Acta Cybernetica **26** (2024) 637–669.

Patient Flow Analysis with a Custom Simulation Engine*

Zoltán Szabó[ab], Emőke Adrienn Hompoth[ac], and Vilmos Bilicki[ad]

Abstract

Patient flow simulation and analysis is one of the oldest IT -based methods used to optimize patient care processes and hospital management. During the pandemic, interest in this domain suddenly increased due to the various constraints and recommendations to reduce the likelihood of further infections in the hospital. Suddenly, metrics such as the number of patients waiting in the same area, the maximum time a patient could stay in a single room, and the minimum distance between patients became important issues to monitor and optimize. Using data and modelling concepts from various hospitals, our team developed a simulation tool that used bpmn models to define an emergency department. We then modified a single day's usual patient flow with various real-world inspired edge cases to evaluate how the simulated flow would change and which stations would become bottlenecks, where the quality of patient care would deteriorate and rooms would become overcrowded. To execute the models, we developed our own tool based on the open-source Camunda modeling tool and the Business Process Model Notation (BPMN) file format. To execute the generated models, we use our own Python-based execution environment based on the SpiffWorkflow library, which permits extensive logging and extensive customization of the attributes analysed. In addition, the modelling toolkit of Camunda was narrowed down and compiled so that it could be easily used by researchers who are not programmers. In the paper, we present both the modeling process and the scenario design process, as well as the results obtained through the runs, including the maximum waiting times during the model runs and the maximum number of

*This research was supported by the EU-funded Hungarian grant GINOP-2.2.1-15-2017-00073; project no. TKP2021-NVA-09 has been implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NVA funding scheme; project no. II-NKFIH-1528-1/2021 has been implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the II-NKFIH-1528-1 funding scheme. The research was also supported by the Ministry of Innovation and Technology NRDI Office within the framework of the Artificial Intelligence National Laboratory Program (RRF-2.3.1-21-2022-00004).

[a]Department of Software Engineering, University of Szeged, Hungary

[b]E-mail: szaboz@inf.u-szeged.hu, ORCID: 0000-0003-3863-7595

[c]E-mail: hompothemoke@gmail.com, ORCID: 0000-0002-7085-3901

[d]E-mail: bilickiv@inf.u-szeged.hu, ORCID: 0000-0002-7793-2661

DOI: 10.14232/actacyb.299391

patients waiting at once, which allowed us to validate the effectiveness of the framework.

Keywords: healthcare, telemedicine, patient flow, simulation, covid, modeling

1 Introduction

As can be read in the NEJM Catalyst short article [18], patient flow technically defines the total time frame that patients spend in and move through the healthcare system from arrival to discharge. In general, we want this time to be as minimal as possible, apart from the time required for the actual examination, diagnosis, and care processes, without compromising patient and provider quality and satisfaction. Improving the flow is essential as it can reduce the workload of medical staff and patient waiting times, but otherwise overcrowding can occur, patient health can deteriorate, while readmission and mortality rates can increase [18, 9]. Improving patient flow was also an area of research in the 1990s. The World Health Organization (WHO) published a study using patient flow analysis (PFA), which helps researchers examine staff utilization, key patient flow characteristics, resource and financial needs, and emerging problems [20]. Another approach has been variability analysis, which involves dividing variables into groups and then determining how to measure them (e.g., severity of illness can be described as the deviation from a perfectly healthy state). The next step is to reduce or even eliminate any variability that is artificial, as it usually arises from dysfunctional processes. This should already lead to an improvement in patient flow. Further progress can be expected if natural variability is also measured and optimally managed [17]. Our approach was to create a patient flow simulation framework that could account for different variables to calculate and measure potential patient flow. To do this, we collected information on commonly used patient flow measures and the variables that may affect patient numbers in the Emergency Department (ED).

In the State of Art section of this study, we provide a summary of numerous research papers that investigate techniques for quantifying the efficacy of ED. Then, in the Motivation section, we review the modeling methodologies and requirements that have arisen throughout the COVID-19 pandemic, summarize the parameters and elements that have been investigated, and highlight those that are pertinent to our simulation. This is followed by the Methodology section, which describes the operation of the ED we aim to represent as well as the modeling and simulation tools produced. In the Simulation section, we first list the scenarios that we planned and then present and analyze their running results, in each case based on the longest waiting times, the greatest number of patients waiting simultaneously, the trends in patient numbers at each station, the average waiting time per triage level, and the relative duration of each waiting time per triage level. The results of the research are reviewed in the Discussion section. Finally, we present an outlook on future goals for the framework in the Conclusion section.

2 State of Art

According to the literature, different patient flow patterns occur under different circumstances. Kang and Park [12] studied the hourly visit pattern and found a bimodal distribution: the peak flow was from 10:00 to 11:00 and from 20:00 to 21:00. The lowest number of visits was between 02:00 and 08:00. In one Hungarian hospital, patient volumes increase from 8:00 and peak around 12:00. Late night hours are the least visited times, but the workload for staff is fairly constant [26]. When daily visit patterns were the focus, Hitzek et. al. [9] found that the peak in patient numbers occurred on weekends (starting on Fridays, with the highest numbers on Saturdays), holidays, and school vacations. The authors suggest that the explanation may be that people tend to engage in risky activities at these times. Varga et al. [26] also examined the difference between patient numbers on weekdays and weekends: They found similar trends, except that weekend nights were slightly more demanding.

There are also seasonal patterns of visits: Hitzek et. al [9] found the highest numbers of patients in spring and the lowest in fall. In contrast, Won, Hwang, Roh, and Chung [27] found the highest number of asthma patients in the fall, especially in September and October, and the lowest from June to August. They also found that more patients visit the ED in spring from year to year.

Linked to seasonality, but with more focus on the actual temperature Otsuki, Murakami, Fujino, Matsumura and Eguchi [19] found that during cold winters less non-urgent patients visited the ED, suggesting that people are less active in the cold weather. In contrast the warmer summer weather raised the patient numbers.

Heat waves can also impact visits to ED. Schramm et. al [22] published a study of the likely impact of a June 25-30, 2021 heat wave, affecting 10 regions of the U.S. that contain 4% of the population but accounted for 15% of heat-related ED visits. From May to June, there were 3,504 heat-related cases at the ED, 79% of which occurred during the heat wave. The peak was on June 28, when 1,038 patients arrived. In comparison, 2 years earlier on the same day, 9 patients had heat-related problems at the ED.

The usual measures of patient flow are bed occupancy rate (it is also suggested to consider the number of outgoing and incoming patients) [13, 8], transfer time (i.e., the time to prepare the bed for a new patient), and patient transfer (how many patients had to be transferred, how much time and phone calls were required to transfer, etc.). Other ED related measures may include: the time a patient spends in the department from admission to discharge, the actual time it takes to discharge a patient and/or refer them to another department, how many patients were treated in a given time interval, the wait time to see a physician or receive treatment, the number of ambulances transferred to another ED, etc. [8].

For example, Varga et. al. [26] measured how much time elapsed before medical care was initiated between different triage levels. The results showed 3.6 ± 5.8 minutes at the first triage level, 7.0 ± 11.8 and 23.2 ± 26.1 minutes at the second and third triage levels, and 37.8 ± 38.3 and 44.2 ± 43.5 minutes at the fourth and fifth triage levels.

Patient flow analysis has also been used as the basis for many research projects using genetic algorithms and in some cases, machine learning, to solve or optimize scheduling issues at various parts of the hospital process.

Yousefi et al. [28] conducted an evaluation of 38 simulation-based optimization experiments for the ED, published between 2007 and 2019. They have given a bibliographic foundation on the topics discussed, compiled data on the methodologies and tools used, and identified significant trends in the area of simulation-based optimization. They have stated that future research should concentrate on improving the effectiveness of multi-objective optimization problems by reducing their time and labor requirements.

In their study, Yang-Kuei Lin and Yin-Yi Chou [16] examined the difficulty of allocating a set of surgical procedures to many multipurpose operating rooms. They have suggested a redesigned mathematical model and four simple heuristics that ensure the efficient discovery of viable solutions to the examined issue. In addition, they provided four local search processes that may greatly enhance a given solution and used a hybrid genetic algorithm (HGA) that combines initial solutions, local search procedures, and an elite search technique to the examined issue.

El-Bouri et al. [7] conducted a literature study on the use of Artificial Intelligence (AI) to hospital patient scheduling. They addressed the many AI strategies described in the literature, such as rule-based systems, decision trees, artificial neural networks, and evolutionary algorithms. In addition, they have examined the many sorts of patient scheduling challenges that have been investigated, including surgery scheduling, appointment scheduling, and emergency department scheduling.

Seunghoon Lee and Young Hoon Lee [14] have suggested using reinforcement learning (RL) to schedule emergency department (ED) patients. They have developed a mathematical model and a Markov decision process (MDP). Then, they developed an RL algorithm based on deep Q-networks (DQN) to identify the ideal scheduling strategy for patients. In the provided cases, they have shown that deep RL outperforms dispatching rules in terms of reducing the weighted waiting time of patients and the penalty score for emergency patients.

Haya Salaha and Sharan Srinivas [21] investigated the usage of a hybrid artificial intelligence system to solve the issue of hospital patient scheduling. To enhance patient scheduling, they have presented a mix of genetic algorithms and an Artificial Neural Network (ANN). They have shown that their hybrid approach can find superior schedules than either Generic Algorithm (GA) or ANN alone, and it has been applied to actual hospital data.

3 Motivation

Research, optimization, and various IT solutions played an important role during the COVID-19 pandemic. While various impacts and metrics of the pandemic itself are still being researched and evaluated by various research teams, another very active area is focused on preparing existing systems to work better and more efficiently in the event of another pandemic.

One need that most research teams agree on is the need for modeling and simulation tools. Currie et. al [4] in their work emphasized the importance of simulations to reduce the impact and severity of the epidemic COVID. They identified the following decision areas as appropriate for optimizing their effectiveness through simulations: the selection of quarantine and isolation strategies, the development of social distancing rules, the construction of lockdown release scenarios, the appropriate method for test distribution and transport, the identification of the most critical demographic groups for vaccine distribution, and the appropriate expansion and allocation of hospital resources.

Similar comments were made by Dieckmann et al. [5], whose work focused on the resources needed for effective simulation and how they can be used. In their view, simulations should focus on three main areas: educating workers about the epidemic, optimising the process of care at the system level, and assessing the needs and mental health workload of health care workers.

Improving hospital systems and patient flow to provide faster patient treatment, efficient resource allocation, and the development of techniques to avoid future infections lies at the junction of the two fields of study. Tavakoli et al.[24] recently published their results on a simulation methodology similar to ours. Although the model and triage levels are much simpler than they should be to prove accurate in simulations of Hungarian hospitals, the metrics and principles established can serve as a model for similar simulations. Terning et al.[25] had similar elements in mind, and although the simulation from their published work is still relatively rudimentary, the formulas and conditions used to evaluate their results provide a very good basis for initial validation of a similar simulation.

One of these key parameters, perhaps the easiest to follow in simulations, is to avoid overcrowding, i.e., to avoid the kind of patient flow where many patients are waiting in an area at the same time. Dinh et al. [6] specifically focused on this importance in their work, attempting to establish principles and rules to avoid unnecessary hospitalizations during an epidemic and to reduce the length of stay in the hospital. In their brief review, Janbabai et al. [10] focused on protecting hospital staff in addition to patients, focusing on preoperative, intraoperative, and postoperative processes within the patient flow. Of course, other approaches have been explored in addition to simulation-based patient flow study and analysis. For example, Arnaud et al, [2] have attempted to use machine learning based on patient flow metrics to determine how to optimise the number of hospital beds and expedite the triage process, to name a few examples.

In the development of various healthcare applications for hospitals and research teams, our team has used the work of Prof. Jose L. Jimenez & Dr. Zhe Peng [11] who, based on various peer-reviewed research, developed an easy-to-use tool to measure the likelihood of COVID infection in different environments based on the size and type of the area in question, as well as the number, behaviour, and condition of the people in it. Based on these results, and taking into account the fact that patients and staff wear masks in the hospital and hospitals use various distancing measures and restrictions, including a strong emphasis on ventilation, our team calculated that the probability of infection for a number of 10 to 20 patients in the area was only 4.39 % after one hour, and even after six hours it only increased to 5.96 %. This means that one of the most important aspects of optimising patient flow for COVID prevention is to keep the number of patients in a given range around or below 10 while trying to speed up the flow itself to avoid congestion.

4 Methodology

4.1 Introduction of the ED

During the early phases of our research, we used the ED model of Leva and Sulis [15], as it proved to be the model most similar to the Hungarian ones based on comparisons between this model and our team's experience and knowledge of the structure and functioning of the ED. Differences include minor changes in terms of which station is served by which staff member, and the introduction of an additional fifth triage level as mandated in the Hungarian system. The model includes 7 different lanes and a sub-process for handling the complex visit process if required by the patient's condition. The first lane is the registration process, where patients are admitted to the hospital and treated according to the severity of their condition upon arrival. They are then admitted to the triage lane where they await initial assessment. For less urgent cases, they may voluntarily leave the process at this stage if they wait too long.

After leaving triage, the triage nurse may decide to refer the patient to an internal clinic; otherwise, the visit process takes place, where the nurse or physician takes a history, takes blood, performs a radiology referral, and then decides the patient's fate based on the results. The outcome of the process may be referral to an internal clinic, admission to the emergency department, referral to an outside facility, discharge, or in a small percentage of cases, death of the patient.

4.2 Modeling and Simulation Tools

To create the hospital simulation, we chose the open-source Camunda Modeler [1], which allows us to create arbitrary processes in a parameterizable, editable, and executable format. The output of Camunda modelling is the BPMN (Business Process Model and Notation) file [3], a text file based on the XML standard that is displayed by Camunda-compatible tools and execution environments with a visual representation.

However, Camunda and BPMN modelling do not always prove suitable. In our various healthcare projects, the issue of clarity and complexity of modelling has often arisen in similar cases, especially when some clinicians and researchers wanted to model and describe processes in a way that was transparent to them, but the Camunda elements were considered too broad and complex. Our goal, therefore, was not only to accurately model the model we created, but also to make it understandable to researchers outside of IT and be able to create similarly simple processes, leaving the more complex parts to scripts and programmers running in the background.

While developing the simulation, we considered using the official Camunda simulation tools and Visual Paradigm, among others. However, we ultimately decided to create our own simulation environment using open source tools to ensure that the simulation settings, configurations, and types of metrics collected were customizable for us.

Our custom simulation is based on the Python library SpiffWorkflow [23], which can process and run bpmn models created with Camunda, among many other inputs. Scalability and robustness were key elements of the hospital workload modelling operating environment. The principle is based on the idea that each patient is a parallel running SpiffWorkflow thread sharing common resources for which we implemented waiting, handover and reservation using semaphores. The measurement and logging of wait and turnaround times in the system is differential, with each thread regularly logging its timestamps as it arrives at and departs from the stations. Our approach was initially based on the naive assumption that the bottleneck is the availability of staff in the ED, and that if the required physician, nurse, or nursing staff is available to perform the task, then the space and equipment are available as well.

Figure 1 illustrates the components of the framework and their precise relationships. The model defining the ED is provided in two bpmn files: Visit.bpmn for the visit subprocess and EDAsIs.bpmn for the ED architecture, which references Visit in the correct place. The framework's starting point is runner.py, which specifies how many patients must be allowed into the system for the simulation, what stop condition must be satisfied to terminate the simulation, and also manages the extraction of the various metrics at the conclusion of the simulation (the latter activity is expected to be handled by a separate module in a future version). The runner.py parses the contents of the bpmn files and utilizes them to generate runner threads for each patient that will execute the steps specified in the bpmn files. The simulations employ playbook.py to execute the simulation of each step and simulation.py to indicate when a shared resource (e.g., doctor, nurse) is required, lock it using semaphore, or set a triage level-based queue if there are no available instances of that resource.

4.3 Modeling

The following section presents the model and elements of the Camunda workflow based on the combination of the Leva and Sulis paper with elements from the Hungarian hospital system. Table 1 shows the content of the first two lanes, registration and triage, the first stations that are the same for every patient in the hospital. The elements of the simulation script are handled either as events, where patients must acquire a shared resource and then perform some processing before proceeding, or as end states, which, when reached, terminate the patient's simulation thread. The required resource in the simulation is a member of the ED staff: Generic Nurse (GN), Hospital Employee (HE), Specialized Nurse (SN), Doctor (DR) and Generic

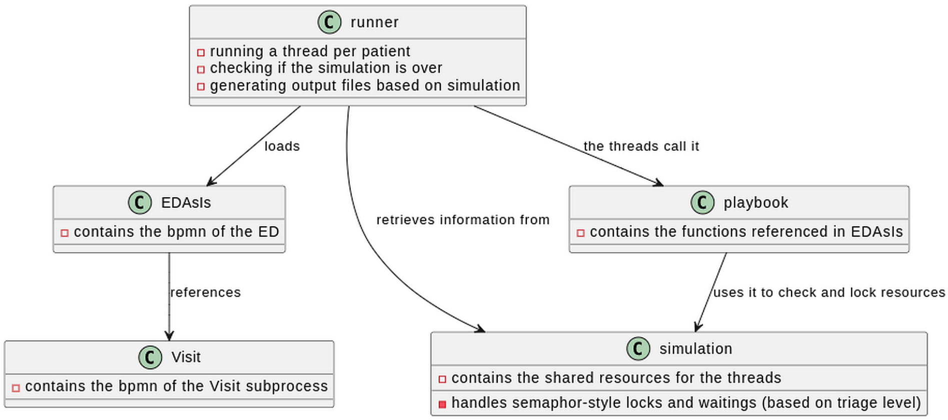

Figure 1: Flowchart of the simulation framework

Operator (GO).

Note that due to the deterministic nature of the simulation, these stations and steps model how a patient is admitted to the hospital and then treated, with the environment assigning almost the entire pathway to the patient at the beginning of the simulation with all the important attributes. Our research team had two main reasons for this: On the one hand, the methodology gives researchers who might use our tools in the future the ability to analyze and debug the expected runtime of the simulation without having to wait for the entire simulation to run. On the other hand, it also gives us the ability to manually enter patients into tables in order of arrival with their severity, and even to examine specific cases in minute detail using the tools. The modeling of these pathways raised a serious research question at the beginning, as the international literature and the original source of this model indicated that the necessary personnel for these pathways are the general nurses, and in order to keep the simulation accurate, we decided to stick with this version. However, in Hungarian hospitals it is much more common to have at least one physician present during these phases. In our further research, collecting more specific information from Hungarian hospitals, including those we have conducted research with, we want to test different modifications of this trace and see how they might change some of the results and trends we have obtained during our previous research. The next major step is the Visit, which is modeled as a separate subprocess. The elements of the Visit model can be seen in Table 2. The main difference from the main lanes is the need for specialized nurses and doctors, and the many optional pathways depending on whether blood tests or radiological examinations are required.

After the Visit subprocess, the only step left in the simulation is the processing of the outcome, which is usually performed by a specialised nurse (SN). There are five possible outcomes defined both in the paper containing the basic version of this model and in the papers analysing Hungarian hospitals: Death, Hospitalisation on

Table 1: Major events of the main ED flow

Name	ID	Description	Type	Req
Patient Registration	register_patient	Admitting the patient to the hospital	event	HE
Evaluate Urgency	evaluate_urgency	The urgency, as a binary information is determined for the patient. Urgent cases are immediately forwarded to pre-visit	event	GN
Abandon	abandon	The patient decides to leave the hospital before pre-visit	end_state	-
Pre-visit	pre_visit	Quick measurements and examination by a nurse to determine the triage level and severity of the case.	event	GN
Assign ESI	assign_esi	Emergency severity index is assigned to the patient, who is either transferred to internal clinic or sent to the full visit process	event	GN
Manage outcome	manage_outcome	Management of results and outcomes resulting from the visit process.	event	SN

Ward, Discharge, Transfer to External Facility, or Transfer to Internal Clinique.

We also achieved the desired simplification in modelling. Since the model was not overly complex, we used only four elements that were visually and practically comprehensible: the start point, the end point, the event, and the decision point. These were simply augmented during design with information about which event gave the patient which additional attributes, and the decision points were then used to select exactly what criteria the patient should use to choose the direction of travel in the simulation. The entire modelling process thus consisted of a total of four elements, plus a few lines of pseudocode description for the events, which is not only simple, but also compatible and interoperable with many other modelling tools. The timer event was considered as a fifth element type, but it was ultimately

Table 2: Elements of the Visit subprocess

Name	ID	Description	Type	Req
Collect History	collect_history	The nurse collects and organizes the healthcare history of the patients via direct interview with the patient and/or database queries	event	SN
Hypothesize Diagnosis	hyp_diag	The doctor evaluates the results so far, examines the patient and either establishes a diagnosis or may require further tests (blood tests and X-rays) before doing so.	event	DR
Take Blood Sample	take_blood	Taking a blood sample as prescribed by the doctor.	event	SN
Laboratory	laboratory	The blood sample is transferred to laboratory examinations. If there is any information available quickly, it is sent back to the doctor.	event	DR
Transfer to radiology	trans_rad	The patient is transferred to radiology and prepared for the X-ray recordings.	event	GO
Radiology	radiology	Usage of diagnostic imaging procedures, such as ultrasound, CT, MR.	event	SN
Establish diagnosis	estab_diag	Based on the results and the information, diagnosis is established	event	DR
Define therapy	def_therap	If possible, therapy is defined by the doctor	event	DR

ruled out due to the system's standardized time management. All wait periods are supplied to the framework as arguments or thresholds with defined values. SpiffWorkflow's scripting and customisation options are restricted in this domain, and the timer event would be conducted in real time regardless of the simulation's time format.

5 Simulation

Our goal was to study how an emergency department ideally operates and how unexpected events can occur, using this operating environment and the modelled emergency department with the number of patients arriving, the severity of their cases, the probabilities and rates for each branch from real data. Or in the case of an epidemic, how to optimise turnaround and wait times (since similar studies in many cases have only looked at similar models in terms of staff time or budget): Is it clearly a good idea to increase staff and the number of rooms and equipment needed to perform each activity?

5.1 Scenarios

To be able to create different situations and scenarios to analyze how small changes in patient flow, staffing, or processing time of the different stages affect the simulation throughput and metrics, we first created a baseline scenario based on real data from the ED of Somogyi Kaposi Mór Practicing Hospital [26] to estimate the rates of patient arrival and distribution between the five triage levels (i.e., the urgency of each case) to model. According to their data from 2015 statistics, the ED sees approximately 90 patients per day. In terms of triage levels, 0.67% of patients had triage level 1, 1.24% had triage level 2, 23.35% had triage level 3, 40.17% had triage level 4, and 34.54% had triage level 5. Triage levels 1 and 2 require immediate treatment, level 3 can tolerate waiting times up to 30 minutes, level 4 up to 60 minutes while level 5 even up to 2 hours.

As for the fate of the incoming patients after treatment: 20.9% were hospitalised, 2.7% voluntarily discharged, 1.5% were referred for triage, 0.4% were transferred to another inpatient facility, 0.4% died and 73.5% were discharged to their home.

The baseline scenario was based on the work of Leva and Sulis and was run with 3 doctors, 2 generic nurses, 3 specialist nurses, 2 clinical staff and 4 generic operators, with a 20% chance of a new patient arriving every minute - this resulted in the most even distribution, the element of the simulation to handle increasing or decreasing patient arrival density at given times is currently being tested and will be included in a next pilot phase. The turnaround times at each station, which depend on the triage level, follow the one-to-one model of Leva and Sulis, considering triage level 3 as the dividing line between urgent and less urgent cases. For all other scenarios, these original distributions and proportions were shifted through a type of exacerbated bias. In some cases, we increased the severity of incoming patient cases, in others we reduced the number of emergency department staff, and in still other scenarios, to approximate the impact of COVID, we used estimations of the need for and duration of decontamination to increase wait and turnaround times at each station in the simulation. The type and number of staff in the emergency department was based on the paper by Leva and Sulis. The various scenarios, their specific configurations and expected results are listed below.

- **SC0**: This is the basis of comparison made by merging of the Somogyi Hospital and the Italian sample. Patients are rarely admitted for urgent triage 1 or 2. The time spent at each station follows the original pattern drawn from the papers. **Specification**: 3 doctors, 3 generic nurses, 2 specialized nurses, 2 clinical staff, 4 generic operators; regular processing times; triage distribution: l1-0.00673, l2-0.01241, l3-0.23359, l4-0.40175, l5-0.34549; 20% patient arrival chance. **Expectation**: Patients with lower triage levels have to wait longer at common stations (registration, triage), where congestion and waiting times increase, but the time spent in the system remains within acceptable limits.

- **SC1**: Scenario inspired by red letter holidays. The staff in the Emergency Department has been significantly reduced, the density of incoming patients is lower, but the incoming patients are almost without exception more urgent, with a triage level higher than 3. **Specification**: 1 doctor, 1 generic nurse, 1 specialized nurse, 1 clinical staff, 2 generic operators; regular processing times; triage distribution: l1-0.0873, l2-0.1241, l3-0.23359, l4-0.00417, l5-0.00345; 20% patient arrival chance. **Expectation**: Even for triage level 3 patients, waiting and turnaround times will increase, while for triage level 1 and 2 cases, times will not be dramatically reduced.

- **SC2**: Summer heat, heatwave inspired scenarios. Patient arrival density is doubled compared to SC0, with a significantly higher probability of arriving triage level 1 and 2 patients. **Specification**: 3 doctors, 2 generic nurses, 3 specialized nurses, 2 clinical staff, 4 generic operators; regular processing times; triage distribution: l1-0.1873, l2-0.2241, l3-0.43359, l4-0.00417, l5-0.00345; 20% patient arrival chance but checked at every half minute instead of every minute. **Expectation**: Significant congestion at higher triage levels, with waiting times of several hours for patients at lower triage levels.

- **SC3**: An epidemic-inspired scenario. With only urgent patients coming to the hospital (less urgent cases are not even admitted), the number of staff in the Emergency Department has been increased, but also the minimum waiting and turnaround times due to disinfection procedures. **Specification**: 6 doctors, 4 generic nurses, 6 specialized nurses, 8 clinical staff, 4 generic operators; processing times are increased with a few minutes to simulate disinfection; triage distribution: l1-0.1873, l2-0.2241, l3-0.13359, l4-0.00417, l5-0.00345 ; 20% patient arrival chance. **Expectation**: Barely any patients from lower triage levels, but significantly increased times for higher levels.

- **SC4**: A benchmarking scenario, using the staff number of SC3 to handle the patient load of SC2 **Specification**: 6 doctors, 4 generic nurses, 6 specialized nurses, 8 clinical staff, 4 generic operators; regular processing times; triage distribution: l1-0.1873, l2-0.2241, l3-0.43359, l4-0.00417, l5-0.00345; 20% patient arrival chance but checked at every half minute instead of every minute. **Expectation**: Critical waiting and turnaround times decreased, the churn and trends of the flow will be similar to the ones of SC0

5.2 Results

Each scenario was run with a load of 90 patients (the daily average based on Somogyi Hospital data) and was intended to fill a 7-8 hour shift in the emergency department.

5.2.1 Scenario0

At SC0 we immediately noticed some interesting differences compared to our first hypothesis. As seen in Figure 2, the Diagnosis Establishment phase had the longest combined times (i.e., waiting and execution combined), while other elements of the first two lanes, such as registration and urgency evaluation, were relatively short.

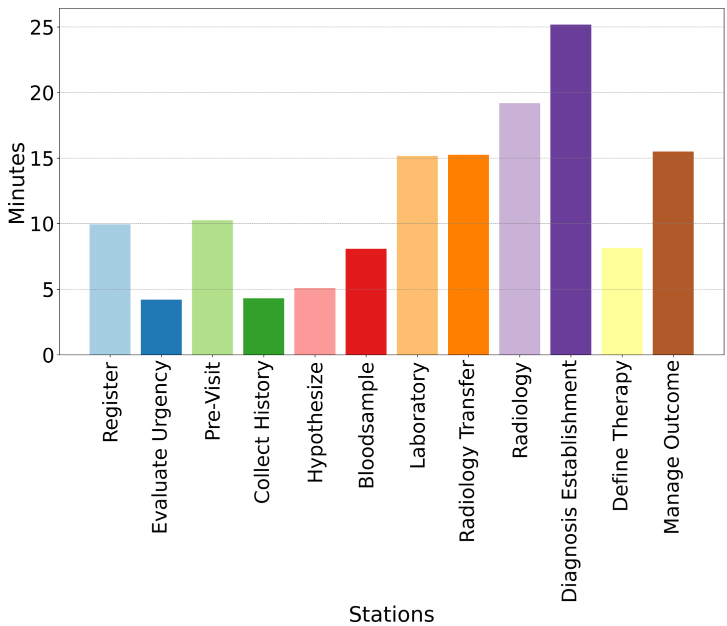

Figure 2: Longest combined times in SC0

Similarly, the maximum number of patients either waited or were studied at the same stage of their simulation. As shown in Figure 3, the longest queues were in

Radiology, Diagnosis Establishment, and Manage Outcomes after the visit.

A more detailed trend chart showcasing the top 3 stations based on maximum number of waiting patients and variability can be seen in Figure 4, which shows how the number of patients at the stations with the strongest bottleneck effect in the given scenario has changed over time in the simulation.

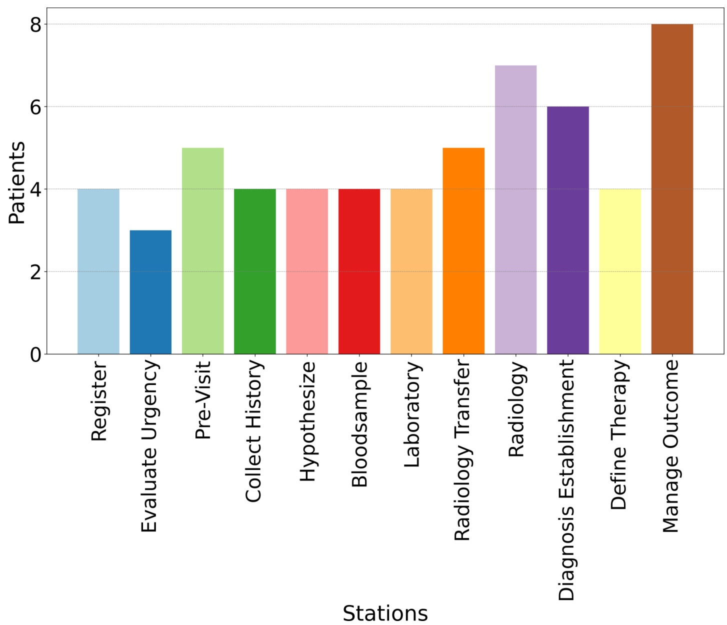

Figure 3: Maximum number of waiting patients in SC0

As for the comparison between the various triage levels, these values can be seen on Figures 5 and 6.

Based on the distributions and probabilities of the Hungarian hospital, not a single triage level 1 patient was admitted to ED during the simulated day. Triage stage 2, of course, had the shortest average time, while others had significantly longer times, with a bottleneck in the outcome management phase after the visit process.

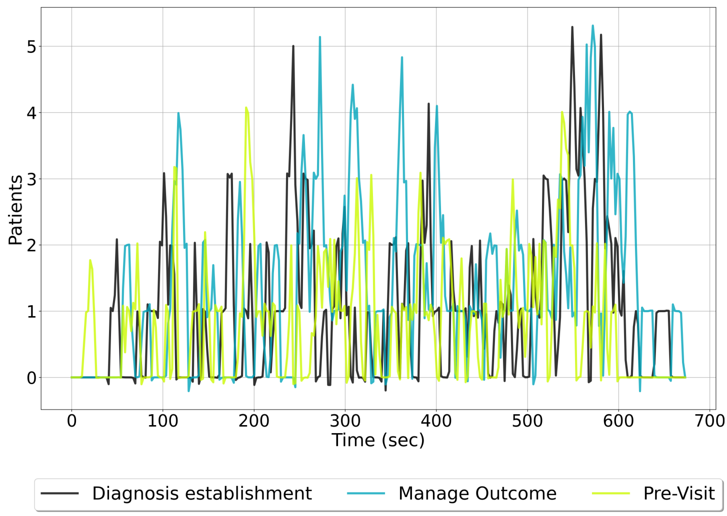

Figure 4: Patient number trends in SC0

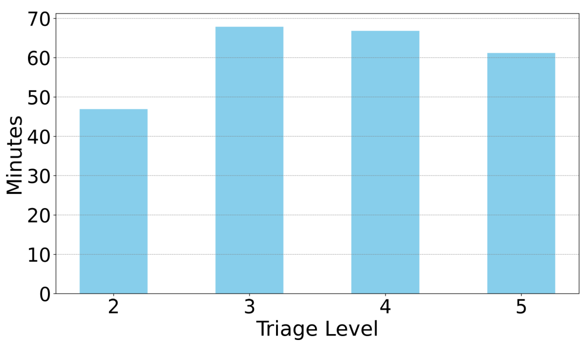

Figure 5: Average times spent in the simulation per triage level in SC0

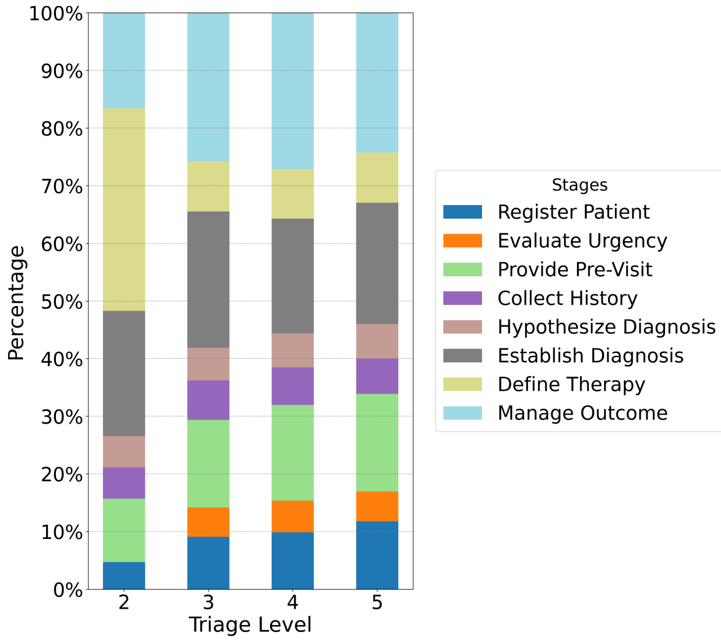

Figure 6: Distribution of times spent at the various stations per triage level in SC0

5.2.2 Scenario1

As with Scenario 1, reducing emergency department staffing and increasing the proportion of urgent patients has led to some interesting results in waiting and turnaround times. As can be clearly seen in Figure 7, the difference from the SC0 results is that the maximum aggregate waiting and turnaround times at each station are almost minimal.

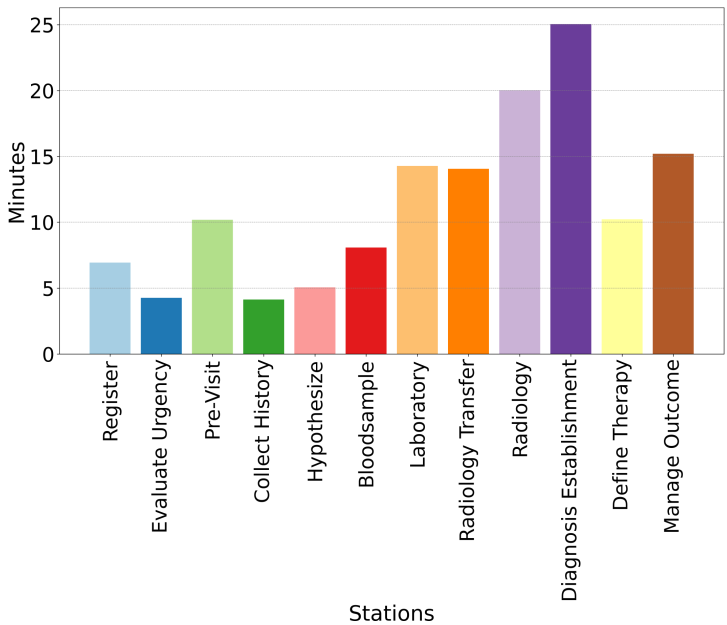

Figure 7: Longest combined times in SC1

The durations hold despite the fact that, as shown in Figure 8, the number of patients waiting at the same time has almost halved compared to the outliers in SC0 due to lower arrival times (with the outliers being the Pre-Visit, Manage Outcome, and Diagnosis Establishment).

Examining the top 3 trends in Figure 9 supports our hypothesis that lower arrival density reduced the number of patients waiting at a site, even at higher triage levels. Here, the peaks occurs when most patients need a clinician (in the simulated case, when preparing or performing radiology examination and during the outcome management), as only one of the clinicians in each role (except the generic operator) is on call during SC1. However, even at the peak, the weighted trend reaches a lower maximum compared to SC0.

The averaged times per triage stage in Figure 10 also confirm one of the key expectations of the simulation. While triage stages 1 and 2 are proportionally faster than the other stages, they are not exceptionally fast-for example, the average duration of Triage Level 5 is not significantly higher than Triage Level 1 (the high

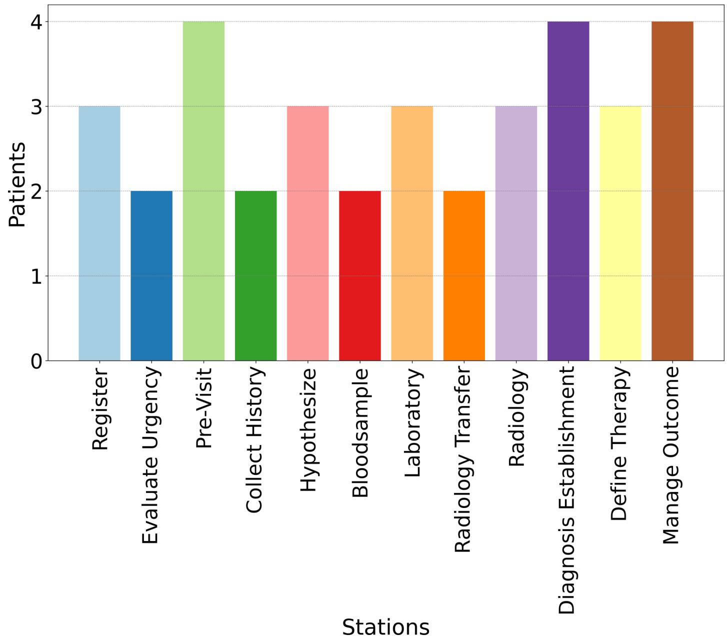

Figure 8: Maximum number of waiting patients in SC1

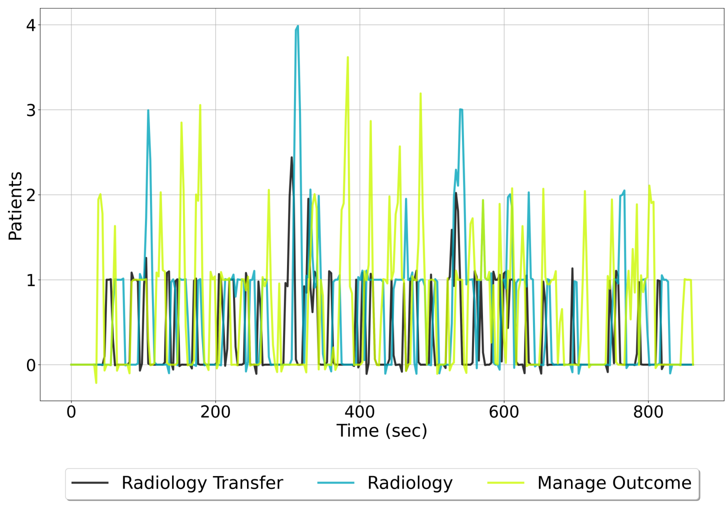

Figure 9: Patient number trends in SC1

values for Triage Levels 3 and 4 may be due to the longer path they follow in the simulation in addition to the arrival time).

They are treated earlier than other patients, they have to wait much less at

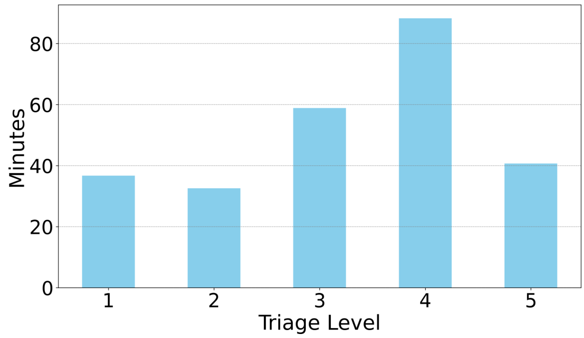

Figure 10: Average times spent in the simulation per triage level in SC1

many bottleneck stations, as can be seen in Figure 11, but their waiting times are not much shorter than those of other patients, and due to congestion and resource constraints, it is not even fully guaranteed that, for example, a level 1 patient will be treated faster than a level 2 patient, because some critical processes cannot be interrupted.

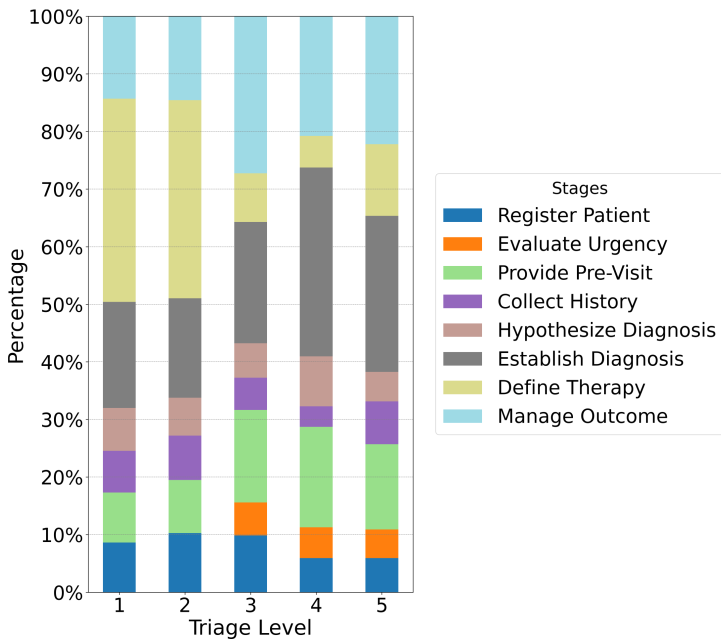

Figure 11: Distribution of times spent at the various stations per triage level in SC1

5.2.3 Scenario2

Scenario 2 represented the next level of workload complexity in patient flow. The goal was to see if the occurrence of higher triage levels and the reduction in hospital staffing would have better highlighted and made visible certain bottlenecks in the model, what the effect would be if more urgent triage levels were more likely than average, and if, despite being fully staffed, the mass of patients arrived in the ED much faster, about twice as fast as normal. And the effect proved to be very interesting. As can be seen immediately in Figure 12, wait times for the stations identified as bottlenecks in the previous scenarios shrink to nearly insignificant amounts compared to registration, where some patients would have to wait up to an almost unrealistic four hours to even be admitted.

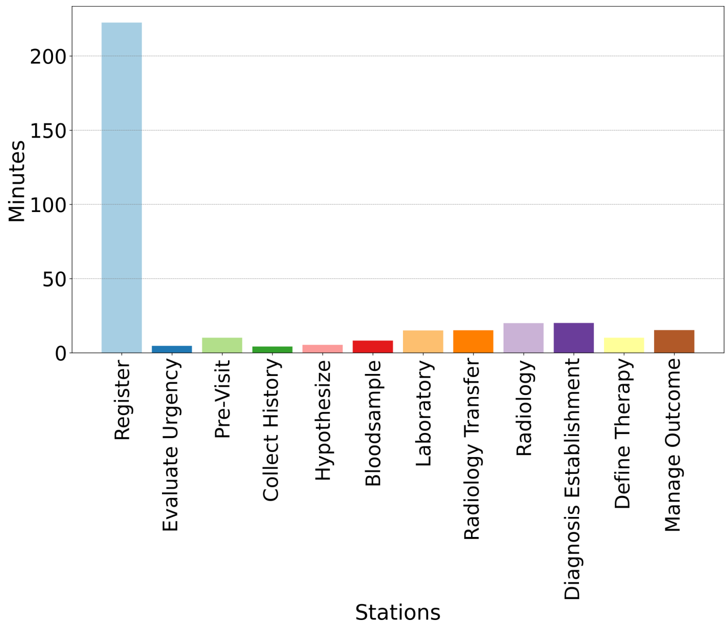

Figure 12: Longest combined times in SC2

This shift is also reflected in the maximum number of patients waiting at the same time, as shown in Figure 13. Diagnosis Establishment, Manage Outcome, and Pre-Visit stations in the simulation can still be considered as outliers, but due to crowding at patient admission, they are significantly exceeded by register_patient and evaluate_urgency.

The exact cause and trajectory of overcrowding are also clearly evident in the weighted trends shown in Figure 14. The number of patients waiting at the same time was slightly higher at the beginning than in the previous scenarios, but around the middle of the simulation there was a huge increase in overcrowding that affected subsequent phases.

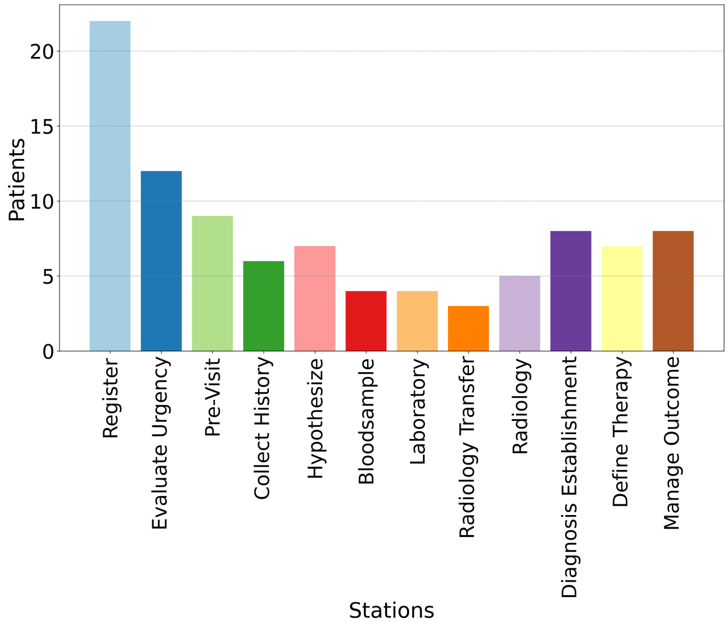

Figure 13: Maximum number of waiting patients in SC2

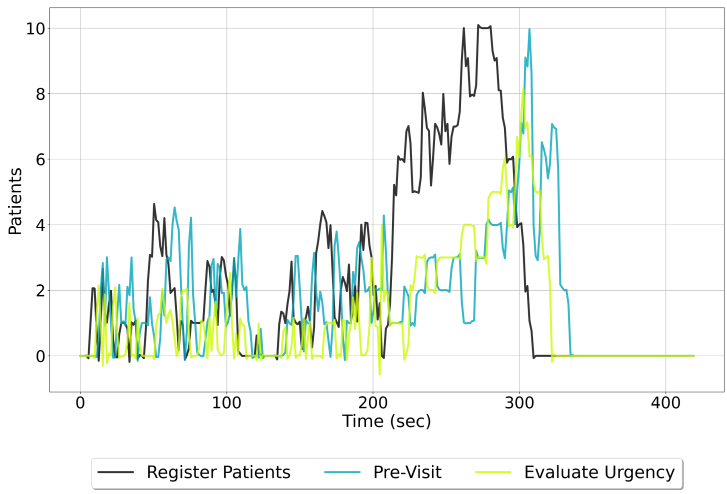

Figure 14: Patient number trends in SC2

However, Figures 15 and 16 also show that prioritisation of triage levels was adhered to despite system congestion. The average turnaround time for patients in triage levels 1 and 2 was remarkably fast in the simulation, spending a relatively large amount of time primarily at the Register and Define Therapy stations, while patients in less urgent cases were in the simulation for up to three to four hours, with a significant amount of waiting time spent at the registration desk.

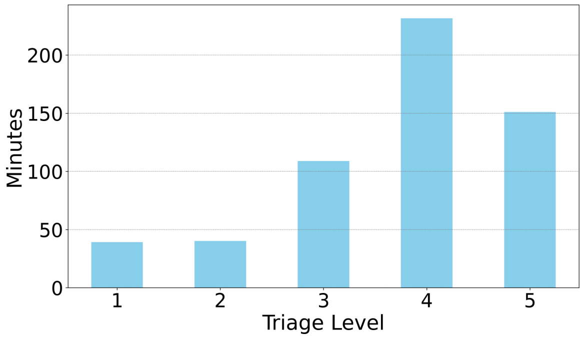

Figure 15: Average times spent in the simulation per triage level in SC2

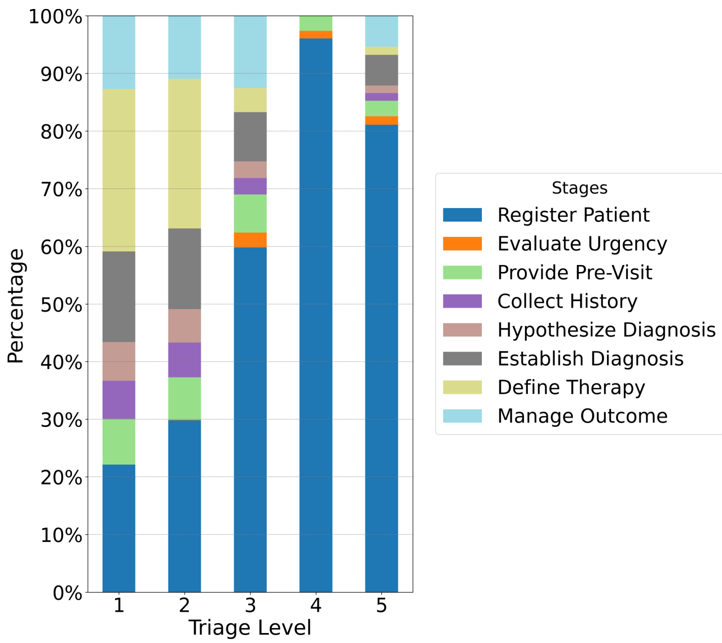

Figure 16: Distribution of times spent at the various stations per triage level in SC2

5.2.4 Scenario3

Scenario 3 was the most critical simulation and the most important for our further research, as we attempted to create a patient flow whose points and biases reflected the characteristics of an actual epidemic compared to the base case. In this case, the emergency department was visited only by urgent patients, typically with a triage level of 3 or higher, and the flow was slowed by the fact that although the number of staff was increased, the passage through the phases was much slower because of mandatory decontamination. Figure 17 shows the primary consequence: average wait and turnaround times per station are significantly longer than for the original cases (especially considering that the majority of cases here required urgent care).

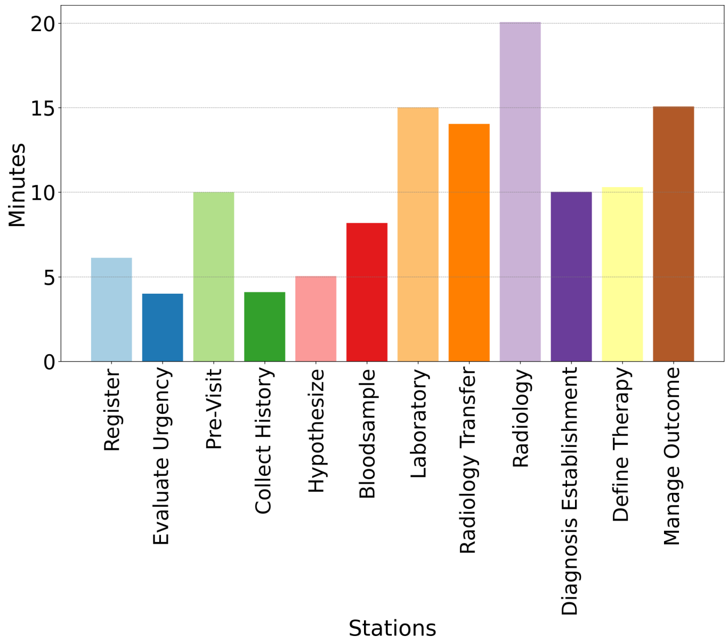

Figure 17: Longest combined times in SC3

The number of patients waiting at the same time has also increased significantly, as can be seen in Figure 18. In addition to the bottlenecks defined so far, the one that stands out is the Hypthosize, the step in the visit process where the patient is first seen by a physician in our model rather than by various nurses and generic operators.

This increase can also be seen in the top 3 trends in Figure 19, which are not only much higher than the results in the previous scenarios, but the peak is not as much of an outlier point as in previous scenarios, but an extended phase that takes up a significant portion of the simulation runtime. In other words, the congestion problem started much earlier, and as the later phases slowed, the number of patients

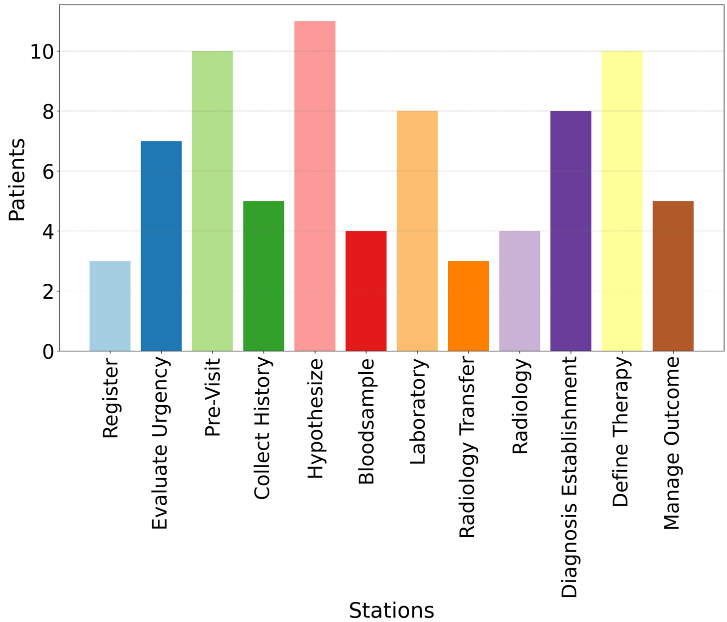

Figure 18: Maximum number of waiting patients in SC3

in the backlog did not start to decline as much as in the earlier cases, even though there should have been more staff available and the patients would have warranted a faster process due to the high triage levels.

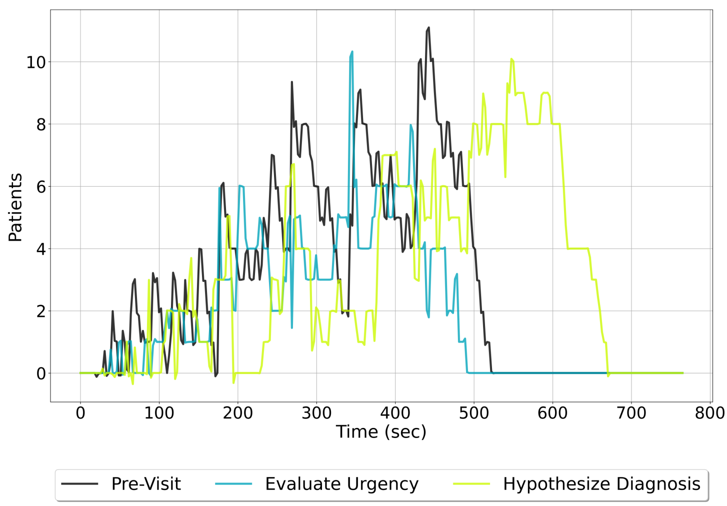

Figure 19: Patient number trends in SC3

Figures 20 and 21 also confirm that, as expected, almost exclusively patients with a triage level of 3 or more were admitted to the emergency department. On average, levels 1 and 2 were completed within an hour. As for the time distribution, it is interesting to note that it is quite similar for the three triage levels, with a significant proportion being spent in Register Patient, Define Therapy and Establish Diagnosis.

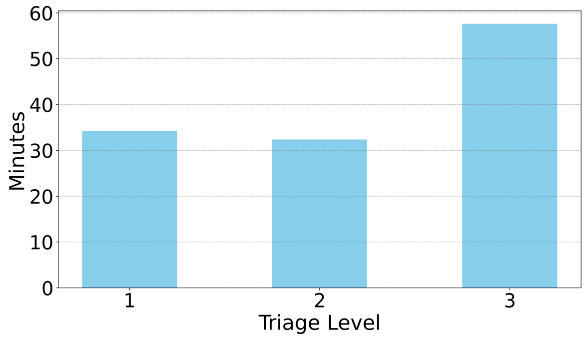

Figure 20: Average times spent in the simulation per triage level in SC3

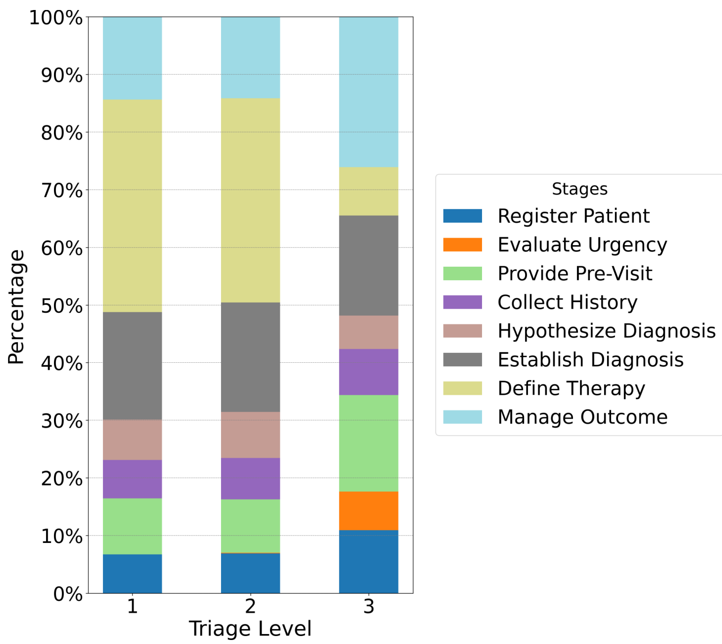

Figure 21: Distribution of times spent at the various stations per triage level in SC3

5.2.5 Scenario4

Based on Scenario3, then, it looked strongly as if simply increasing staffing would not lead to proportionally faster patient flow in the emergency department model. However, in our analyses, we felt that this should not be a generalisation. Therefore, for the fourth scenario, we essentially combined two earlier scenarios by combining the most extreme and in some respects worst Scenario2 with the significantly increased staffing of Scenario3. Figure 22 already shows that the average waiting and turnaround times are much more similar to the baseline Scenario0 than to the original Scenario2.

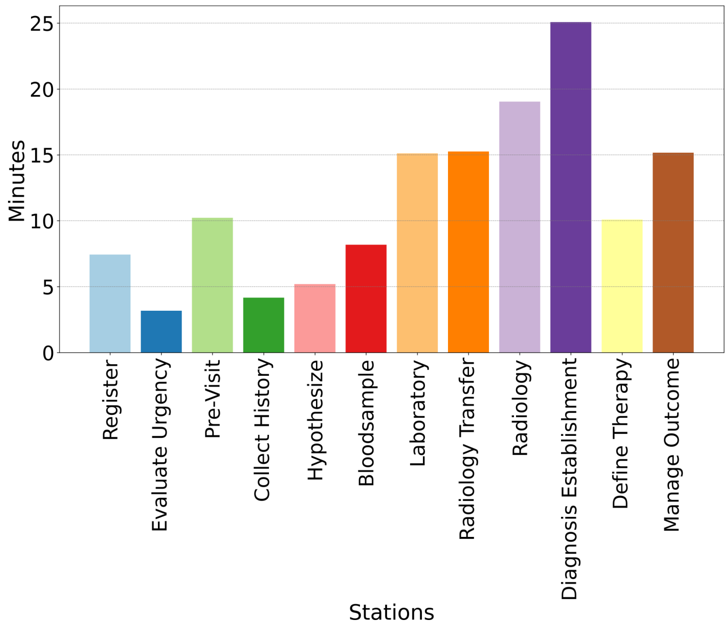

Figure 22: Longest combined times in SC4

As can be seen in Figure 23, the metric of most patients waiting in one place at one time have also changed: Although most stations still have 4-6 patients waiting at the same time, the number is much more balanced because bottleneck stations no longer have as many patients stopped at the same time as they did originally.

This smoothing and apparent reweighting is also reflected in the top 3 trends in Figure 24. The trend almost follows a pattern, with patient peaks at each phase pccuring in a nearly synchronous manner, rather than showing larger and more severe outliers as the simulation nears its end.

Looking at the metrics in Figures 25 and 26, it is noticeable that this time we had no patients in triage level 5, i.e., the least severe patient category, which makes the improved results in the previous figures even more obvious. Triage levels 1 and 2 took almost half as much time as the least severe patients. In addition, the first

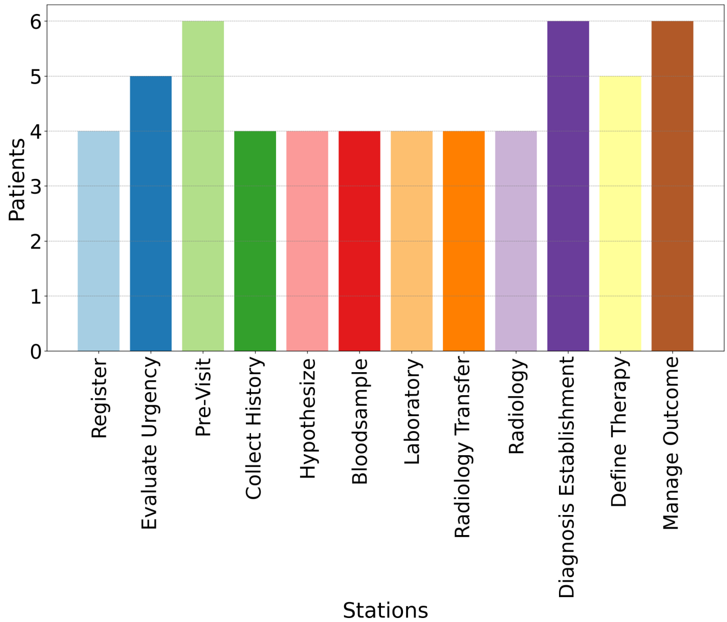

Figure 23: Maximum number of waiting patients in SC4

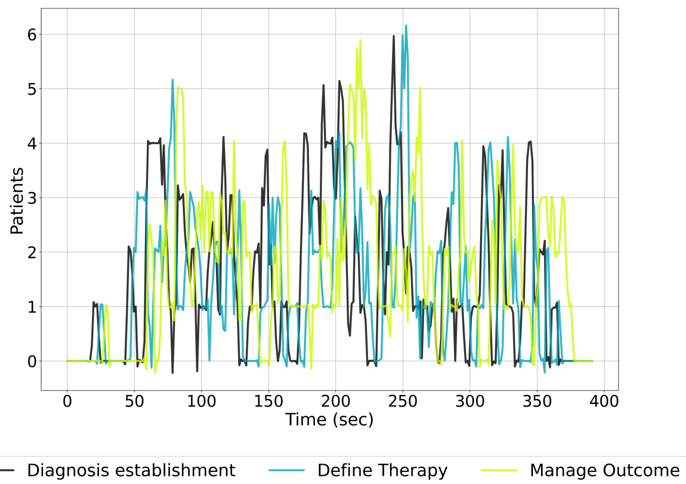

Figure 24: Patient number trends in SC4

two triage levels are similar in their time distribution, compared to levels 3 and 4, where several factors contributed to a longer duration relative to each other. Thus, while in Scenario3 the increase in staffing did not appear to reduce phase wait and turnaround times as much proportionately due to longer durations, here the increase proved to be significantly effective.

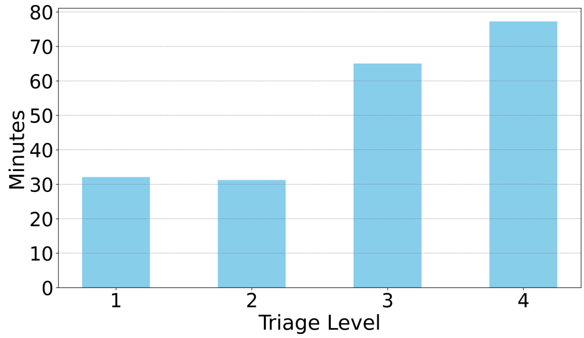

Figure 25: Average times spent in the simulation per triage level in SC4

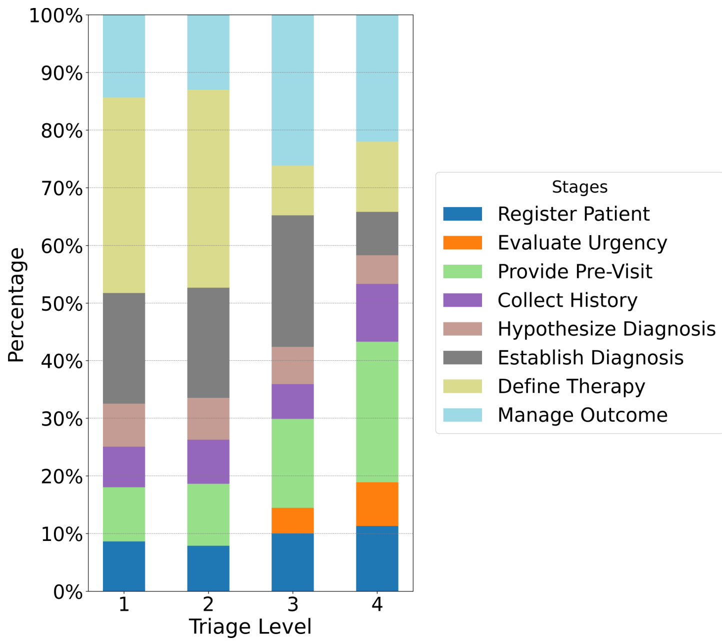

Figure 26: Distribution of times spent at the various stations per triage level in SC4

5.3 Discussion

In reviewing the simulation results, we found the following. First, the expectations for each scenario were either met or deviations occurred that can be interpreted based on the simulation run. Thus, despite the fact that the modelling toolbox itself has been simplified in line with our original objective, the model functions with the same accuracy. For example, with input from Hungarian and Italian hospital sources, the results of Scenario0 meet all specifications, from patient waiting times through triage level prioritisation to the maximum number of patients waiting at any one location. For the additional scenarios, the biases also yielded the expected results, so we can say that ***the reduced modelling toolkit, supported by scripts based on the SpiffWorkflow library, met our expectations and can be used for further simulations.***

The second most important observation from the results is that ***an increase in the number of staff in the emergency department does not necessarily mean an automatic acceleration of patient flow*** (and at this point, the simulation has not even been extended to include elements such as the limited space available for equipment and testing). However, ***there are some points where increased staffing does provide a boost, such as in patient registration and emergency assessment***, where increased staffing not only prevents these stations from becoming bottlenecks, but also helps prevent overcrowding in later stages of patient flow.

Examination of the trend plots for each scenario also shows that overcrowding does not start immediately, with the exception of SC3, where overcrowding was much more continuous due to the time gained from decontamination, with the peak typically occurring in the middle of the Emergency Department simulation, typically at points where throughput was already more critical. So, ***if the goal is to allocate resources more efficiently, it is certainly worthwhile to increase staffing during these periods and decrease it thereafter***.

Finally, in all scenarios, it has been shown that the most problematic phases in the patient flow are determining therapy and waiting for the visit to be evaluated. If the goal is to comply with COVID recommendations and reduce potential infection rates, these are the stages of patient flow where it is worth either reducing the time spent in the waiting room or, ***if this is not possible, providing patients with more separate, well-ventilated waiting rooms where they can wait for results without risking an extended stay that could reduce the effectiveness of infection prevention***.

Moreover, unlike many commercial solutions such as Visual Paradigm[1], Simcad Pro Health Simulation Software[2], or Simul8[3], our solution is a significant improve-

[1] Visual Paradigm — How to Create BPMN Diagram? URL: https://www.visual-paradigm.com/tutorials/how-to-create-bpmn-diagram/. [Accessed 24-Jan-2023]

[2] Simcad Simulation Software — Patient Flow Simulation: Predictive Modeling and Analytics, URL: https://www.createasoft.com/patient-flow-simulation. [Accessed 24-Jan-2023]

[3] Simul8 — Improve patient flow and enhance service quality. URL: https://www.simul8.com/applications/healthcare/improving-emergency-department-processes-with-simulation. [Accessed 24-Jan-2023]

ment in that it provides both free modeling and model execution, the source code of the modules used can be modified freely, as can the simulation's exact elements and output. In addition, the simplified modelling toolset and the BPMN file format do not restrict the usage of the created model, so if a research team has access to alternative simulation systems, the generated model may be utilized as-is or with minor modification. And its usage in its current form, maybe with minor enhancements, enables it to be utilized in conjunction with other simulation and assessment tools or by other processes. For instance, the data may be automatically merged with the hospital's measurements, which can then be run through Jimenez and Peng's tool [11] to create an accurate picture of the possibility of COVID spreading in a particular department or institution. Since the output is customizable, it may be used to study a broad variety of optimisation tasks, answering the demand mentioned by Yousefi et al [28]. The output and Python-based framework will presumably be of great value for reinforcement learning.

6 Conclusions

Our research is therefore currently at a stage where we have a modelling toolkit that can be readily used by researchers in other fields, as well as a simulation environment built from open-source components that can capture metrics and observations according to our needs, and which has already been able to provide valuable observations in its current form. There are, of course, a number of active directions in which we would like to develop this simulation solution further. The first and most obvious would of course be to obtain real Hungarian hospital data, with particular emphasis on data collected during the pandemic, to refine and improve the simulation, which is subject to ethical approval, and the application process is already underway. In addition, we plan to extend the simulation in its current form to include additional agents, such as constraints on equipment and rooms and more complex manipulation of patient arrival density using distribution and probability methods published in literature. We hope that the tools, method, and model resulting from our research will advance to the point where they become a valuable source of information for healthcare professionals to reduce the impact of cases such as COVID and improve hospital efficiency and patient care.

References

[1] About Modeler — Camunda Platform 8 Docs. URL: `https://docs.camunda.io/docs/components/modeler/about-modeler/`. [Accessed 15-Sep-2022].

[2] Arnaud, E., Elbattah, M., Ammirati, C., Dequen, G., and Ghazali, D. Use of artificial intelligence to manage patient flow in emergency department during the COVID-19 pandemic: A prospective, single-center study. *International Journal of Environmental Research and Public Health*, 19(15):9667, 2022. DOI: `10.3390/ijerph19159667`.

[3] Business Process Model And Notation Specification Version 2.0. URL: https://www.omg.org/spec/BPMN/2.0/. [Accessed 10-Sep-2022].

[4] Currie, C., Fowler, J., Kotiadis, K., Monks, T., Onggo, B., Robertson, D., and Tako, A. How simulation modelling can help reduce the impact of COVID-19. *Journal of Simulation*, 14(2):83–97, 2020. DOI: 10.1080/17477778.2020.1751570.

[5] Dieckmann, P., Torgeirsen, K., Qvindesland, S., Thomas, L., Bushell, V., and Langli Ersdal, H. The use of simulation to prepare and improve responses to infectious disease outbreaks like COVID-19: Practical tips and resources from Norway, Denmark, and the UK. *Advances in Simulation*, 5(1):1–10, 2020. DOI: 10.1186/s41077-020-00121-5.

[6] Dinh, M. and Berendsen Russell, S. Overcrowding kills: How COVID-19 could reshape emergency department patient flow in the new normal. *Emergency Medicine Australasia*, 33(1):175–177, 2021. DOI: 10.1111/1742-6723.13700.

[7] El-Bouri, R., Taylor, T., Youssef, A., Zhu, T., and Clifton, D. Machine learning in patient flow: A review. *Progress in Biomedical Engineering*, 3(2):022002, 2021. DOI: 10.1088/2516-1091/abddc5.

[8] He, L., Madathil, S., Oberoi, A., Servis, G., and Khasawneh, M. A systematic review of research design and modeling techniques in inpatient bed management. *Computers & Industrial Engineering*, 127:451–466, 2019. DOI: 10.1016/j.cie.2018.10.033.

[9] Hitzek, J., Fischer-Rosinskỳ, A., Möckel, M., Kuhlmann, S., and Slagman, A. Influence of weekday and seasonal trends on urgency and in-hospital mortality of emergency department patients. *Frontiers in Public Health*, 10, 2022. DOI: 10.3389/fpubh.2022.711235.

[10] Janbabai, G., Razavi, S., and Dabbagh, A. How to manage perioperative patient flow during COVID-19 pandemic: A narrative review. *Journal of Cellular & Molecular Anesthesia*, 5(1):47–56, 2020. DOI: 10.22037/jcma.v5i1.29789.

[11] Jimenez, J. and Peng, Z. Covid-19 airborne transmission tool available. URL: https://cires.colorado.edu/news/covid-19-airborne-transmission-tool-available, 2020.

[12] Kang, S. and Park, H. Emergency department visit volume variability. *Clinical and experimental emergency medicine*, 2(3):150, 2015. DOI: 10.15441/ceem.14.044.

[13] Karakusevic, S. Understanding patient flow in hospitals. URL: https://www.abhi.org.uk/media/1215/understanding_patient_flow_in_hospitals-nuffield-trust.pdf, 2016.

[14] Lee, S. and Lee, Y. Improving emergency department efficiency by patient scheduling using deep reinforcement learning. *Healthcare*, 8(2):77, 2020. DOI: 10.3390/healthcare8020077.

[15] Leva, D. and Sulis, E. A business process methodology to investigate organization management: A hospital case study. *WSEAS Transactions on Business and Economics*, 14:100–109, 2017. URL: https://www.wseas.org/multimedia/journals/economics/2017/a225807-059.php.

[16] Lin, Y.-K. and Chou, Y.-Y. A hybrid genetic algorithm for operating room scheduling. *Health Care Management Science*, 23(2):249–263, 2020. DOI: 10.1007/s10729-019-09481-5.

[17] Litvak, E. and Long, M. Cost and quality under managed care: Irreconcilable differences. *American Journal of Managed Care*, 6(3):305–12, 2000. URL: https://pubmed.ncbi.nlm.nih.gov/10977431/.

[18] NEJM Catalyst. What Is Patient Flow? URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.18.0289. DOI: 10.1056/CAT.18.0289, [Accessed 29-Sep-2022].

[19] Otsuki, H., Murakami, Y., Fujino, K., Matsumura, K., and Eguchi, Y. Analysis of seasonal differences in emergency department attendance in Shiga Prefecture, Japan between 2007 and 2010. *Acute Medicine & Surgery*, 3(2):74–80, 2016. DOI: 10.1002/ams2.140.

[20] Patient flow analysis: An overview of national and international application. URL: https://apps.who.int/iris/handle/10665/59190?locale-attribute=es&. [Accessed 29-Sep-2022].

[21] Salah, H. and Srinivas, S. Predict, then schedule: Prescriptive analytics approach for machine learning-enabled sequential clinical scheduling. *Computers & Industrial Engineering*, page 108270, 2022. DOI: 10.1016/j.cie.2022.108270.

[22] Schramm, P., Vaidyanathan, A., Radhakrishnan, L., Gates, A., Hartnett, K., and Breysse, P. Heat-related emergency department visits during the northwestern heat wave—United States, June 2021. *Morbidity and Mortality Weekly Report*, 70(29):1020–1021, 2021. DOI: 10.15585/mmwr.mm7029e1.

[23] SpiffWorkflow 1.1.6 documentation. URL: https://spiffworkflow.readthedocs.io/en/latest/, 2014. [Accessed 22-Sep-2022].

[24] Tavakoli, M., Tavakkoli-Moghaddam, R., Mesbahi, R., Ghanavati-Nejad, M., and Tajally, A. Simulation of the COVID-19 patient flow and investigation of the future patient arrival using a time-series prediction model: A real-case study. *Medical & Biological Engineering & Computing*, 60(4):969–990, 2022. DOI: 10.1007/s11517-022-02525-z.

[25] Terning, G., Brun, E., and El-Thalji, I. Modeling patient flow in an emergency department under COVID-19 pandemic conditions: A hybrid modeling approach. *Healthcare*, 10(5):840, 2022. DOI: 10.3390/healthcare10050840.

[26] Varga, C., Lelovics, Z., Soós, V., and Oláh, T. Betegforgalmi trendek multidiszciplináris sürgősségi osztályon. *Orvosi Hetilap*, 158(21):811–822, 2017. DOI: 10.1556/650.2017.30749.

[27] Won, Y., Ho Hwang, T., Roh, E., and Chung, E. Seasonal patterns of asthma in children and adolescents presenting at emergency departments in Korea. *Allergy, asthma & immunology research*, 8(3):223–229, 2016. DOI: 10.4168/aair.2016.8.3.223.

[28] Yousefi, M., Yousefi, M., and Fogliatto, F. Simulation-based optimization methods applied in hospital emergency departments: A systematic review. *Simulation*, 96(10):791–806, 2020. DOI: 10.1177/0037549720944483.