, R., Paredis, C. J. J., Liang, V. C., & Khosla, P. K. (2001). Modeling and Simulation Methods for Design of Engineering Systems. *Journal of Computing and Information Science in Engineering*, *1*(1), 84–91. https://doi.org/10.1115/1.1344877

Srivastava, H. M. (2003). Queueing Theory. Encyclopedia of Physical Science and Technology, 481–495. https://doi.org/10.1016/B0-12-227410-5/00632-3

Steuben, J. C., Birnbaum, A. J., Michopoulos, J. G., & Iliopoulos, A. P. (2019). Enriched analytical solutions for additive manufacturing modeling and simulation. *Additive Manufacturing*, *25*, 437–447. https://doi.org/10.1016/J.ADDMA.2018.10.017

Takakuwa, S., & Shiozaki, H. (2004). Functional analysis for operating emergency department of a general hospital. *Proceedings - Winter Simulation Conference*, *2*, 2003–2011. https://doi.org/10.1109/WSC.2004.1371562

Tan, K. W., Wang, C., & Lau, H. C. (2012). Improving patient flow in emergency department through dynamic priority queue. *IEEE International Conference on Automation Science and Engineering*, 125–130. https://doi.org/10.1109/COASE.2012.6386409

Urbani, M., Brunelli, M., & Collan, M. (2020). A Comparison of Maintenance Policies for Multi-Component Systems through Discrete Event Simulation of Faults. *IEEE Access*, *8*, 143654–143664. https://doi.org/10.1109/ACCESS.2020.3014147

Vázquez-Serrano, J. I., Peimbert-García, R. E., & Cárdenas-Barrón, L. E. (2021). Discrete-Event Simulation Modeling in Healthcare: A Comprehensive Review. *International Journal of Environmental Research and Public Health*, *18*(22). https://doi.org/10.3390/IJERPH182212262

Wang, J., Li, J., Tussey, K., & Ross, K. (2012). Reducing length of stay in emergency department: A simulation study at a community hospital. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, *42*(6), 1314–1322. https://doi.org/10.1109/TSMCA.2012.2210204

Weiss, E. N., Cohen, M. A., & Hershey, J. C. (1982). An Iterative Estimation and Validation Procedure for Specification of Semi-Markov Models with Application to Hospital Patient Flow. *Https://Doi.Org/10.1287/Opre.30.6.1082*

Welch, S. J., Asplin, B. R., Stone-Griffith, S., Davidson, S. J., Augustine, J., & Schuur, J. (2011). Emergency department operational metrics, measures and definitions: Results of the second performance measures and benchmarking summit. *Annals of Emergency Medicine*, *58*(1), 33–40. https://doi.org/10.1016/J.ANNEMERGMED.2010.08.040

Yin, C., & McKay, A. (2018). *Introduction to Modeling and Simulation Techniques*. 2–06. http://eprints.whiterose.ac.uk/135646/

Zarei, E., Daneshkohan, A., Pouragha, B., Marzban, S., & Arab, M. (2015). An Empirical Study of the Impact of Service Quality on Patient Satisfaction in Private Hospitals, Iran. *Global Journal of Health Science*, *7*(1), 1. https://doi.org/10.5539/GJHS.V7N1P1

Zeng, Z., Ma, X., Hu, Y., Li, J., & Bryant, D. (2012). A Simulation Study to Improve Quality of Care in the Emergency Department of a Community Hospital. *Journal of Emergency Nursing*, *38*(4), 322–328. https://doi.org/10.1016/j.jen.2011.03.005

Zhang, C., Grandits, T., Härenstam, K. P., Hauge, J. B., & Meijer, S. (2018). A systematic literature review of simulation models for non-technical skill training in healthcare logistics. *Advances in Simulation 2018 3:1*, *3*(1), 1–16. https://doi.org/10.1186/S41077-018-0072-7

Zhang, X. (2018). Application of discrete event simulation in health care: A systematic review. *BMC Health Services Research*, *18*(1), 1–11. https://doi.org/10.1186/S12913-018-3456-4/FIGURES/5

Annexes

Appendix 1 – *Results of the KPIs regarding the individual and collective analysis.*

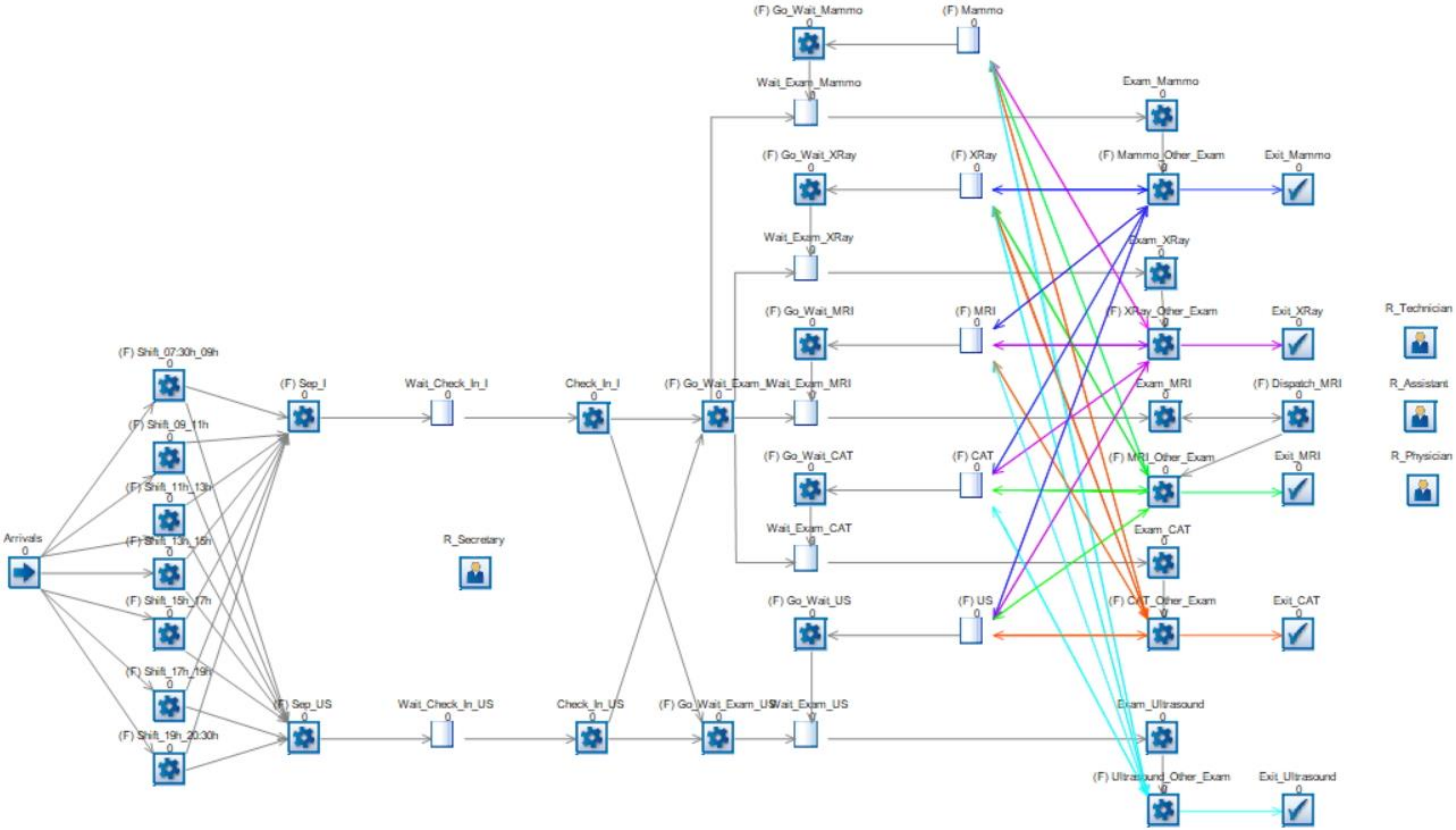
	Activities	1st more significant benefit		2nd more significant benefit	
Number of Replicas	*(A) Exam_Mammo*	3		4	
	(A) Exam_XRay	4		5	
	(A) Exam_MRI	7		9	
	(A) Exam_CAT	5		6	
	(A) Exam_Ultrasound	6		8	
	(A) Check_In_I	6		7	
	(A) Check_In_US	2		3	
Number of Resources	*R_Technician*	19		24	
	R_Assistant	22		28	
	R_Physician	6		8	
	R_Secretary	8		10	
Simulation Object	**Performance Measure**	Individual Analysis	Collective Analysis	Individual Analysis	Collective Analysis
(Q) Wait_Check_In_I	*Average Queuing Time*	1.77	1.77	0.55	0.55
	Maximum Queuing Time	29.60	29.60	17.33	17.33
(Q) Wait_Check_In_US	*Average Queuing Time*	3.90	3.90	0.58	0.58
	Maximum Queuing Time	57.58	57.58	23.64	23.64
(Q) Wait_Exam_Mammo	*Average Queuing Time*	4.51	6.54	0.83	1.33
	Maximum Queuing Time	52.60	77.54	24.90	29.19
(Q) Wait_Exam_XRay	*Average Queuing Time*	3.05	4.39	0.66	1.15
	Maximum Queuing Time	43.31	65.54	18.80	21.94
(Q) Wait_Exam_MRI	*Average Queuing Time*	12.84	17.35	2.42	3.97
	Maximum Queuing Time	97.49	150.70	34.50	52.84
(Q) Wait_Exam_CAT	*Average Queuing Time*	2.90	3.90	0.91	1.32
	Maximum Queuing Time	50.95	49.31	26.25	28.28
(Q) Wait_Exam_Ultrasound	*Average Queuing Time*	5.01	10.01	0.37	1.04
	Maximum Queuing Time	52.74	85.71	12.25	24.54
R_Tecnician	*Utilization (%)*	86.99	77.10	82.13	68.22
R_Assistant		87.20	78.65	82.03	67.97
R_Physician		76.06	78.69	57.06	58.89
R_Secretary		76.31	62.23	66.32	50.59

Appendix 2 – *The KPIs' results in the absence of one resource per experience.*

		2nd more significant benefit			
	Activities	**Initial**	**1 Reduction**	**2 Reduction**	**3 Reduction**
Number of Replicas	***(A) Exam_Mammo***	4	4	4	4
	(A) Exam_XRay	5	5	5	5
	(A) Exam_MRI	9	9	9	9
	(A) Exam_CAT	6	6	6	6
	(A) Exam_Ultrasound	8	8	8	8
	(A) Check_In_I	7	7	7	7
	(A) Check_In_US	3	3	3	3
Number of Resources	***R_Technician***	24	23	22	21
	R_Assistant	28	27	26	25
	R_Physician	8	7	6	5
	R_Secretary	10	9	8	7
Simulation Object	**Performance Measure**	**Collective Analysis**	**Collective Analysis**	**Collective Analysis**	**Collective Analysis**
(Q) Wait_Check_In_I	***Average Queuing Time***	0.55	0.67	1.19	3.87
	Maximum Queuing Time	17.33	18.79	24.43	44.42
(Q) Wait_Check_In_US	***Average Queuing Time***	0.58	0.73	1.12	1.99
	Maximum Queuing Time	23.64	25.82	29.49	35.14
(Q) Wait_Exam_Mammo	***Average Queuing Time***	1.33	1.55	1.44	1.30
	Maximum Queuing Time	29.19	29.61	27.40	46.12
(Q) Wait_Exam_XRay	***Average Queuing Time***	1.15	1.25	1.30	1.21
	Maximum Queuing Time	21.94	19.72	25.52	20.78
(Q) Wait_Exam_MRI	***Average Queuing Time***	3.97	3.75	3.53	3.27
	Maximum Queuing Time	52.84	46.50	42.83	37.34
(Q) Wait_Exam_CAT	***Average Queuing Time***	1.32	1.61	1.47	1.49
	Maximum Queuing Time	28.28	36.16	31.95	44.23
(Q) Wait_Exam_Ultrasound	***Average Queuing Time***	1.04	2.78	10.94	46.82
	Maximum Queuing Time	24.54	35.94	92.17	198.55
R_Tecnician	***Utilization (%)***	68.22	71.45	73.89	76.32
R_Assistant		67.97	70.59	72.66	74.36
R_Physician		58.89	66.97	78.26	91.99
R_Secretary		50.59	56.21	63.23	72.27

Appendix 3 – *The KPIs' results in the absence of three resources, except for the Physician.*

	Activities	Initial	3 Reduction (Except Physician, with 1 Reduction)
Number of Replicas	*(A) Exam_Mammo*	4	4
	(A) Exam_XRay	5	5
	(A) Exam_MRI	9	9
	(A) Exam_CAT	6	6
	(A) Exam_Ultrasound	8	8
	(A) Check_In_I	7	7
	(A) Check_In_US	3	3
Number of Resources	*R_Technician*	24	21
	R_Assistant	28	25
	R_Physician	8	7
	R_Secretary	10	7
Simulation Object	Performance Measure	Collective Analysis	Collective Analysis
(Q) Wait_Check_In_I	*Average Queuing Time*	0.55	3.87
	Maximum Queuing Time	17.33	44.42
(Q) Wait_Check_In_US	*Average Queuing Time*	0.58	1.99
	Maximum Queuing Time	23.64	35.14
(Q) Wait_Exam_Mammo	*Average Queuing Time*	1.33	1.86
	Maximum Queuing Time	29.19	33.62
(Q) Wait_Exam_XRay	*Average Queuing Time*	1.15	1.78
	Maximum Queuing Time	21.94	24.72
(Q) Wait_Exam_MRI	*Average Queuing Time*	3.97	3.57
	Maximum Queuing Time	52.84	54.20
(Q) Wait_Exam_CAT	*Average Queuing Time*	1.32	1.58
	Maximum Queuing Time	28.28	25.09
(Q) Wait_Exam_Ultrasound	*Average Queuing Time*	1.04	2.31
	Maximum Queuing Time	24.54	29.62
R_Tecnician	*Utilization (%)*	68.22	75.90
R_Assistant		67.97	74.33
R_Physician		58.89	67.32
R_Secretary		50.59	72.27



Appendix 4 – *Overview of the model implementation in SIMUL8.*

Other Appendixes may be found at:

https://drive.google.com/file/d/1ED-HuSAEbWsPurZgLsGQfBTUIN2uzkiM/view?usp=share_link