8-2022

Improving Patient Safety, Patient Flow and Physician Well-Being in Emergency Departments

Vishnunarayan Girishan Prabhu
Clemson University, vgirish@clemson.edu

IMPROVING PATIENT SAFETY, PATIENT FLOW, AND PHYSICIAN WELL-BEING IN EMERGENCY DEPARTMENTS

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Industrial Engineering

---

by
Vishnunarayan Girishan Prabhu
August 2022

Accepted by:
Dr. Kevin M Taaffe, Committee Chair
Dr. Ronald G Pirrallo
Dr. Sudeep Hegde
Dr. Tugce Isik

Abstract

Over 151 million people visit US Emergency Departments (EDs) annually. The diverse nature and overwhelming volume of patient visits make the ED one of the most complicated settings in healthcare to study. ED overcrowding is a recognized worldwide public health problem, and its negative impacts include patient safety concerns, increased patient length of stay, medical errors, patients left without being seen, ambulance diversions, and increased health system expenditure. Additionally, ED crowding has been identified as a leading contributor to patient morbidity and mortality. Furthermore, this chaotic working environment affects the well-being of all ED staff through increased frustration, workload, stress, and higher rates of burnout which has a direct impact on patient safety.

This research takes a step-by-step approach to address these issues by first forecasting the daily and hourly patient arrivals, including their Emergency Severity Index (ESI) levels, to an ED utilizing time series forecasting models and machine learning models. Next, we developed an agent-based discrete event simulation model where both patients and physicians are modeled as unique agents for capturing activities representative of ED. Using this model, we develop various physician shift schedules, including restriction policies and overlapping policies, to improve patient safety and patient flow in the ED. Using the number of handoffs as the patient safety metric, which represents the number of patients transferred from one physician to another, patient time in the ED, and throughput for patient flow, we compare the new policies to the current practices. Additionally, using this model, we also compare the current patient assignment algorithm used by the partner ED to a novel approach where physicians determine patient assignment considering their workload, time remaining in their shift, etc.

Further, to identify the optimal physician staffing required for the ED for any given hour of the day, we develop a Mixed Integer Linear Programming (MILP) model where the objective is to minimize the combined cost of physician staffing in the ED, patient waiting time, and handoffs. To develop operations schedules, we surveyed over 70 ED physicians and incorporated their feedback into the MILP model. After developing multiple weekly schedules, these were tested in the validated simulation model to evaluate their efficacy in improving patient safety and patient flow while accounting for the ED staffing budget.

Finally, in the last phase, to comprehend the stress and burnout among attending and resident physicians working in the ED for the shift, we collected over 100 hours of physiological responses from 12 ED physicians along with subjective metrics on stress and burnout during ED shifts. We compared the physiological signals and subjective metrics to comprehend the difference between attending and resident physicians. Further, we developed machine learning models to detect the early onset of stress to assist physicians in decision-making.

Dedication

This dissertation is dedicated to my loving family: Girish Prabhu, Sandhya Kamath, and Shyamnarayan Girishan Prabhu, for believing, supporting, and motivating me during tough times.

Acknowledgements

I am deeply indebted and grateful to my advisor Dr. Kevin Taaffe for the extensive support and guidance provided throughout my graduate program, without whom this work would not have been possible. I had the luxury of spending time in the Emergency Department at Greenville Memorial Hospital, which provided invaluable insights and led to developing practical solutions. I want to express my deepest appreciation to Dr. Kevin Taaffe, Dr. Ronald Pirrallo, Dr. William Jackson, Dr. Mike Ramsay, and Dr. Jess Hobbs for providing this opportunity, their guidance, and a wonderful experience at the Emergency Department. Next, I would like to thank my committee members, Dr. Tugce Isik and Dr. Sudeep Hegde, for taking time off their schedules and providing constructive feedback on improving this dissertation.

I am extremely grateful to Dr. Laura Stanley, Dr. Robert Morgan, and Dr. Dotan Shvorin, who sparked my initial interest in healthcare research and provided various research opportunities at PRISMA Health, which made a meaningful impact on patient's life. I would also like to thank Dr. WIlliam Hand, Dr. Tugce Isik, and Dr. Yongjia Song for involving me in research projects beyond the dissertation work and guiding me throughout those projects. I would also like to thank all the other faculties at the Dept. of Industrial Engineering for their support and guidance and for making my graduate program an unforgetful experience.

Last but not least, I would like to thank my friends and fellow students Marisa Shehan, Nitin Srinath, Apostolos Kalatzis, Srinivasan Nagarajan, Chakara Rajan, Devesh Kumar, and Clemson Creative Inquiry students for their support and valuable insights in improving this dissertation.

Table of Contents

List of Figures

List of Tables

1. Chapter One

1.1 Introduction

The Emergency Department (ED) is a critical segment in the US health system as it acts as an essential patient entry point that contributes to approximately 70% of hospital admissions [1]. As society's health care safety net, patients with no other options for medical care access the ED because the federal government mandates an ED to provide screening and stabilizing care to all patients regardless of their ability to pay [2]. The total number of patients visiting EDs is increasing annually, and according to the latest report, over 151 million people visit US EDs annually [3]. The sheer volume and diverse nature of patient visits make the ED predisposed to crowding. The American College of Emergency Physicians defines crowding as a situation in which the identified need for emergency services exceeds available resources for patient care in the emergency department, hospital, or both [4]. ED crowding is a patient safety issue as well as a public health problem. Crowding is a result of multiple factors, including high patient volumes, inadequate staffing, and bed shortages resulting in a longer patient length of stay and slower discharge rates [5–7]. Additionally, the reduced availability of ED beds due to admitted patients awaiting transfer into an in-hospital bed restricts an ED's capacity to accept new patients, resulting in higher patient boarding time and delays in providing patient care [8]. Thus, crowding disrupts the ED patient care processes and negatively impacts patient safety, ED staff wellbeing, and health system costs [5,9–11].

The impact of ED crowding and patient flow has been studied for decades, and the researchers have used a variety of operations research approaches and other methodologies to address the issue; however, the ED overcrowding crisis is still prevalent [5,10]. Hence, it is essential to conduct a detailed literature review to identify less explored research areas and well-researched topics to apply new methodologies and techniques. To include these areas of interest that could potentially impact this proposed study, we divide our brief literature review and organize the chapters based on three broad topics. First, we start with 1) Forecasting Patient Arrivals to the ED, followed by 2) Patient-Physician Assignment and Patient Flow in the ED, and finally, 3) Physician Stress and Well-being in the ED.

Day-to-day ED operations involve numerous medical and nonmedical decisions, starting from patient arrivals to the ED until final disposition from the ED. Forecasting the patient arrivals to an ED is critical as there is an apparent daily, weekly, and monthly pattern of patient arrivals to the ED that affects various operational decisions. Specifically, having these forecasts can assist ED administrators in making informed decisions for daily operations. Apart from the daily decision-making, to ensure efficient functioning of the ED, administrators make a lot of decisions as early as three months in advance, including decisions on the number of beds required, pod's functioning, physicians required, and shift schedule. Hence it is critical to have a model that predicts patient arrivals to schedule and maintain an adequate number of physicians and beds.

Often, the ED administrators have to make real-time adjustments and add surge capacity beds when the ED is overcrowded and close pods in case of a low number of patients. Both have consequences, but overcrowding and lack of beds could lead to patient mortality, increased wait times, and higher chances of medical error. Given the significance of the topic, researchers have used various time series forecasting models and a few machine learning models to forecast patient arrivals to an ED. However, most studies have focused on long-term predictions such as weekly or daily arrival volumes, with very few studies focusing on hourly arrival predictions. Moreover, to our knowledge, none of the studies have focused on predicting the hourly patient arrivals with their ESI levels. Adding the Emergency Severity Index (ESI) level information to the forecasting models could further help larger EDs with multiple pods in bed planning and staff scheduling.

Following the patient's arrival, the next patient activity in the ED is triaging, where a patient is assigned with the standardized Emergency Severity Index (ESI), which defines the patient's severity level [12]. Once a patient is assigned an ESI level, they are assigned to a pod in the ED based on the severity, where ESI 1 represents an urgent patient who needs immediate care, whereas ESI 5 represents a non-urgent case that can wait in the absence of bed. Pods in ED represent a section with specific bed capacity, physician availability, and equipment availability. Therefore, some pods are limited to providing care for only low acuity patients (ESI 3, 4, and 5) because of the lack of expert physicians and equipment in the pod. Hence patient assignment to the pod is crucial to maintain the smooth operation of the ED and avoid overcrowding. After the patient assignment to a pod, the physician provides care for the patient on their bed multiple times based on the patient's ESI levels and orders various ancillary tests (Labs, radiology, consults, etc.) as necessary. Finally, when ready, the physician makes a disposition decision to either admit or discharge the patient. Thus, it is crucial to study each interaction from patient arrival to the ED until disposition as these have a significant impact on patient flow, patient safety, patient satisfaction, and the time each patient spends in the ED. Although a variety of approaches have been adopted to improve ED patient throughput and efficiencies (e.g., using hallways as additional bed resources, scheduling additional physicians, vertical care in triage, calling off-shift nurses, etc.), these fixes appear to be only short-term solutions for this crisis. Hence further research in this direction is warranted to improve ED performance. Finally, from a physician’s perspective, frequent exposure to overcrowded ED results in increased workload and stress and could eventually lead to physician burnout. Prior studies have identified ED physicians as one of the top specialties among physicians who experience burnout.

A variety of mathematical modeling and simulation modeling approaches have been used by researchers to address the topics of patient triaging, patient assignment, bed planning, and staffing and scheduling to improve the patient flow within the ED [13–20]. However, very few studies have combined these individual topics, and to our knowledge, none of the studies utilizing modeling techniques have focused on incorporating the idea of patient flow, patient safety, and physician workload management. Similarly, researchers have used various techniques, including physiological measures, to understand physician stress; however, none of these have focused on detecting the early onset of physician stress in the ED for decision making. Although taking a microscopic view and focusing on one of these topics provides a sound estimate for specific performance measures within the ED, it fails to capture all the effects and interactions of the ED, leading to a limited application of the model for decision making. To address this issue and provide our partner ED with the right tools to assist in decision-making, we propose a step-by-step approach utilizing novel modeling techniques that captures all these crucial components of ED.

The first chapter focuses on forecasting patient arrivals to the ED using time series forecasting models and machine learning models. We developed six different models to generate long-term (90 days ahead) daily and short-term (one week ahead) hourly forecasts to predict the total patient census to the ED and their ESI levels. In the second chapter, we develop and validate an agent-based hybrid discrete event simulation model representative of PRISMA Health Greenville Memorial Hospital ED for testing various physician shift scheduling policies to improve patient safety and patient flow. Along with testing various overlapping policies and restriction policies for physician staffing, we also investigate policies where patient-to-physician assignments are based on physician workload and time remaining on their shift. Next, in the third chapter, we develop a Mixed Integer Linear Program (MILP) model to identify optimal physician staffing levels to reduce the combined cost of staffing the physicians, patient wait time, and handoff costs. To develop practical policies and incorporate physician preferences, we surveyed over 70 physicians in the partner ED. Five different staffing schedules were developed and tested in the validated simulation model to understand the efficacy of each physician staffing schedule in improving patient safety and patient flow while accounting for the ED staffing budget. Finally, in the fourth chapter, we investigate physician stress and well-being in the ED as they are critical factors leading to burnout. Since our partner is an academic ED, we collected 100+ hours of objective physiological data and subjective feedback from attending and resident physicians working an entire 8-hour ED shift and compared the stress and burnout levels among the two populations. Further, we develop a recurrent neural network-based machine learning model using the long-short term memory (LSTM) approach for early detection of physician stress using physiological measures to reduce the likelihood of burnout.

2. Chapter Two

2.1 Forecasting Patient Arrivals to Emergency Departments

2.1.1 Introduction

Emergency Department is one of the primary patient entry points into a hospital and acts as the frontline for delivering emergency services. Patient arrivals to EDs in the US have increased from 96.5 million annual visits in 1995 to 115.3 million in 2005 and 151 million in 2019 [3,21]. At the same time, the number of EDs in the US has decreased by over 15% in the last decade [22]. The ever-increasing patient volume and the decreasing number of EDs lead to mismatch predisposing EDs to crowding [8,23,24]. The American College of Emergency Physicians (ACEP) defines ED crowding as the situation that "occurs when the identified need for emergency services exceeds available resources for patient care in the ED, hospital, or both" [4]. Crowding in ED is a global concern, and studies have often linked this as a factor leading to suboptimal patient care, delays in care, and higher chances of medical errors [23,25]. A few leading causes of ED overcrowding include high patient census (patient arrivals), inadequate resources (beds, medical devices, etc.), inadequate planning, and poor ED design [10,26]. Some of the most commonly adopted solutions to avoid ED crowding include expanding ED capacity, stopping boarding admitted patients in ED, hallway beds, on-call providers, and adding temporary resources [27]. A recent study investigating ED crowding identified that access to future patient demands (arrivals to ED) during the time of shift scheduling and resource allocation can improve ED planning and potentially avoid the chances of ED crowding [24]. Although patient arrivals to the ED are affected by factors beyond the ED clinicians' and administrators' control, prior studies have found consistent hour-of-the-day and day-of-the-week patterns in patient arrivals, allowing for developing robust time series forecasting models [28].

Time series forecasting focuses on developing mathematical models to predict future values based on previously observed data [29]. Specifically, it focuses on understanding the patterns associated with a series of data points of the variable of interest over time to make predictions. The ability of the time series model to forecast future values has led to its adoption in various research areas, including healthcare, finance, banking, weather, traffic flow, energy, and manufacturing [30–34]. In healthcare, researchers have implemented various time series forecasting techniques to forecast surgical case volume, disease progression, stress detection, risk of disease over time, identify early onset of diseases, mortality, disease management, inpatient admissions, patient arrivals (census) to the ED, etc. [35–39]. In terms of the methodology used in developing forecasting models for healthcare applications, studies have used various models, from the simple persistence model to complex deep-learning models [35–39]. In most studies, persistence models are used as the baseline model as they can account for only the autoregressive term of the time series. Compared to persistence models, Autoregressive integrated moving average (ARIMA) is a slightly advanced model which can account for both autoregressive and moving average components. Additionally, seasonal autoregressive integrated moving average (SARIMA), an advanced version of ARIMA, is used to handle time-series data with seasonality. Although these models are effective and widely used, one of the primary limitations of these models when applying to healthcare datasets is the underlying linearity assumption which makes these models undesirable as most healthcare data sets are non-linear time series data [40].

Over the last several years, various non-linear forecasting methods, artificial neural networks (ANN), support vector regression (SVR), and fuzzy models have been implemented for forecasting time series data [41–45]. Since these models do not assume linearity for the time series data, they perform well in forecasting both linear and non-linear time series data. More recently, researchers have explored using memory-based recurrent neural networks (RNNs) and convolutional neural networks (CNNs) for various healthcare applications [36,46,47]. The capability of memory-based RNNs to account for the temporal nature of the data, work around the vanishing gradient problem and store important information in their memory cell allows these models to generate a long-term forecast with high accuracy. Additionally, a few studies have used the ensembling approach, where predictions from various modeling approaches are combined by assigning weights to generate forecast [48,49]. Although the standard approach of ensembling is to assign equal weights to each forecasting model, researchers have developed advanced approaches, including least-square regression (LSR) average of in-sample weights (AIW) that aims to generate optimal weights by minimizing the errors. Although ANNs, SVRs, and ensembling approaches usually tend to perform better than the traditional time series forecasting models, researchers should be cautious while using these advanced methods to avoid overfitting [50].

2.1.2 Background and Literature

Forecasting patient arrivals to the ED has been an active area of research across the world over the last two decades, given the public health and patient safety issue of ED crowding. Researchers have used a variety of forecasting methods to predict patient arrivals to the ED for different horizons [51–63]. Additionally, a few studies have focused on forecasting specific types of patient arrivals to the ED (primarily patients with respiratory diseases) [64,65]. In terms of the methodology used for forecasting patient arrivals to ED, ARMA, VARMA, Holt-Winters, linear regression, multiple linear regression (MLR), ARIMA, SARIMAX, ANNs, and RNNs have been used extensively. All the above-cited studies investigating patient arrivals to the ED except two have used at least two different forecasting methods to evaluate the predictions generated and choose the best performing model. In terms of forecasting horizon, researchers have forecasted hourly, daily, and monthly patient arrivals. Surprisingly, none of the studies in the US have aimed to include ESI levels of the patients while generating the forecasts. Table 2.1 below provides a snapshot of prior studies with the forecasting horizon, forecasting methods, and inclusion of ESI/similar severity index in forecasts.

A majority of prior studies have focused on forecasting the daily patient arrivals to the ED for various time horizons varying from 1-month forecast to 6 months. Among these studies, most have reported ARIMA or SARIMA models as the best performing models except for two recent studies, which reported neural networks (ANN and LSTMs) as the best performing models [53,58]. However, it should be noted that one study among these two has not compared the machine learning models to the traditional time-series forecasting models, including ARIMA and SARIMA, and the other one used additional input variables beyond the time of patient arrival to the ED, which could have led to a non-linear relationship. Irrespective of the specific underlying reason, advanced machine learning models usually tend to outperform the traditional time series models when using multiple input variables [66,67]. Although using multiple input variables can improve patient forecasts, some of these inputs are difficult to identify without proper feature selection approaches and require expert manipulation, making them difficult for use by ED administrators and stakeholders. Hence our primary focus was to use only two features a) patient arrival time and b) patient ESI levels. These two factors were selected because a) they can be readily extracted from Electronic Health Record (EHR) database and b) resource requirements vary by ESI, thus allowing for better planning.

Table 2.1: List of prior studies forecasting patient arrivals to the ED.

Study	Forecasting Horizon	Forecasting Methods	Inclusion of Severity Index	Best Model
Sarfo et al.,	Monthly for 24 months	VARMA, ARMA, Holt-Winters	Australian Triage Scale	VARMA
Khaldi	Weekly for 2 Months	ANN, ARIMA	--	ANN
Silva et al.,	Weekly, Daily for 1 month	ARIMA, Holt-Winters	--	ARIMA
Whitt	Daily for 3 months	SARIMA, SARIMAX, MLP	--	SARIMAX
Batal et al.,	Daily 3 months	Regression	--	Single Approach
Jones et al.,	Daily for 1 month	Regression, SARIMA, Exponential Smoothing, ANN	--	Regression with calendar variables
Sun et al.,	Daily for 6 months	Regression, ARIMA	P1,P2, P3	ARIMA
Xu et al.,	Daily for 1 month	ANN, MLR, Regression	Only 3 & 4	ANN
Kadri et al.,	Daily for 1 Month	ARMA, ARIMA	--	ARIMA
Zhang et al.,	Daily, Hourly for 3 Months	ARIMA, KNN, SVR, Ridge, XGBoost, Random Forest, AdaBoost, LSTM	--	LSTM
Choudhury et al.,	Daily Hourly for 1 month	ARIMA, Holt-Winters, NN, Regression	--	ARIMA
Hertzum	Hourly for 1 Month	Regression, ARIMA	--	ARIMA
Cote et al.,	Hourly, Daily, Monthly, Yearly for Yearly	Regression	--	Single Approach

Among four studies that focused on forecasting hourly patient arrivals to the ED, two studies identified ARIMA models to be the best performing model, while another reported LSTMs (a type of RNN) to be better than ARIMA, and the final study utilized only a single approach (regression model). For the study using a single approach, the adjusted R-square for the models for hourly and daily predictions were 46.8 and 32.8, and forecasted values varied significantly from the actual values [57]. This can be attributed to the lack of capability of simple linear regression models to account for the seasonality of patient arrivals to the ED, as reported by other studies [60,61]. For the study that reported LSTMs to outperform ARIMA, the potential reason that LSTMs outperform ARIMA could be because of two key reasons a) the use of multiple input variables and b) the capability of LSTMs to store important temporal behavior in the memory cell. Multiple prior studies have identified LSTMs to perform well for short-horizon forecasts where data varies quickly between the time frames [68–70].

For the two studies that identified ARIMA as the best performing model for the hourly forecast, one reported ARIMA outperformed regression models, whereas the other study reported ARIMA outperformed Neural Nets, Regression, and Holt-Winters forecasting models. In the first study, although the ARIMA model outperformed the regression model, the hourly prediction varied significantly where the reported mean percentage error varied between 49-58% for the ARIMA model. Although it is expected that mean percent errors for smaller forecast horizons would be higher, the reported results limit the ability of the forecasting model to inform sound decision-making. The second study was able to develop a robust ARIMA model where the reported mean error (ME) and a root mean square error (RMSE) were 1.01 and 1.55. Although these observations are promising, the model was generating forecasts daily. This could be useful for immediate fixes such as adding hallway beds etc., in real-time but most EDs generate staffing schedules and resource allocation plans in advance (2-3 months ahead), thus failing to inform long-term planning. Moreover, as briefly mentioned earlier, none of these studies that forecasted hourly patient arrivals have considered patient severity (ESI) in their forecasts which is critical for resource allocation as each severity index requires different types of resources [12].

To address these gaps, our research proposes a two-step modeling approach that utilizes patient arrival times and their severity index to forecast daily long-term (3 months ahead) and hourly short-term (weekly) forecasts. Three months duration for the long-term was decided based on the feedback from our partner ED which generates their shift schedules and resource allocation plans for 3 months ahead. Both the forecasts will provide the total expected patient arrivals along with their ESI levels to assist in resource allocations, including staffing schedules. To achieve this goal using the data from our partner hospital, we developed various traditional time series models along with machine learning models.

2.1.3 Methods

2.1.3.1 Data

Input data for the model, including the number of patient arrivals to the ED and their ESI levels, were gathered from the PRISMA Health Greenville Memorial Hospital (GMH), Greenville, SC. PRISMA Health is the largest healthcare provider in South Carolina and serves as a tertiary referral center for the entire Upstate region, and the flagship GMH academic Department of Emergency Medicine is an Adult Level 1 Trauma Center. Patient arrival data from January 2017 - December 2020, totaling 309,430 visits, were retrospectively accessed from the hospital's EHR database.

We first introduce Figure 2.1 below, which represents the hourly average patient arrivals per day to the GMH ED over the four years. The first thing to notice here is the consistent hourly pattern across the four years. It can be observed that the patient arrivals are low during the early hours (12:00-7:00 am) and slowly start picking up from 7:00 am until 11:00 am -12:00 pm when they reach the maximum and stay high until 6:00 pm. This patient arrival trend is consistent with a lot of other EDs, and prior studies have reported the same [28,71]. Specifically, in this dataset, it can be observed that approximately 60% of the patient arrivals occur during the 8-hour time window between 9:00 am - 5:00 pm. Although the physicians and ED administrators are aware of this general trend of patient arrivals to the ED, it's crucial to have a robust prediction model to maintain adequate staffing, beds, and other resources at a given point in time to ensure the smooth function of the ED. This is where forecasting patient severity (ESI) adds value, as patient arrivals and ESI patterns are not the same throughout the day. Forecasting of ESI thus provides more insights to providers and administrators regarding what resources (physicians vs. physician assistants, hallway bed vs. high acuity bed, etc.). should be allocated to which pod, etc.

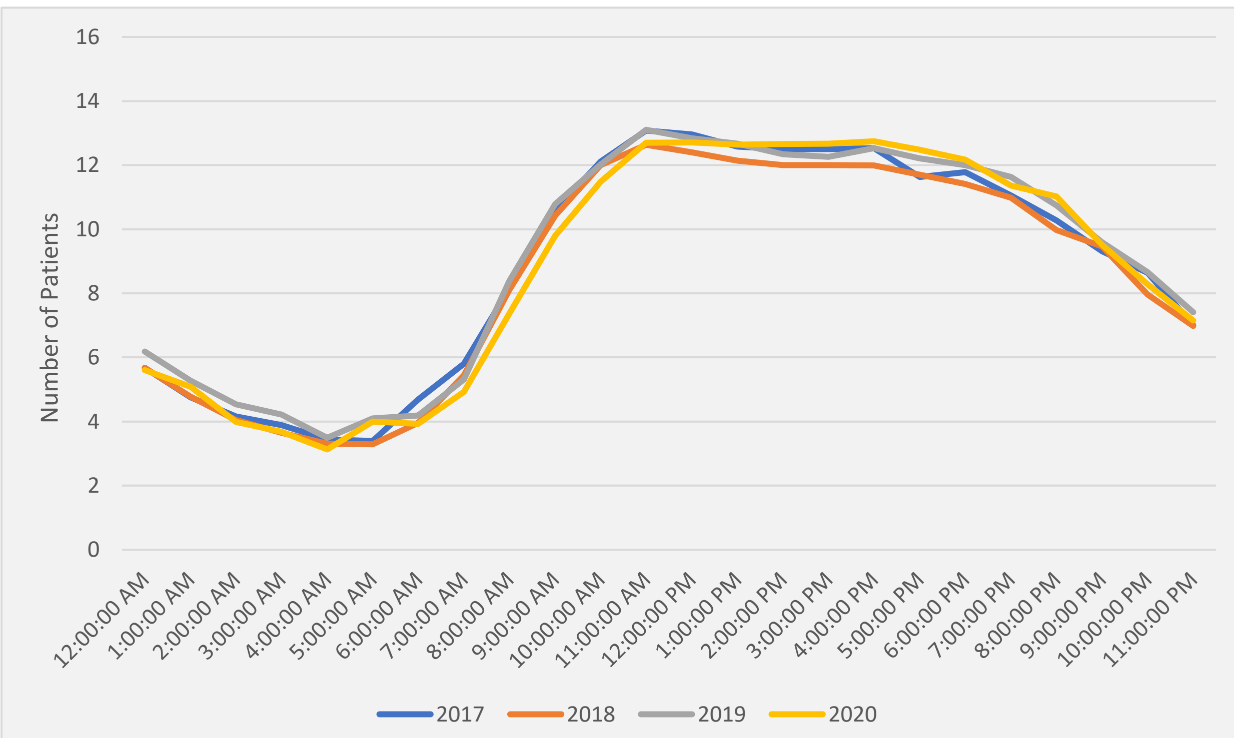


Figure 2.1: Hourly patient arrival per day to GMH ED.

Next, we introduce Table 2.2 below, which represents the percentage contribution of each ESI level on the patient arrivals for the four years. Similar to the hourly pattern, the first thing to notice here is the constant ESI pattern over the four years. It can be observed that about 50% of patient arrivals to the ED are ESI-3 patients, followed by ESI 2 and ESI 4, which contributed 25% and 18% of the patient arrivals. Finally, ESI 1 and 5 each contributed only 2-3% of the total arrivals. ESI 1 refers to severely unstable patients who need immediate intervention, and ESI 5 patients are the most stable patients and may be treated non-urgently and mostly require the least resources. Although patient arrivals varied over the four years where the least arrivals were observed during 2020 because of the COVID-19 pandemic, as mentioned earlier, the percentage of each ESI level contributing to the total patient arrivals has stayed the same over the years. It should be noted that for each year, at most 2% of patient arrivals were recorded without an ESI level in the EHR, and we interpolated these values using the rest of the data as ESI levels were important for our forecasting models.

Table 2.2: Percentage contribution of each ESI level on the patient arrivals to the ED.

Year	ESI 1	ESI 2	ESI 3	ESI 4	ESI 5
2017	3%	25%	51%	18%	2%
2018	3%	26%	51%	18%	2%
2019	4%	25%	50%	18%	2%
2020	4%	24%	53%	17%	3%

Next, we introduce Table 2.3 and Figure 2.2 to provide insights into the seasonality associated with the day of the week. Table 2.3 represents the average patient arrivals for each day of the week for each year, and Figure 2.2 represents the average patient arrivals for each day of the week over the four years. It can be observed that the patient arrivals are the highest on Mondays and least on the weekends. For weekdays other than Monday, the average patient arrival to the ED doesn’t vary drastically and stays within a specific range. Most EDs across the world have reported a similar weekly pattern [72].

Table 2.3: Average patient arrivals for each day of the week.

Year	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
2017	204	233	222	223	216	219	204
2018	198	226	218	212	206	212	195
2019	209	237	228	221	218	220	204
2020	196	233	221	218	215	215	203

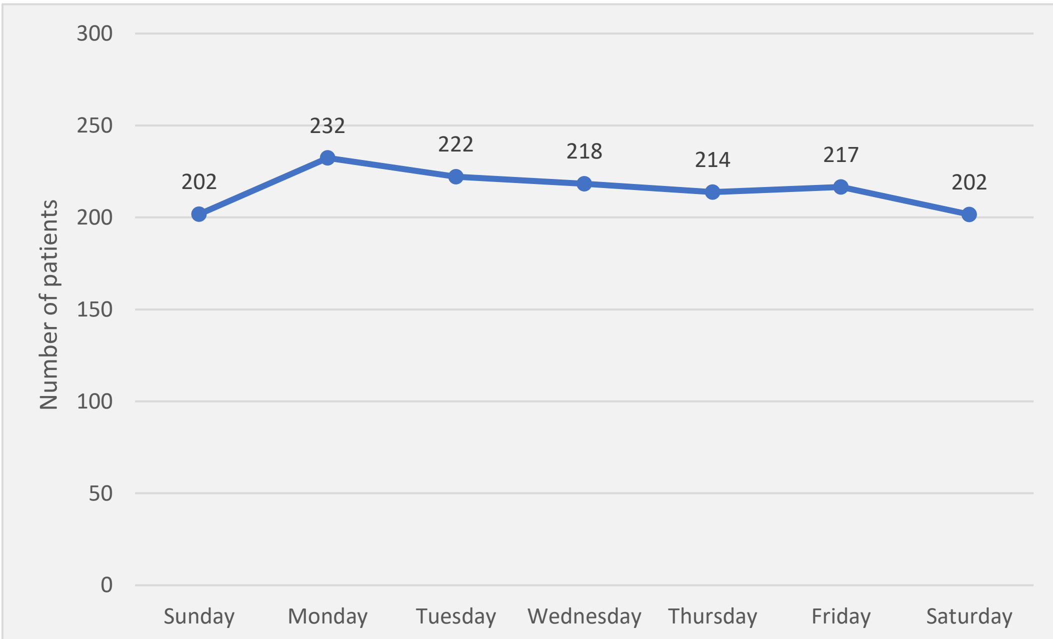


Figure 2.2: Average daily patient arrival rate to GMH ED (2017-2020).

Finally, we introduce Figure 2.3, which represents the total monthly patient arrivals to the GMH ED for each year. The most important thing to notice here is the significant drop in patient arrivals in April and May 2020 because of the COVID-19 pandemic. Although these are outlier months, we did not exclude or extrapolate the data to capture the natural variability and be representative of the ED. Apart from these months, the arrival pattern did not vary significantly (p-value > 0.09).

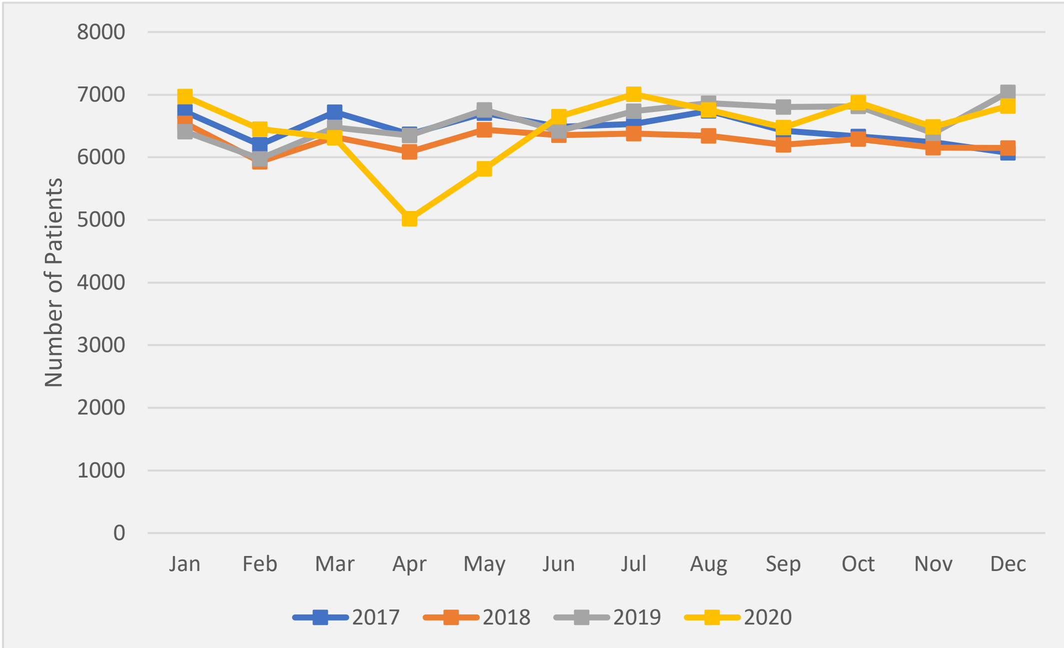


Figure 2.3: Monthly patient arrivals to GMH ED.

For training the models and testing the potential of various models to forecast patient arrivals, we split the data into training and test set. We used the last 90 days of 2020 as the test set, and the rest of the data was used for training. For daily data, each row in the data set corresponds to a specific date, and the dataset had six columns ESI 1, ESI 2, ESI 3, ESI 4, ESI 5, and the Total. For the hourly data, each row represented a specific hour of the day and had the same six columns. Figure 2.4 below represents the first five rows of hourly (left) and daily (right) datasets.

Hours	ESI 1	ESI 2	ESI 3	ESI 4	ESI 5	Total
2017-01-01 00:00:00	0	0	4	1	0	5
2017-01-01 01:00:00	2	0	6	1	0	9
2017-01-01 02:00:00	0	1	2	1	0	4
2017-01-01 03:00:00	0	3	1	0	1	5
2017-01-01 04:00:00	0	0	1	0	0	1

Date	ESI1	ESI2	ESI3	ESI4	ESI5	Total
2017-01-01	6	38	116	47	2	209
2017-02-01	5	31	112	66	2	216
2017-03-01	5	43	140	52	4	244
2017-04-01	5	39	163	42	3	252
2017-05-01	2	43	138	47	1	231

Figure 2.4: Sample hourly and daily file inputted into the model.

2.1.3.2 Model Development and Evaluation

From the data discussed above, it is evident that naïve forecasting models will not be useful in forecasting the patient arrivals to the ED in this case, given the seasonality. However, we used the moving average model as our benchmark model to compare the forecasts from other models. Multiple studies have reported ARIMA and SARIMA models to be effective in forecasting time series data, especially SARIMA when the data is considered to have seasonality [56,59–63]. Based on these observations, as well as our high-level data analysis, we decided to develop both ARIMA and SARIMA models in this study.

2.1.3.2.1 ARIMA and SARIMA Forecasting Model

The ARIMA model has three parameters (*p, d, q*) which should be tuned according to the characteristics of input data to develop a robust ARIMA model. Here *p* represents the order of the autoregressive components, which are the lags or the previous values that should be considered to predict the next value, *d* represents the number of differentiation required to make the data stationary if the initial data is non-stationary, and *q* represents the order of moving average component which is the number of past error terms that should be used for prediction. There are various validated methods to estimate the best *p, d, q* values, and in this research, we followed the approach suggested by Box and Jenkins [73]. First, to check if the data can be considered to be stationary, we performed a Dickey-Fuller test [74]. For the Dickey-Fuller test, if the p-value was less than the alpha level (0.05), we reject the null hypothesis (H_0 = This time series has a unit root and is not stationary). If non-stationary, the data was differentiated until stationarity was achieved, and the number of differentiations will be used for *d.* To determine the AR and MA components, *p* and *q*, the auto-correlation function and partial auto-correlation function were plotted. Based on if the partial correlation function and autocorrelation function vanish (stay within the confidence interval), we choose the p and q. If these functions do not vanish (go beyond the confidence interval), we used those patterns to set the upper bound and lower bound of p and q. After selecting the parameters, the best combinations were selected based on the Akaike information criterion (AIC) values. After which the goodness of fit of the model was evaluated using the Jarque-Bera test. Additionally, we plotted standardized residuals over time and Q-Q plot to evaluate the residuals were normal. After evaluating all these criteria, the model was used to forecast the patient arrivals for the forecast horizon.

Next, to account for the various seasonality components involved in the ED patient arrival data, we developed the Seasonal ARIMA (SARIMA) model for predicting patient arrivals. The SARIMA model is an extension of the ARIMA model that adds new parameters to account for the seasonal element in data. Similar to ARIMA, the SARIMA model can be represented using its parameters: $(p,d,q)(P, D, Q)_{m,}$ where p and P represent the order of autocorrelation at the nonseasonal and seasonal levels, respectively, d and D represent the degree of nonseasonal/seasonal differencing, and q and Q represent the order of the moving average process at the non-seasonal, and seasonal levels and the m represents the seasonality of the data. For developing and evaluating the SARIMA model, we followed the same approach as for ARIMA, and the m was set as 7 because our data was for daily prediction. In the case of hourly prediction, m was set as 24.

In this section, we will discuss how we set the parameters for our ARIMA and SARIMA models used for long-term daily prediction. As mentioned above, the first step was to check if the data was stationary, and for this, we performed the Dickey-Fuller test and observed that the p-value < 0.001. Hence, we rejected the null hypothesis and determined that the data was stationary. Hence our value for d=0. Next, to determine p and q, we plot the partial autocorrelation function and auto-correlation function. Figure 2.5 and Figure 2.6 below represent the autocorrelation and partial autocorrelation plot, and we see that both the partial correlation function and the autocorrelation function do not vanish, so we try to model it as an ARMA sequence. Here we select both p and m to vary from 1 to 6, and these numbers were based on the seasonality of the data. Based on the AIC values, ARIMA (3,0,1) is the best model, whose AIC was 11221.67.

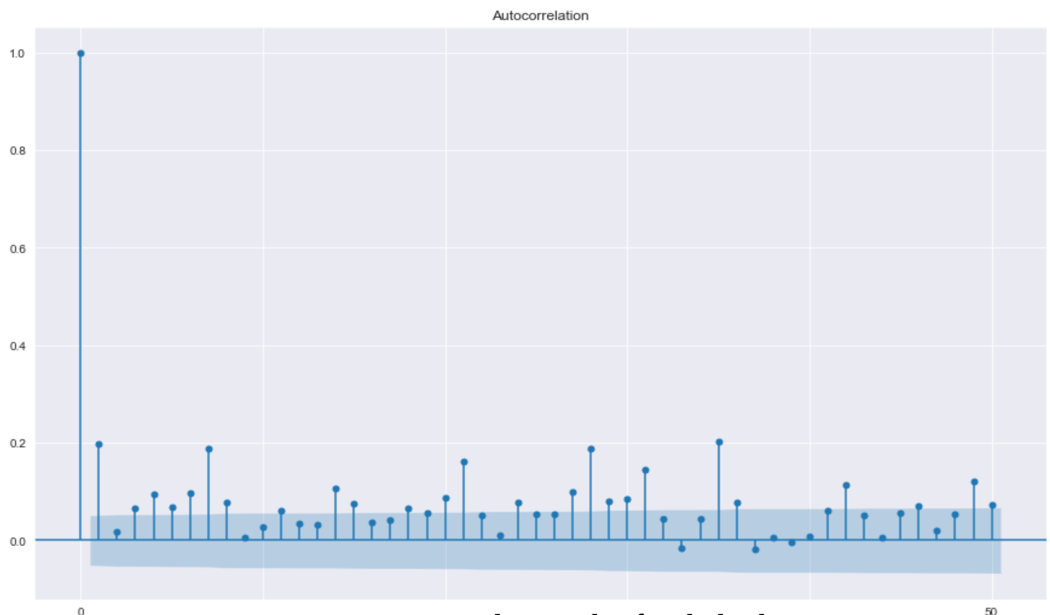


Figure 2.5: Autocorrelation plot for daily data.

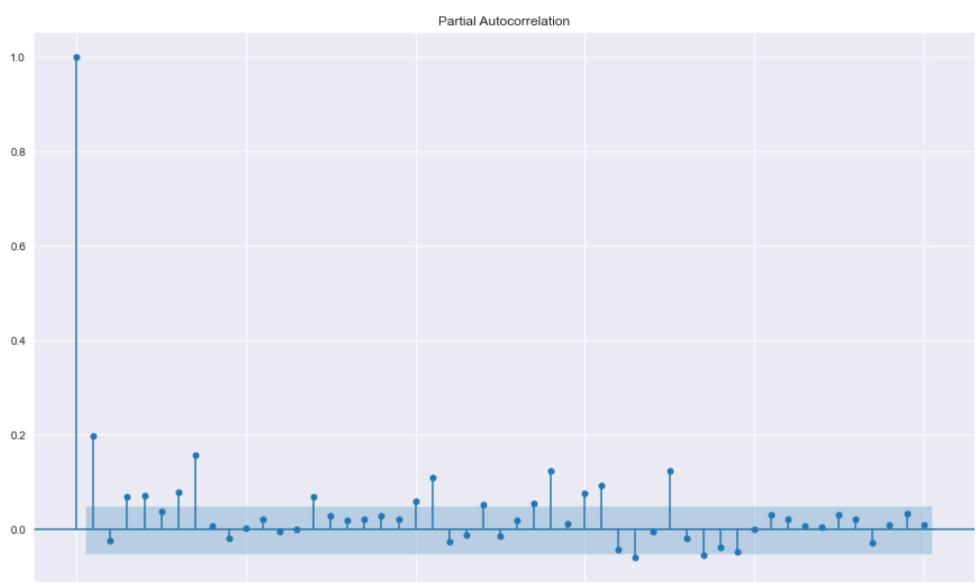


Figure 2.6: Partial autocorrelation plot for daily data.

From the Jarque-Bera test results, we verified that the model was not skewed or had excess kurtosis. Although these are performed to verify normality, we also plotted standardized residuals over time and a Q-Q plot to evaluate whether the residuals were normal. Figure 2.7 represents the standardized residuals over time, and Q-Q plot, it is evident that the residuals are normal.

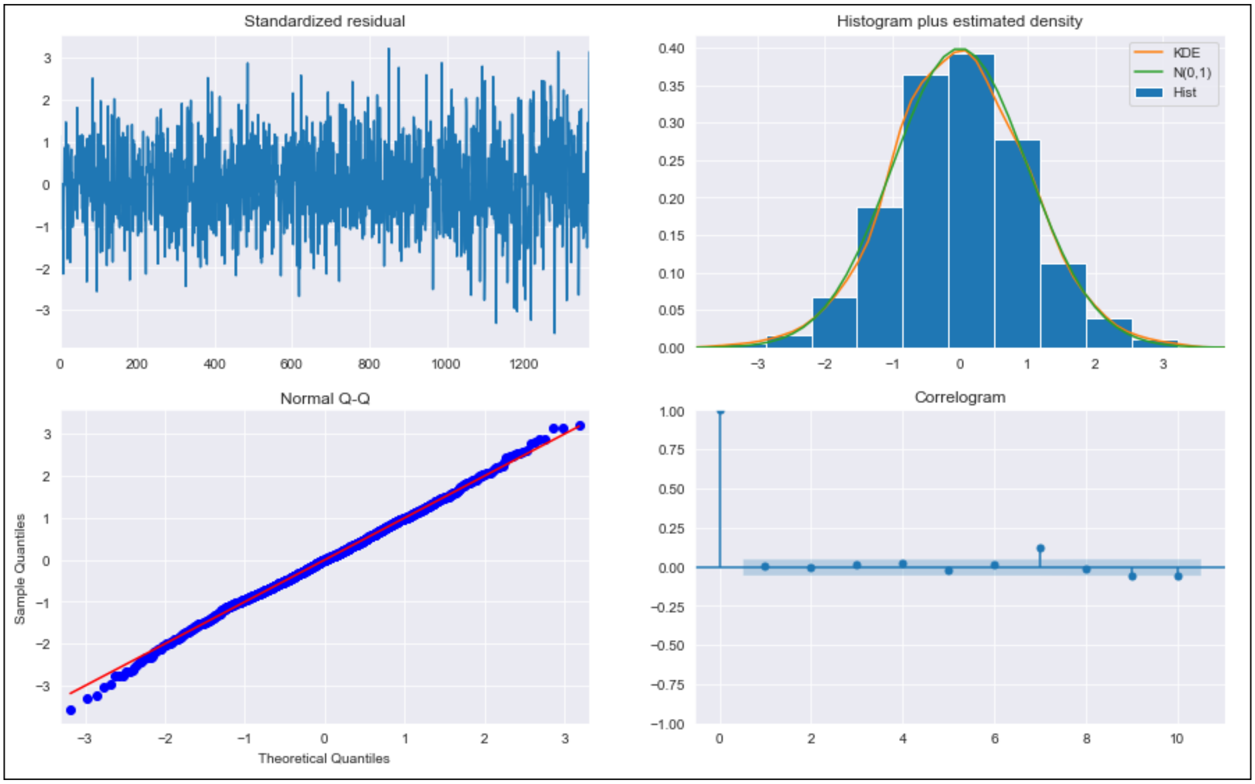


Figure 2.7: Standardized residuals over time and Q-Q plot (ARIMA).

Our next step was to develop the SARIMA model, and the exact same steps were followed. Based on the AIC values, SARIMA (3,0,1) $(1,0,1)_7$ is the best model, whose AIC was 10959.25. From the Jarque-Bera test results, we verified that the model was not skewed or had excess kurtosis. We also plotted standardized residuals over time and a Q-Q plot to evaluate whether the residuals were normal. Figure 2.8 represents the standardized residuals over time and Q-Q plot, and it can be observed that the residuals are normal

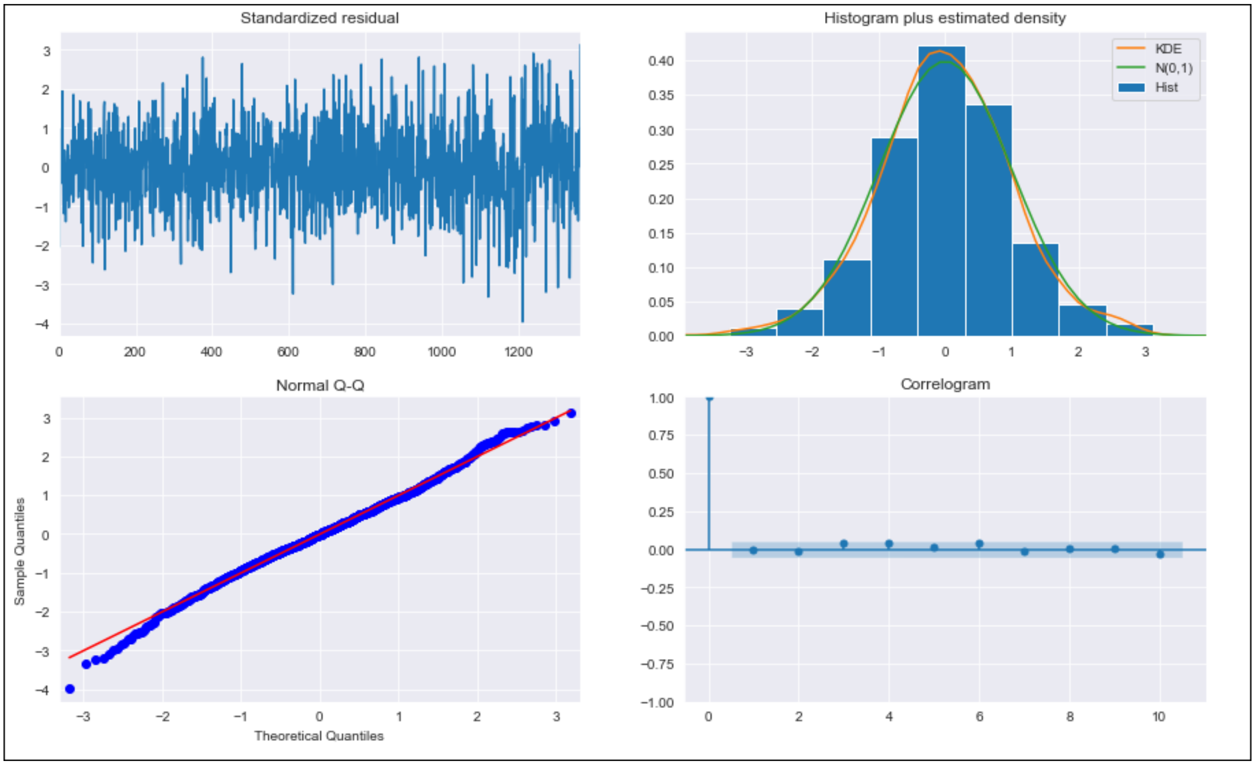


Figure 2.8: Standardized residuals over time and Q-Q plot (SARIMA).

Next, for developing the short-term hourly forecasting models, we repeated the same testing as discussed above and first observed that the data was stationary (p-value < 0.001). Next, we plotted the auto-correlation function and partial auto-correlation function to determine *p* and *q*. Figure 2.9 and Figure 2.10 represent the autocorrelation and partial autocorrelation plot, and we see that neither the partial correlation function nor autocorrelation function vanishes, so we model it as an ARMA sequence. Here we select *p* to vary from 1 to 12 and *q* to vary from 1 to 12, and these numbers were based on the seasonality of the data. Based on the AIC values, ARIMA (2,0,3) is the best model, whose AIC was 4595.95. Similarly, for SARIMA based on the AIC values, SARIMA (3,0,1) $(5,1,0)_{24}$ was the best model, whose AIC was 4204.09. From the Jarque-Bera test results, we verified that the model was not skewed or had excess kurtosis. Additionally, we also plotted standardized residuals over time and a Q-Q plot and observed that the residuals were normal.

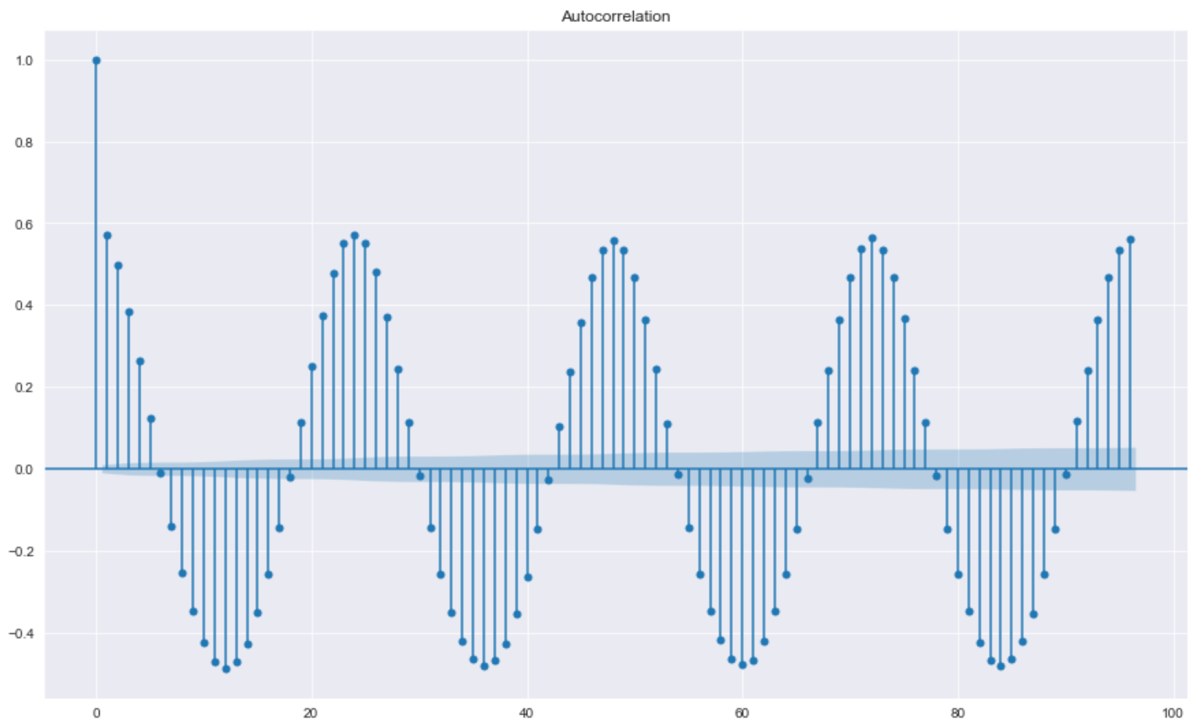


Figure 2.9: Autocorrelation plot for hourly data.

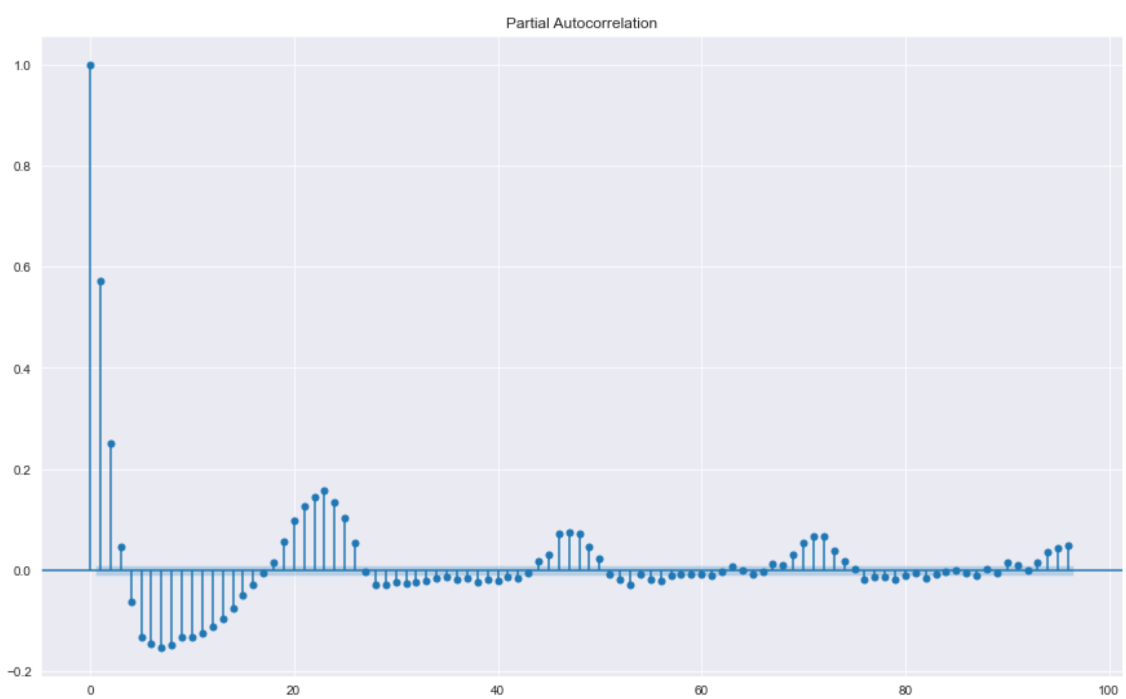


Figure 2.10: Partial autocorrelation plot for hourly data.

2.1.3.2.2 Holt-Winters Forecasting Model

For our next model, we developed a Holt-Winters forecasting model as this approach can account for the level, trend, and seasonality component in the time-series data. A Holt-Winters or triple exponential smoothing model has three parameters: Alpha, Beta, and Gamma, where Alpha specifies the coefficient for the level smoothing, Beta specifies the coefficient for the trend smoothing, and Gamma specifies the coefficient for the seasonal smoothing [75,76]. Additionally, a parameter representing the type of seasonality "*m*" is also included in model [77]. To develop this model, we used the open-source statsmodel package as this allowed us to run multiple models in parallel with different parameters and identify the best model fits based on AIC values. For the trend and seasonality and we have two options of smoothing, either additive or multiplicative. Since our data did not show an exponential increase over time, we used additive smoothing for both. However, to account for the decrease in patient arrivals because of the COVID-19 pandemic, we used a dampening method on the trend smoothing. Figure 2.11 below represents the best Holt-Winters model parameters based on AIC (8299.37) for generating the daily long-term forecast. A model with different parameters and seasonality was used to develop the hourly forecasting model.

	Name	Parameters	Optimized
smoothing_level	alpha	0.075714	True
smoothing_trend	beta	0.010816	True
smoothing_seasonal	gamma	0.034233	True
initial_level	l.0	2137.390327	True
initial_trend	b.0	-22.981985	True
damping_trend	phi	0.990000	True
initial_seasons.0	s.0	-59.800484	True
initial_seasons.1	s.1	172.254333	True
initial_seasons.2	s.2	142.038579	True
initial_seasons.3	s.3	-18.480048	True
initial_seasons.4	s.4	-22.166515	True
initial_seasons.5	s.5	34.798383	True
initial_seasons.6	s.6	-248.644248	True

Figure 2.11: Holt-Winters model parameters for daily long-term forecasting.

2.1.3.2.3 Machine Learning Models

Finally, we also developed two machine learning models: Extreme Gradient Boosting (XGBoost) and Random Forest Regression model. Both are decision tree machine learning algorithms and require a supervised learning approach where each input requires an output pair within the training model for the model to learn and later predict. However, the foundation of each algorithm is different where Random Forest Regression uses a bagging technique, whereas the XGBoost uses a boosting technique for learning. The Random Forest algorithm generates multiple decision trees in the beginning with equal weight and runs in parallel, whereas the XGBoost follows an iterative approach where each tree starts with a single leaf and then expands to multiple trees based on the information gain (learning). For the XGBoost algorithm, although there are numerous parameters, there are six major/primary hyperparameters, which are a) Number of sub-trees, b) maximum tree depth, c) learning rate, d) L1 (reg_alpha) and L2 (reg_ lambda) (e) the complexity control (gamma=γ) and, (6) minimum child weight. Here, the number of sub-trees informs the algorithm when to stop learning, and the maximum depth represents how many splits should be generated from each tree. For identifying the number of sub trees, we used the treelimit function within the XGBoost opensource library, and for depth, we restricted the value to 3 to protect from overfitting by increasing the number. The learning rate represents the constant that is multiplied by the weight in each tree to continue learning. We tested a few values and used 0.01 as the learning rate as this allowed for the best performance on our dataset. The consensus is that a lower learning rate generates a better model fit at a higher computational cost. L1 and L2 are regularization parameters used to avoid overfitting models by lowering variance while increasing some bias. However, since we had a single feature, we used the default values. gamma=γ was initially set at 0, and based on the training and testing speed, we varied this value to control for the complexity from loss. Finally, to avoid overfitting, we utilized a minimum child weight of 1 as this is considered a “safe” practice [78,79].

The Random Forest Regression model also has six core hyperparameters, similar to XGBoost. Starting with the first one, n_estimators, which represents the number of decision trees that will be used in the model. We identified this number by running multiple scenarios and choosing the one that returned the leaset mean absolute error (MAE). The next hyper parameter was criterion which represents the performance metric such as MAE, root mean square error (RMSE), etc., to calculate the loss function. We developed the model using each and observed the one using MAE outperformed the RMSE. The third criterion is are max_depth, which is the same as the one mentioned for XGBoost. The fourth one is max_ features and represents the maximum number of features the model should consider when determining a split, and in our case, we had only a single feature. The last two are bootstrap and max_samples. Bootstrapping process randomly takes a set of samples from the data, learns and makes predictions out of it, and replaces the sample back in the dataset. The idea of this method is to infer population results from the small subsets of the data. These predicted results are then averaged to potentially obtain better results. The max_samples hyperparameter represents the maximum number of samples from the training dataset that will be given for any individual tree for bootstrapping. In our models, we used bootstrapping to improve the efficacy of the forecast and used max_samples of 90 or 168 based on long-term or short-term forecasts.

For each model discussed above, the models were trained using the total arrivals, and the ESI levels were not used. However, for the forecasts, the model developed using the total arrivals was used to forecast each ESI level. Although this is not the ideal scenario, this provides an opportunity for physicians and administrators to get some sense of the ESI levels of arriving patients.

2.1.3.2.4 Prediction Evaluation Metrics

To evaluate the performance of each model, we utilized three performance metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percent Error (MAPE).

Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\sum_{t=1}^{n} \frac{(y_t - \hat{y}_t)^2}{n}}$$

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{t=1}^{n} |y_t - \hat{y}_t|$$

Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|y_t - \hat{y}_t|}{y_t}$$

Here y_t represents the actual value and $\hat{y}_t$ represents the forecasted value in time t, and n is the number of time units. Since each metric has its own advantages and limitations, we decided to report all three metrics to capture the whole picture. Both RMSE and MAE are not independent scale metric and does not allow for direct comparison over various time series. RMSE is built around the mean and penalizes the outlier heavily, whereas MAE is built around the median and protects against penalizing the outliers heavily. However, as mentioned earlier, MAE does not allow comparison across different scales and has a higher chance of bias. To allow for comparison across multiple time series, we report MAPE. Thus, reporting these different metrics allowed for presenting the advantages and drawbacks of each forecasting model.

2.1.4 Results

Upon developing the model and tuning the parameters to achieve the best performance metric on the training data, the next step was to use these models to forecast the future patient arrivals to the ED. We first discuss the findings from the long-term forecasting model, which forecasts the patient arrivals to the ED for the next 90 days. The predicted model output was compared against the actual data to calculate each performance metric. Table 4 below represents the performance metrics score for each model for the long-term forecasts.

Table 2.4: Model performance for the 90-day forecast.

Model	RMSE	MAE	MAPE
MA	30.1	23.6	14.2%
ARIMA	27.2	21.6	10.6%
SARIMA	25.6	19.2	9.9%
Holt Winters	26.8	19.8	10.0%
XGBoost	16.6	14.1	5.9%
Random Forest	17.4	14.6	6.4%

From Table 2.4 above, it is evident that both machine learning models outperformed the naïve model and other traditional time series models. However, it is interesting to notice that the Holt-Winters approach outperformed the ARIMA model, and this can be primarily attributed to the fact that the Holt-Winters model can account for seasonality. However, comparing the SARIMA model to the Holt-Winters model, SARIMA was slightly better.

The most significant improvements were observed with the machine learning models, where the MAPE value was reduced by half compared to the traditional time series forecasting model. Among the two machine learning models, XGBoost outperformed the Random Forest model for all the performance metrics. One of the key observations here is the high RMSE values irrespective of the forecasting approach, which could be potential because of the extreme values (outliers). To investigate this, we plotted the forecasts from the best-performing model against the actual values. Figure 2.12 below presents the daily forecasted values against the actual values for the 90-day period. It is clear from the data that there are few outliers, and the pattern is not as seasonal as compared to the prior data. This is primarily because of two reasons: a) the last 3 months of 2020 observed a significant variation in ED patient demands, and b) these months led to increased patient demands due to COVID-19 patients. On a positive note, even with a significant change in patient demands and arrivals, the machine learning models forecast were robust (based on RMSE, MAE, and MAPE) as models with a MAPE value of 5.0% are considered excellent [80]. However, to avoid bias and over-relying on one value, we look at RMSE (16.6), which is comparatively low given the daily arrivals vary from 150 patients a day to as high as 270 patients.

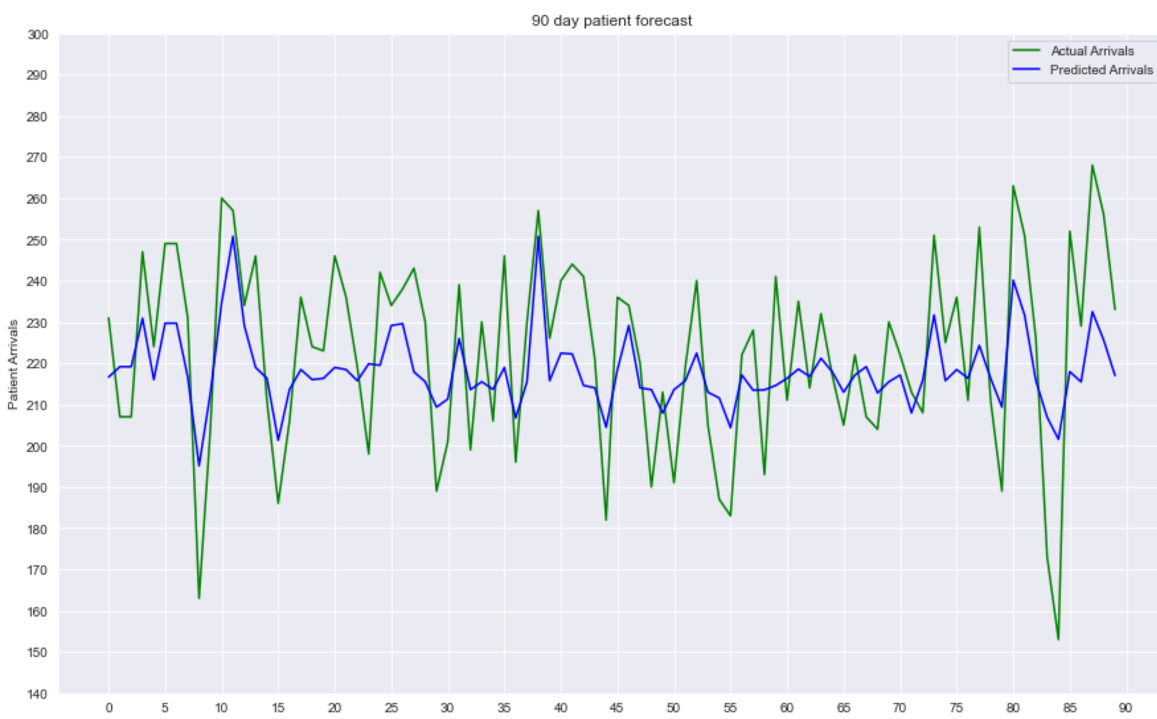


Figure 2.12: Daily XGBoost forecast against actual values.

After identifying the best-performing model, the next step was to look at the ESI predictions for the 90-day forecast. Table 2.5 below represents the performance metrics score for each ESI level from the XGBoost forecast.

Table 2.5: XGBoost ESI level forecast for 90 days.

ESI	RMSE	MAE	MAPE
ESI 1	3.1	2.8	38.0%
ESI 2	8.9	7.0	12.5%
ESI 3	13.4	10.1	8.4%
ESI 4	6.0	5.1	15.5%
ESI 5	1.9	1.4	46.1%

The first thing to notice from Table 2.5 above is the varying RMSE, MAE, and MAPE values across the ESI levels. Specifically, it can be noticed that MAPE values are high for ESI 1 and 5 and minimum for ESI 3, whereas the RMSE and MAE behave vice versa. This represents the bias associated with each metric where MAPE penalizes heavily when the forecasted values are smaller as it is a percentage value. For example, if the residual is 1 and the actual value is 5 then the percent error here is 20%, but if actual values are larger, the percent error value will shrink. Since ESI 1 and 5 together contributes only towards 5% of daily arrivals, a small variation in prediction is penalized heavily by MAPE values. However, by using a combination of three performance metrics, we can identify that the ESI-level forecasts from the model are still robust.

Next, we forecast our short-term prediction, which is an hourly prediction a week ahead. Table 2.6 below represents the performance metrics score for each model and it is evident that both machine learning models outperformed the naïve model and other traditional time series models. Similar to the long-term predictions, the Holt-Winters approach outperformed the ARIMA model because of the seasonality component in the Holt-Winters approach. However, among the traditional time-series models, SARIMA was still the best performing model. The most significant improvements were again observed with the machine learning models, where the RMSE, MAE, and MAPE values reduced significantly compared to the traditional time series forecasting model. Among the two machine learning models, XGBoost slightly outperformed the Random Forest model for all the performance metrics.

Table 2.6: Model performance metric for the hourly forecast.

Model	RMSE	MAE	MAPE
MA	4.2	3.5	49.3%
ARIMA	3.4	2.6	44.1%
SARIMA	3.3	2.5	39.8%
Holt Winters	3.4	2.5	42.4%
XGBoost	2.2	1.4	32.4%
Random Forest	2.3	1.5	34.2%

Similar to the observations we had while predicting daily ESIs that contributed very less to the patient arrivals, the MAPE values are really high. As a benchmark, we compared these numbers to a prior study forecasting hourly patient arrivals to the ED and observed the same pattern, and MAPE values from our best performing model were lower than their best performing model's MAPE [53]. To further investigate and ensure that our forecasts were not significantly different from the actual observations, we plotted the hourly prediction generated for a week against the actual values. Figure 2.13 on the next page presents the hourly forecasted values against the actual values for one week.

It is evident from the figure that there are some hours of the day where the forecasted values and actual values show a mismatch. However, in general, the forecasts track the actual patient arrivals for most of the week. Although some prior studies have used the approach of adjusting extreme values/outliers, we did not follow that approach to maintain the data integrity and be representative of the variability observed in the ED. Next, using the XGBoost model, we forecasted the ESI-level hourly forecast for the same time frame.

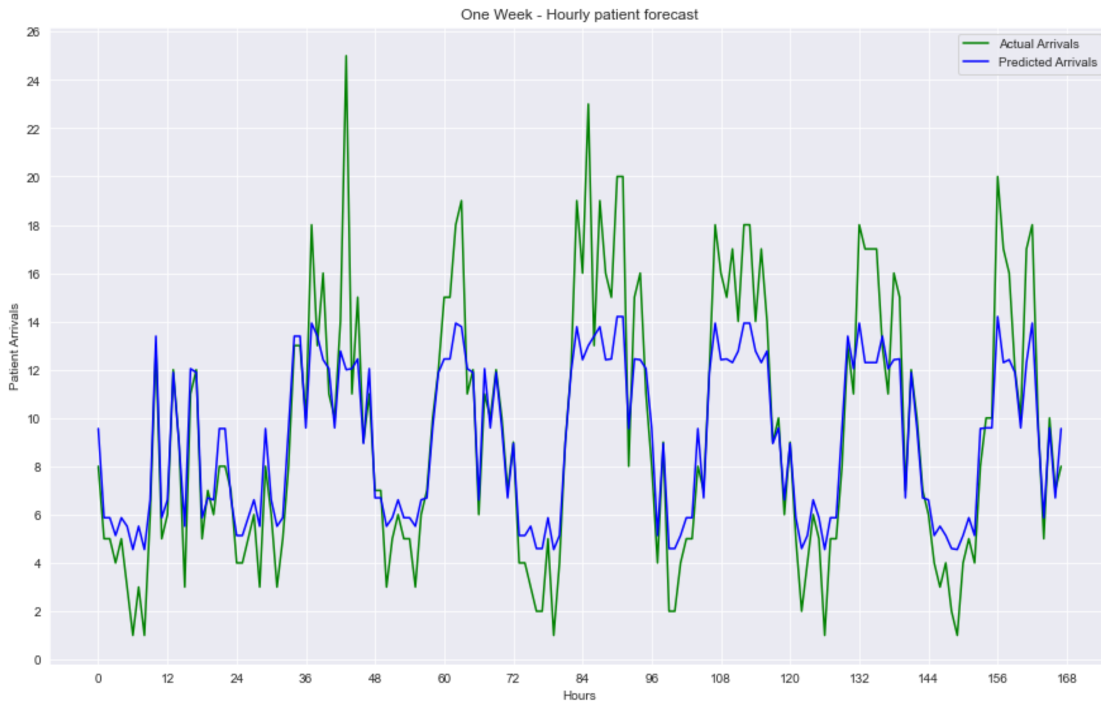


Figure 2.13: Hourly XGBoost forecast against actual values.

Table 2.7 below represents the performance metrics score for each ESI level from the hourly XGBoost forecast. We removed the MAPE from the table for hourly ESI level calculations as those values were significantly high because the range of values (arrivals) for some ESI was as low as 0. It can be observed that both MAPE and RMSE values are low, given the hourly patient arrival range was as low as 1 to as high as 16 for certain ESI levels.

Table 2.7: XGBoost ESI level hourly forecast for one week.

ESI	RMSE	MAE
ESI 1	0.5	0.44
ESI 2	1.2	0.93
ESI 3	1.8	1.40
ESI 4	1.2	0.99
ESI 5	0.5	0.30

2.1.5 Discussions and Conclusions

Protecting the ED from crowding is one of the highest public health priorities, and each ED across the US uses both long and short-term planning to mitigate the consequences of ED crowding. The solutions could vary from short-term fixes such as adding temporary beds to long-term planning of a complete overhaul of shift designs. Beyond the short and long-term planning, sometimes it requires ad-hoc actions such as adding hallway beds, etc. Although sometimes ad-hoc actions are required because of unexpected issues such as evacuations and natural disasters, most of the time, these are required because of inadequate short and long-term planning. One of the most important inputs required for robust planning is the future patient census (arrivals) to the ED. Over the last decades, several studies have applied numerous approaches for forecasting patient arrivals to the ED and have generated acceptable results [51–63]. However, most of these studies have focused on predicting daily patient arrivals to the ED except for two recent studies that have explored hourly patient arrival forecasting [53,60]. Moreover, among these two, only one study has investigated both long-and short-term forecasting [53]. Surprisingly, neither of these studies have included ESI levels of forecasted patient arrivals, and the latter study only explored machine learning algorithms.

Our research developed traditional time-series models and machine learning models to forecast long-term (daily forecast – 90 days ahead) and short-term (hourly forecast – one week ahead) patient arrivals to the partner ED with the patient's ESI levels. This study used two simple input variables: patient arrival time and ESI levels, exportable from any hospital EHR database, and forecasted the daily and hourly arrivals using traditional time-series approaches including ARIMA, SARIMA, Holt-Winters and two machine learning algorithms XGBoost and Random Forest regression. Machine learning algorithms outperformed the traditional time-series, forecasting models. XGBoost generated the best long-term and short-term forecasts with MAPE values of 5.9% and 32.4% outperforming prior studies. Moreover, we forecast ESI levels of these arrivals for the long-term and short-term with maximum RMSE values of 13.4 and 1.4. These findings are promising especially given the simple input variables and the realistic time horizon of the forecasts to inform both long- and short-term planning.

Future research will focus on incorporating other simple parameters that can be exported from EHR to investigate if the model predictions can be improved. Additionally, a hierarchical forecasting approach with an optimization function could potentially improve ESI-level forecasting. Lastly, these forecasts are only practically useful if there exist scheduling tools to input the forecast output and estimate the necessary changes in resource allocation. In the next two chapters, we discuss two such tools that can be used for estimating resource allocation requirements by using the patient arrival census as input.

3. Chapter Three

3.1 Overlapping Shifts to Improving Patient Safety and Patient Flow in ED

3.1.1 Introduction

According to the 2017 Centers for Disease Control and Prevention (CDC) reports, approximately 151 million visits are made to Emergency Departments (EDs) in the US annually [3]. These numbers are expected to increase based on the current trends where ED patient arrivals have seen a 24.7% increase over the last ten years [81]. Although studies have reported the lack of primary care access, substance use disorders, and psychotic issues as a significant cause of increased patient arrivals to EDs, the key reason driving the increasing patient arrivals are underserved patients [82]. While the Affordable Care Act has helped reduce ED visits by uninsured patients, ED access by the underserved population has increased significantly [83]. Moreover, the federal mandate, Emergency Medical Treatment and Active Labor Act (EMTLA), which requires an ED physician to provide stabilizing care to a patient irrespective of their ability to pay, makes ED the healthcare safety net [2]. According to the latest reports, about 70% of inpatient hospital admission occurs through the ED, and an additional 3% are transferred to a different hospital for inpatient admission [84].

The diverse nature of patients seeking medical care and the overwhelming volumes of arrivals make ED one of the most complex healthcare environments predisposed to crowding. The American College of Emergency Physicians defines crowding as a situation in which the identified need for emergency services exceeds available resources for patient care in the emergency department, hospital, or both [4]. ED crowding is a patient safety issue as well as a public health problem. Crowding results from multiple factors, including high patient volumes, inadequate staffing, bed shortages, etc., and negatively impacts patient flow, patient safety, and health system expenditure [5–7,9–11]. Additionally, the reduced availability of ED beds due to admitted patients awaiting transfer into an in-hospital bed restricts an ED's capacity to accept new patients, resulting in higher patient boarding time and delays in providing patient care [8]. It is imperative that ED resource allocation, which includes staffing and availability of providers, number of beds, and ancillary units, including labs, radiology, etc., are well planned to improve the patient flow and patient safety within the health system and avoid ED crowding. Although the most intuitive solution to address crowding is adding extra resources, including beds, staff, and ancillary units, adding new hospital resources could be very expensive. Moreover, researchers have observed that rather than adding physical resources (e.g., bed, equipment, machines, etc.), temporarily adding or changing staff schedules are comparatively cheaper options. However, the schedules should be generated carefully such that they can improve patient flow and patient safety in the ED without overstaffing.

Patient safety is an integral aspect of the ED as continuous patient flow and interactions with multiple departments and providers make it prone to errors. Recent studies have observed ED as one of the hospital departments with high error rates. Some of the common sources of ED errors are interruptions, miscommunications, and loss of information. Handoffs, transfer of a patient's care and responsibility from one physician to another, are fraught with miscommunications, omissions, errors, and information loss [85,86]. However, handoffs are unavoidable in EDs as they operate throughout the day, and a physician ending their shift is required to transfer their current patients to the newly arriving physician. Although unavoidable, handoffs should be minimized, as it is a significant patient safety concern.

A recent study where thirty-six ED physicians were shadowed for over 100 hours observed that a physician's likelihood of making an error while prescribing was significantly higher when interrupted [87]. Similarly, studies have observed that approximately 80% of serious medical errors involve miscommunication during the patient handoff [88]. Additionally, poor handoffs, which involve miscommunication, can lead to conflicting expectations for information and contribute to delayed patient onboarding and conditions that can pose safety threats [89]. Further, studies that specifically investigated ED shift-change handoffs observed that for approximately 75% of the patients, the vital signs were not communicated, and errors were observed in about 60% of cases [90]. Finally, insurance claims involving missed ED diagnoses that harmed patients reported that 24% of the cases involved inadequate handoffs [91].

From the literature, it is evident that ED patient handoffs have a negative impact on patient safety. Hence, while developing ED physician staffing schedules, it is crucial to consider handoffs as a performance metric along with other patient flow metrics. To our knowledge, most of the prior studies that used simulation modeling or mathematical modeling approaches have focused only on the patient flow in the ED, and none have considered patient safety metrics as a performance indicator of the ED. This research developed a novel hybrid simulation model for identifying shift policies that can improve patient safety and patient flow in the ED while not negatively affecting other Centers for Medicare & Medicaid Services (CMS) core metrics.

3.1.2 Background and Literature

The contribution of operations research models and methodologies has had a significant impact on improving EDs throughout the world. A variety of approaches, including mathematical and optimization models, queuing theory, simulation modeling, and probabilistic models, have been used to address a variety of ED issues, including resource allocation, patient streaming, fast track ED, staffing, and scheduling, etc. Although various tools have been used to improve ED operations, researchers have endorsed simulation models as one of the best tools to model different phases of patient flow (arrival to departure) in the ED because of the complexity and nature of ED [92,93]. Specifically, researchers have identified discrete event simulation (DES) to be efficacious in representing and simulating ED activities [94,95]. Additionally, the ability of simulation tools to model different ED processes, phases of patient flow, and test "what-if" scenarios make it an essential tool to investigate staffing and scheduling, resource allocation, and overall process improvement before implementing changes [96].

One of the earliest studies that utilized DES for bed allocation was half a century ago [17]. It investigated various scenarios that compared the impact of the grouping of patients and its impact on bed utilization. Additionally, simulation models' capability to delve into the micro details of processes has helped in understanding the bottleneck leading to an increased length of stay and, thus, assisting in resource allocation [97,98]. Further, studies have used simulation models to test each variable's impact, including different resources in the ED, to identify their impact. Specifically, one study identified that adding a single doctor and nurse during ED peak hours was found to impact patient waiting times the most [99]. Similarly, a recent study used the DES modeling approach to identify the number of different resources it would require, including beds, staff, equipment etc., to meet specific key performance metrics such that the desired patient flow is achieved [100]. Moreover, researchers have used the DES modeling approach to compare a pod versus unit-based ED and observed that pod-based ED improves the quality of care metrics by slightly increasing resource utilization [13]. Finally, studies have combined simulation modeling with optimization to identify the optimal amount of ED staff and other resources required to improve the patient flow [101–103]. One particular study has observed that without any new addition of resources, a simulation optimization model was able to reduce the patient wait time by 40% and increase the throughput by 28% [101]. Similarly, studies have used linear optimization models to identify a resource's contribution to ED workflow and test a variety of shift schedules. One study investigated the impact of a staggering shift schedule and observed that it reduced the LOS and the number of staff required [104].

All these studies aimed to improve the patient flow in the ED using different approaches but mainly focusing on resource allocation. Hence, most studies considered physicians as resources, thus failing to capture different physician activities in the ED. Similarly, a majority of the prior studies modeled the physician-patient encounter as a single visit with a time delay that fails to capture the multiple physician-patient interactions, physician placing patient orders, the possibility of handoffs, etc., as observed in the ED. We observed one specific study that modeled physicians as agents for the physician-patient assignment, but the study was limited to investigating patient onboarding time and did not consider the other performance metrics [105].

Although numerous studies have used simulation models to test different staffing schedules, staffing levels, and resource planning in the ED to improve patient flow in the ED, none of the studies has considered combining patient safety and patient flow metrics to evaluate the ED. Similarly, from the medical literature, studies that investigated the negative impact of handoffs have focused on improving the quality of handoffs by standardizing handoffs using templates, using dedicated space for handoffs, bedside handoffs, etc., rather than reducing the number of handoffs [106–109]. One of the earliest recommendations on interventions to reduce the number of handoffs was a decade ago when a group of ED physicians recommended that overlapping shifts could potentially reduce the number of handoffs [110]. However, no observation or intervention was performed to investigate the recommendation.

We identified a recent study that implemented and investigated the impact of overlapping shifts on handoffs [111]. The study was conducted in a pediatric academic ED with an annual volume of 46,000 patients, and ED physician shifts with three-hour overlap were implemented. Compared to the non-overlapping policy, the new overlapping policy that restricts physicians from taking new patients during the end of their shift was able to reduce the percentage of handoffs by 25%, with a non-significant increase in the patient time in the ED. However, translating such policies to larger Level 1 trauma centers like our partner ED, which sees over 106,000 patients over multiple pods, is expensive and requires validation before implementation. Moreover, implementing ED physician shifts with a three-hour overlap in multiple pods can also lead to a higher cost burden for larger systems. Hence, a risk-cost-benefit analysis of the same is warranted before implementing such policies.

As the first step in this direction to investigate if such policies can be translated to other larger EDs, we developed a proof-of-concept model using a publicly available dataset and tested a few policies [15]. In this proof-of-concept model, we did not model individual pods and other details pertinent to our partner ED but considered the whole system as a single unit with bed capacities and staffing representing our partner ED. Further, we tested only the impact of restricting the leaving physician from taking high-severity patients during the last hour of the shift and did not consider overlapping shifts.

3.2 Phase One

3.2.1 Methods

3.2.1.1 Data

Data used in this study (average door to physician time, wait time in the ED, treatment time, and total time in the system) was obtained from the publicly available National Hospital

Ambulatory Medical Care Survey (NHAMCS) 2011 and 2015. NHAMCS is a Centers for Disease Control and Prevention (CDC) initiative to collect data on the utilization and provision of ambulatory care services in hospital emergency and outpatient departments and ambulatory surgery locations. Findings are based on a national sample of visits to these departments. The data regarding the shift schedule, the capacity of the ED, the number of physicians available for a shift, etc., were obtained from the ED of the partner hospital, Greenville Memorial Hospital (GMH) in South Carolina. The research team included an ED physician working in the GMH, SC, for guidance and developing policies, which are discussed later in the paper.

We first introduce Figure 3.1, which represents the total time spent by a patient in the system based on the data from the NHAMCS mentioned above. From Figure 3.1, we split the data into evaluation time and additional care time, as shown in Table 3.1. Evaluation is the time spent by a physician observing the patient (direct contact with the patient), whereas additional care is the time spent by a nurse (running tests, providing meds, etc.) or time spent with a consulting physician if requested.

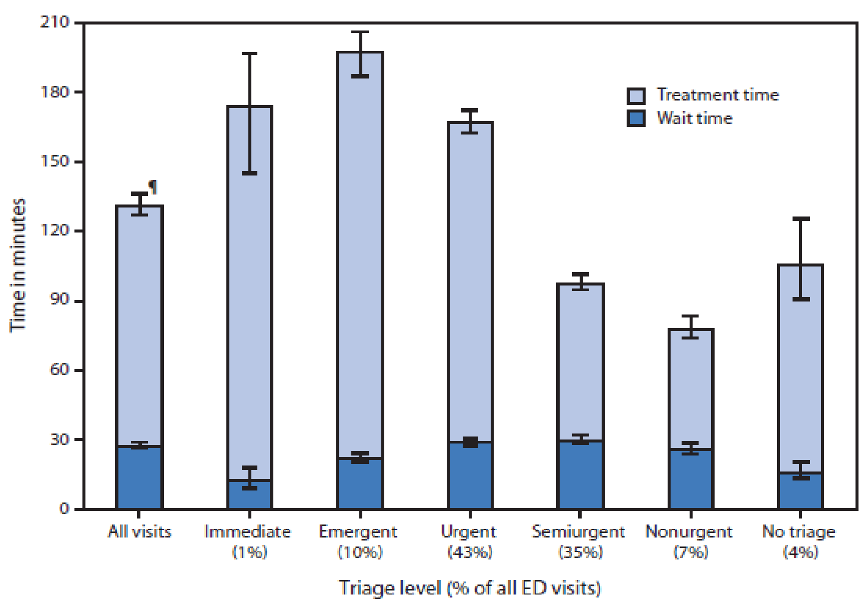


Figure 3.1: Patient time spent in the ED for different triage levels.

As seen in the table, the total evaluation time for a patient is comparatively lower than the total time spent providing additional care, which is consistent with prior studies (Hollingsworth et al. 1998; Hill et al. 2013) and observations from the GMH ED. The evaluation time and time for additional care were split based on the severity of the patient. In the case of level one patients where the condition is critical, physicians spend more time during the first evaluation trying to stabilize the patient. Whereas in cases two and three, the physician initially runs a few tests to comprehend the ailment; hence the initial evaluation is lesser than the second evaluation. For cases four and five, which are less severe, the physician spends almost the same time for the first and second evaluations. In general, the total evaluation time contributed to 30-50% of the total time.

Table 3.1: Time spent by a patient in the ED.

	Severity				
Activity	**1**	**2**	**3**	**4**	**5**
Evaluation 1	TRIA(33,35,37)	TRIA(13,15,17)	TRIA(8,10,12)	TRIA(12,14,16)	TRIA(8,10,12)
Additional Care 1	TRIA(28,30,32)	TRIA(28,30,32)	TRIA(23,25,27)	TRIA(23,25,27)	TRIA(20,22,24)
Evaluation 2	TRIA(23,25,27)	TRIA(23,25,27)	TRIA(20,22,24)	TRIA(6,8,10)	TRIA(6,8,10)
Additional Care 2	TRIA(28,30,32)	TRIA(38,40,42)	TRIA(18,20,22)	TRIA(21,23,25)	TRIA(8,10,12)
Evaluation 3	TRIA(20,22,24)	TRIA(11,13,15)	TRIA(8,10,12)	**N/A**	**N/A**
Additional Care 3	TRIA(21,23,25)	TRIA(20,22,24)	TRIA(18,20,22)	**N/A**	**N/A**
Evaluation 4	**N/A**	TRIA(8,10,12)	TRIA(8,10,12)	**N/A**	**N/A**
Additional Care 4	**N/A**	TRIA(18,20,22)	TRIA(18,20,22)	**N/A**	**N/A**

Patient arrivals are represented in Figure 3.2 based on the data from a previous study [28]. Note that activity is low in the early morning hours, but there is a steady increase from 7:30 am until 12:00 pm, at which point patient arrivals remain consistent until 5:00 pm.

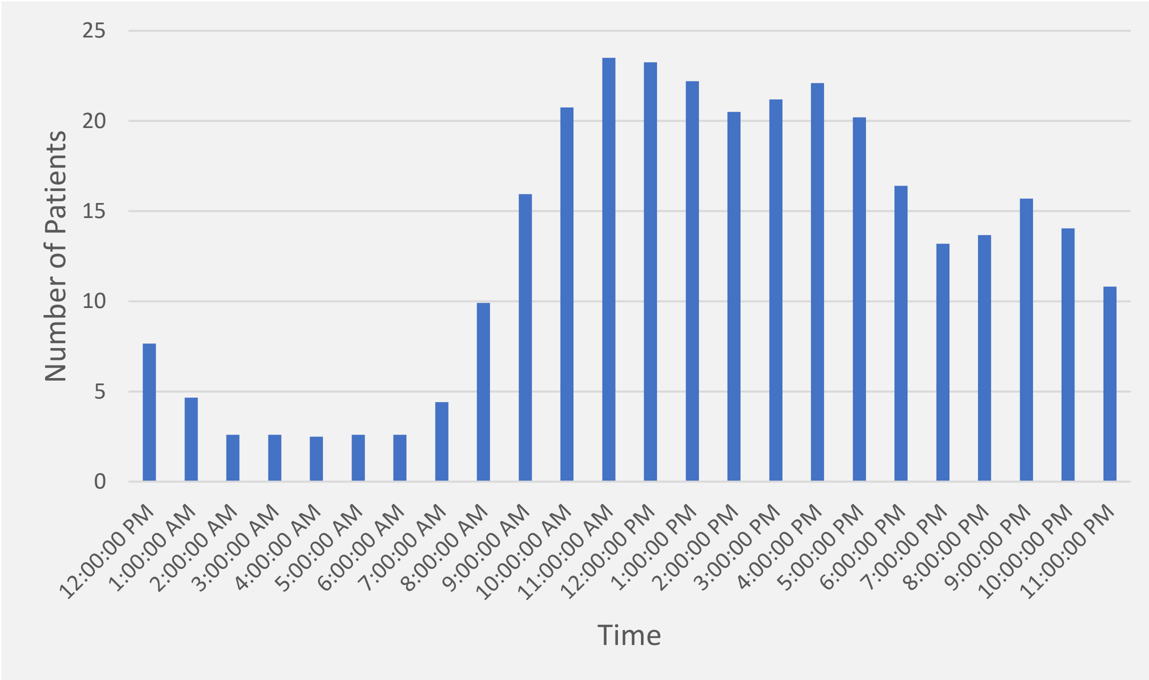


Figure 3.2: Patient arrival rate to the Emergency Department.

3.2.1.2 Initial Simulation Model

The proof-of-concept model that aimed to minimize the number of handoffs was built in Arena using a discrete event modeling approach. However, unlike the traditional modeling approach adopted in most previous research, which considered the physicians as a resource and the patient as an entity receiving treatment, this simulation model considers the physician and the patient as agents that flow in the ED. Under this modeling method, physicians and patients carry unique attributes and can contribute to their own actions and the actions of others. This modeling approach provides the flexibility of replicating physician activities in the ED, such as searching and accepting a patient, interacting with patients based on their severity, performing patient handoffs, and charting, which would be difficult to accommodate if physicians were modeled as a resource.

The model is initialized by creating patients and physicians. Upon patient arrival, the severity of a patient is assessed on a level of 1-5 by a triage nurse, where 1 demands immediate attention and five is considered the least urgent. The patient is next registered into the hospital electronic health record and waits in the waiting room, where they are prioritized based on the initial triage-severity level assigned. The ED nurse collects the patients from the waiting room based on their severity level, availability of a physician, and ED rooms. In the case of patients with a severity level of 1, they are taken to the trauma bay rather than to the normal ED room. The ED rooms and trauma bays are modeled as resources where the capacities of these resources are the same as their capacities in the GMH ED.

Upon a physician's arrival on shift, the physician who will be leaving the ED must transfer his patients to the arriving physician. As mentioned earlier, this process of transferring the care of a patient is defined as a handoff. Post handoff, the physician decides on taking a new patient depending on the current number of cases handled. In current practice, it is not common for the oncoming physician to check how many patients the other physicians are currently handling. Thus, the physician only considers whether or not he can accommodate another patient instead of trying to balance the workload among physicians.

If the physician accepts a new patient, the physician meets the patient in the ED room for the first evaluation, after which the physician returns to the station to document in the medical record, order tests, medicines, consult, etc. The nurse then completes their required documentation, physician-ordered tasks, medication administration, and runs bedside tests or ordered interventions. Patient care often includes diagnostic imaging that may require the patient to be moved out of the ED to the radiology suite. Following the drug administration, imaging, and diagnostic testing, the physician returns to the patient for the subsequent evaluation, and the physician remains focused on that patient until clinically stabilized. After the evaluation, the patient is either discharged or admitted as an inpatient to the hospital, and the physician may take on a new patient. Although the patient time of day arrival is relatively predictable, the variability of patient acuity is not. Thus, the ED physician’s workload and the need to take on a new patient is influenced by triage severity level regardless of the number of patients currently under the physician's care.

As mentioned earlier, our primary goal in this research was to reduce the number of handoffs, and hence our focus was to replicate the physician's behavior in the ED successfully. The modeling approach adopted was able to satisfy this goal successfully. Although we do not consider the triage nurses, nurses, consults, and in-hospital bed placement as specific entities or resources in the simulation model, the delays associated with each process were incorporated as probability distributions. This approach was adopted as it does not affect the efficacy of the model to replicate the physician behavior in the ED.

3.2.1.3 Simulation Policies

To comprehend the best policy to reduce the number of handoffs in the ED, the current GMH physician-patient assignment policy was considered as the baseline policy. To make sure that ED performance was not influenced by new policies, three performance measures were used based on the prior studies which used the same to measure the performance of an ED [112,113].

Table 3.2 below represents the three performance metrics and their definition.

Performance measures	Definition
Number of handoffs	Number of patients transferred b/w physicians
Throughput	Total number of patients discharged
Treatment time	The time between first physician contact and patient disposition discharge or admission to hospital

Table 3.2: Performance measures.

3.2.1.3.1 Policy 1 (Baseline policy)

This policy depicts the current policy adopted by the physicians working in GMH for patient management in the ED. The arriving ED physician has a minimum of a two-hour overlap with physicians working on the prior shift. Hence, upon a physician's arrival, they wait for the physician who is leaving in the physician's station for the patient handoffs. In this model, a physician, after their arrival, waited for 5 minutes on average in the physician station for the departing physician to arrive and start the handoff. Post handoff, depending on the number of patients managed, the physician decides on taking a new patient or evaluating an existing patient. In this policy, a physician handles no more than six patients at a time, and new patients can be accepted only after discharging an existing patient. In the present scenario, after receiving a new patient, the physician evaluates the patient in the ED room and returns to the physician station to document the medical record, order tests and medicines, and consult depending on the situation. For the subsequent visits to a patient, the physician may not necessarily return to the physician station after each evaluation. However, the physicians working also make sure that they return to the station and take new patients so that the ED rooms are not left vacant. Although this policy maintains a restriction regarding the maximum number of patients that a physician could manage at a time, it does not restrict the physicians from receiving the patients irrespective of the time remaining in their shift.

3.2.1.3.2 Policy 2

In this policy, we restrict the physicians from signing up a new patient during the last 15 minutes of the shift. Additionally, to reduce the possibility of handoffs, we restrict the physician from accepting high acuity cases (level 1, 2 & 3) that needs longer treatment time and reduce the maximum number of patients that can be managed by a physician to four for the last 120 minutes. Moreover, another reason for restricting physicians from accepting high acuity patients is based on prior studies, which have proved that physician's productivity decreases as the shift progress and increases the chances of errors (Jeanmonod et al. 2008; Silverman 2011).

3.2.1.3.3 Policy 3

In this policy, we reduce the maximum number of patients that can be handled by a physician to five, and we restrict the physicians from signing up a new patient during the last 15 minutes of the shift. However, no specific measures were adopted to restrict physicians from accepting high-severity patients during the end of their shift.

3.2.1.3.4 Policy 4

In this policy, we reduce the maximum number of patients that can be handled by a physician to five, and we restrict the physicians from signing up a new patient during the last 15 minutes of the shift. Additionally, we restrict the physician from accepting high acuity cases (level 1, 2 & 3) that needs longer treatment time and reduce the maximum number of patients that can be managed by a physician to four for the last 120 minutes.

3.2.2 Results

The four policies were tested and compared using a simulation model. The model performance under each policy was tested using the performance measures detailed in Table 3.2. As explained earlier, the changes in the policies included the maximum number of patients a physician could handle and restrictions regarding accepting a new patient. For testing purposes, the simulation was run for a week and over 600 replications such that a half-width of 5 minutes on treatment time was achieved (as seen in Table 3.3). Note that handoffs were reduced considerably under each of the alternative policies compared to the first policy. All other performance measures also improved or stayed the same under the new policies.

Table 3.3: Initial results.

Policy	**#Handoffs per day**	**Throughput per physician**	**Treatment time (mins)**
1	47.8 (1.8)	6.3 (1.6)	246.5 (1.1)
2	41.1 (1.5)	6.2 (1.6)	262.0 (0.8)
3	42.3 (1.5)	6.4 (1.7)	212.7 (1.1)
4	37.4 (1.3)	6.3 (1.6)	226.5 (0.9)

From Figure 3.3, the handoff decreased by 21.8% in policy 4 compared to policy 1. Even though we introduced various restrictions into the policy, the throughput per physician showed slight improvement under the third and fourth policies, where we reduced the maximum number of patients handled by a physician. This restriction on accepting new patients requires the physicians to evaluate and discharge the existing patients, thereby increasing the throughput. Moreover, in the second policy, where we restrict a physician from handling high acuity cases in

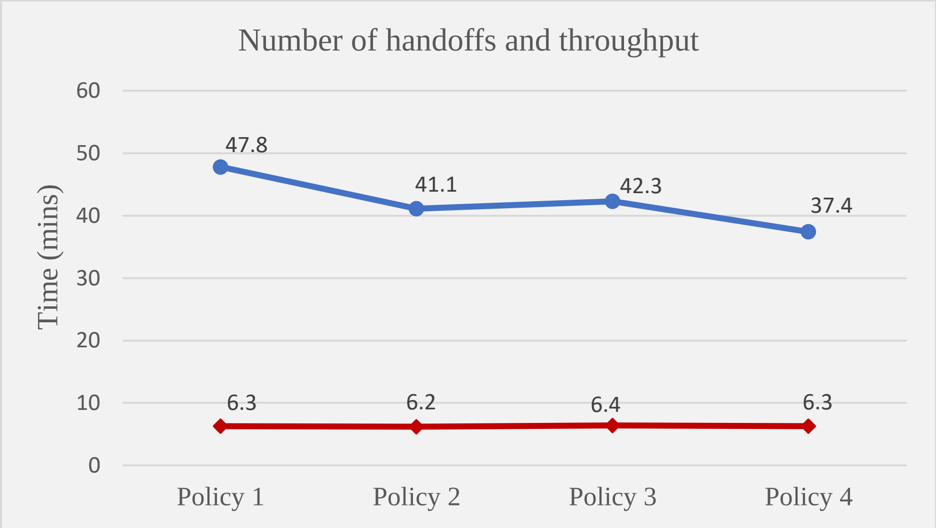
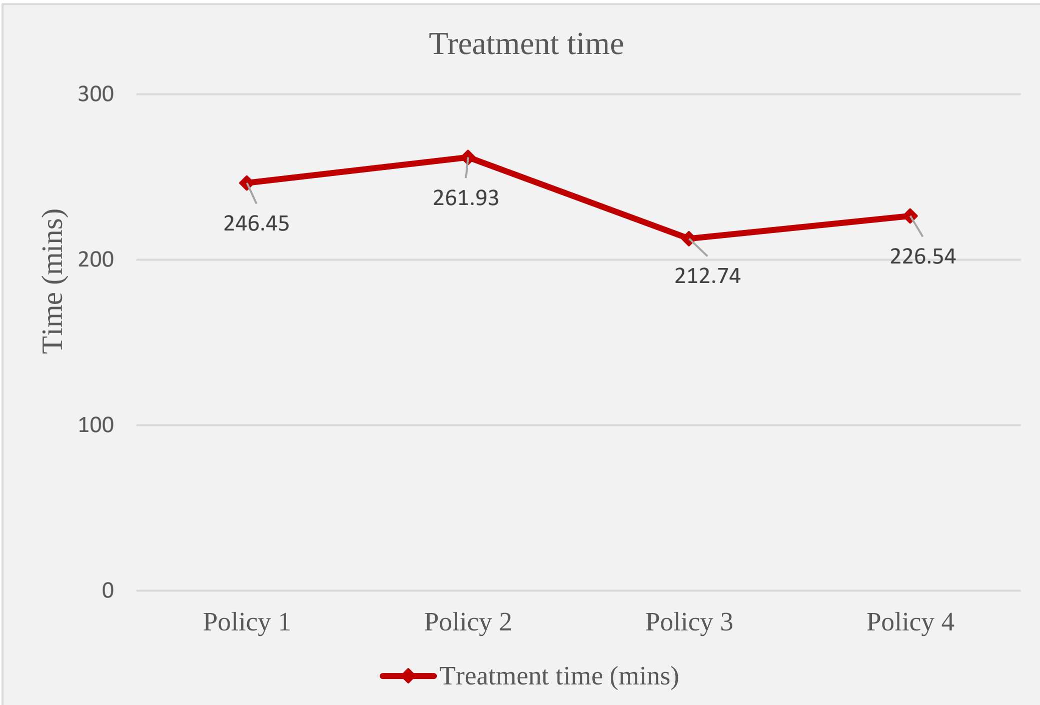


Figure 3.3: Average number of handoffs and throughput per physician.

the last 120 minutes, a reduction in the number of handoffs is observed. However, policy 4 shows the most reduction in handoffs where we restrict the maximum number of patients a physician can manage to 5 and further restrict it to four during the last 120 minutes where the physician handles only level 4 and level 5 patients.



These observations are also consistent with the performance measures in Figure 3.4, which display a decreasing treatment time. The treatment time under policy 4 decreased by 8.1% compared to the current policy. This is because the maximum number of patients a physician can manage is higher in the first two policies.

Figure 3.4: Average treatment time and length of stay for a patient.

The simulation modeling framework enabled the testing of multiple policies on patient care management, and findings from testing various restriction policies identified the potential for reducing handoffs in the ED by over 22% compared to the current practices [15]. This motivated us to the next step to build a simulation model representative of PRISMA Health GMH ED, including pods, patient arrivals, etc. Further, we also tested the impact of combining overlapping shifts and restriction policies on patient safety and patient flow.

3.3 Phase Two

3.3.1 Methods

3.3.1.1 Data

Input data for the model, including the number of beds, physician shifts, patient arrivals, ESI level of the patients, patient time in the ED, and the number of interactions between physicians and patients, were gathered from the partner ED. Additionally, observations were conducted in the ED, and the research team included ED physicians working in the partner ED for guidance, developing policies, and addressing any other physician-dependent activities in the ED to be included in the model. Our partner ED is the largest healthcare provider in the state and serves as a tertiary referral center for the entire Upstate region. The flagship academic Department of Emergency Medicine is an Adult Level 1 and Pediatric Level 2 Trauma Center, Stroke and ST-Elevation Myocardial Infarction (STEMI) Comprehensive Center seeing over 106,000 patients annually over four different pods.

We first introduce Figure 3.5, which represents the patient arrivals to the partner ED based on the day of the week. As seen in the image, the patient arrivals are low during the early hours and slowly start picking up from 7:00 am until 12:00 pm, when they reach the maximum and stay the same until 7:00 pm. This patient arrival trend is universal, and prior studies have reported the same [28,71]. Moreover, it can be noted from Figure 3.5 that weekdays have higher patient arrivals compared to the weekends, and Mondays have the highest patient arrivals. Although it would be ideal to develop a simulation model with an entire year of data, the variability among the patient arrivals for each month impacts the average time a patient spends in the ED, making it challenging for model validation. Hence, we created clusters of 3 months and used the cluster with the highest patient arrivals for this research (July '19 – September 19). Additionally, based on expert opinions from the ED physicians, we wanted to use the pre-COVID-19 data in our model, as these numbers were more representative of the ED patient arrivals. For the modeling patient arrivals, we consider

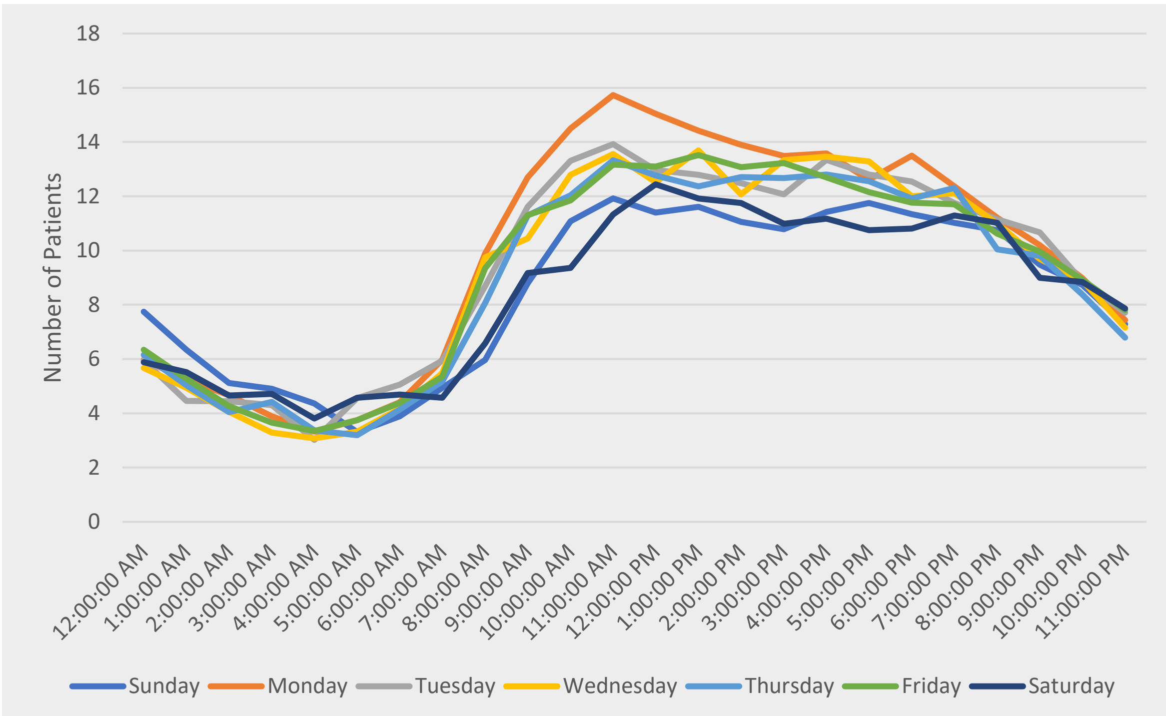


Figure 3.5: Hourly patient arrivals to the partner ED.

24-time slots for a day, each corresponding to an hour of the day. Based on the three months of data, the hourly patient arrival rate for each day of the week was modeled using a Poisson arrival.

Next, we introduce Figure 3.6, which represents the time a patient spends in the ED based on their ESI levels. We define time in the ED as the time a patient completes the registration and gets a bed in the ED until they leave the ED. As seen below, we split the data into two parts: "Bed to Disposition" and "Disposition to ED Departure." Bed to disposition represents the time for which a patient occupies an ED bed and is provided care by physicians and other medical providers, including performing tests, providing medicines, blood draws, etc. Although for some time, patients will be waiting in their beds during this period without receiving direct care while waiting for test results, etc. However, all these delays are related to patients' medical care. In general, this represents the period a patient first occupies a bed in the ED until the physicians make a disposition decision (admit, discharge or transfer). The second part, "Disposition to Departure," is the period during which a patient occupies the ED bed from the time the physician makes a disposition decision until they are physically moved from the ED (discharged, admitted, or transferred). Hence, these are logistical delays where a patient can be either waiting until a bed is available in the hospital (admission) or waiting for transportation (discharged or transfer). As seen in the figure, the disposition to discharge time for ESI-1 patients, which represents the most urgent patients, is the highest and higher than their bed to disposition time because most ESI-1 patients are admitted to the hospital. Hence, they have to wait in the ED until a bed is available. However, as the severity reduces, the disposition to departure time also reduces as most of the low-severity patients are discharged, and the delay we observe here is usually a result of patients waiting for transportation from ED. Finally, after a patient vacates an ED bed, a bed turnover time was included in the simulation model as the bed needs to be prepared and will not be immediately available for the next patient. This bed turnover time does not add to the patient time as the patient leaves the ED, and as observed in the ED, this delay is only on beds being unavailable for the next patient to occupy.

As mentioned earlier, the entire bed to disposition time of a patient is not spent with a physician as it includes other activities. Based on literature and discussions with ED physicians, we used between 15-30% of bed to disposition time as the care time where a patient would be cared for by a physician [114]. The percentages were assigned based on severity, such that the total time spent with an ESI-1 patient was the highest and that with an ESI 4 or 5 patient was the lowest, as seen in Table 3.4. This approach was used mainly for two reasons: lack of detailed visit-by-visit data available to support detailed modeling and to reduce the complexity of modeling individual delays and processes, which are beyond the control of ED administrators and the scope of this research. Here, TRIA represents the triangular distribution, a type of continuous probability distribution where TRIA (a,b,c) represents a distribution with a lower limit a, upper limit b, and mode c, where a < b and a ≤ c ≤ b. Further details on model development and validation are discussed in the later sections.

Severity	Bed to Dispo mi	% time with a physician	Visit 1	Visit 2	Visit 3	Visit 4
ESI 1	11				)	TRIA(2,3,4)
ESI 2	18				)	--
ESI 3	17				)	--
ESI 4	90					--
ESI 5	10					--

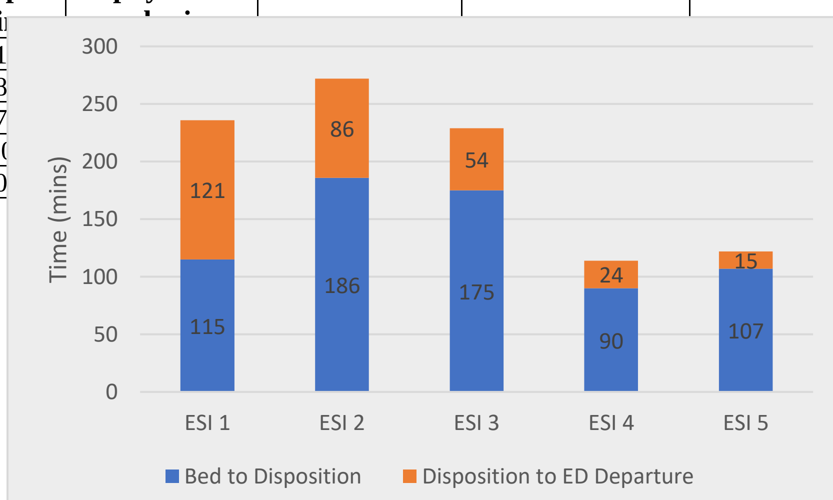


Figure 3.6: Patient time in the ED.

Table 3.4: Percent time a patient spends with a physician based on their assigned severity.

To build a model representative of ED operations where a physician visits patients multiple times based on their severity (ESI- level), we split the care time into multiple smaller windows. Based on our past observational studies and discussion with ED faculties and physicians, on average, an ES1-1 patient was visited four times by a physician, ESI-2 and 3 were visited three times, and ES1-4 and 5 were visited two times [115]. As represented in Table 5, the time spent with a patient for each visit based on their severity was provided to the model as unique distributions. Additionally, it can be noticed from Table 1 that for ESI-1, 4, and 5, the first visit is the longest. This is based on expert opinion and observation, as most of the time, the medical condition of these patients is recognizable during their first visit, and the physician can start providing care. However, in ESI 2 and 3 patients where medical conditions are not easily recognizable, the physician most likely orders a test and hence a lower time for the first visit. However, their second visit time is higher because the physician will start providing care and spend more time with the patients.

3.3.1.2 Model Development and Validation

As discussed in the background, prior studies have observed DES as one of the best methods to simulate an ED where various daily ED activities are modeled as a discrete sequence of events in time. In a traditional DES modeling approach used to model an ED, patients are considered agents that flow in the ED, each with unique attributes cared for by health care providers modeled as a resource. This traditional DES approach would suffice to address the issues at a high level, including bed planning and staffing requirements. However, to meet the aim of this research, which focuses on improving patient safety by minimizing the number of handoffs and identifying the impact of overlapping shifts on the patient flow, this approach would not incorporate the impact of the physician's decision-making capabilities based on current conditions in the ED. Hence, in this study, we used a novel approach where physicians are modeled as agents with unique parameters and abilities, allowing them to make informed decisions based on rules and policies practiced in the Greenville Memorial Hospital (GMH) ED.

Using this approach allowed replicating a physician's activity as realistic as observed in the ED, unlike the traditional DES approach where a patient would seize a physician resource just once for a particular amount of time and release them to move on to the following process. To provide further insight into the modeling approach adopted for this study, we introduce Figure 3.7 below, which captures the essence and capabilities of various ED physician activities that the model can simulate. In the figure below, dashed lines represent patients, and the solid lines represent the physicians moving in the ED. A patient arriving at the ED undergoes various onboarding processes (discussed in the next paragraph) before being assigned an ED room. Each room in the ED has a single bed that a patient will occupy from room assignment until the physician makes a disposition decision. Each arriving physician has an arrival time, shift end time, and pod assignment in the ED to provide medical care during their shift. Upon arrival to their specific pod, a physician goes to the physician station, and if another physician is leaving the same pod, the patients from the leaving physician are transferred to the new physician – that is, patient handoffs occur. If no physician leaves the same pod, the arriving physician starts assigning themselves new patients who are waiting in the ED without a room assignment.

The physician will also spend time in the station reviewing the patient's medical record before visiting each patient. When ready, the physician visits the patient in their room, with the time required depending on the patient's ESI level. Following the patient visit, the physician returns to their station to order tests, labs, and imaging as necessary while the patient waits in the bed for the requested tests. Secondary care, including labs, medicines, imaging, etc., are either performed while patients are on the bed, or in a few instances, patients might be rolled out of the ED, but the bed/room will not be assigned to another patient (based on observations in the ED). After the first visit with a patient, our approach links a physician and the specific patient based on their unique IDs. This ensures that the same physician will provide the subsequent care for the patient unless

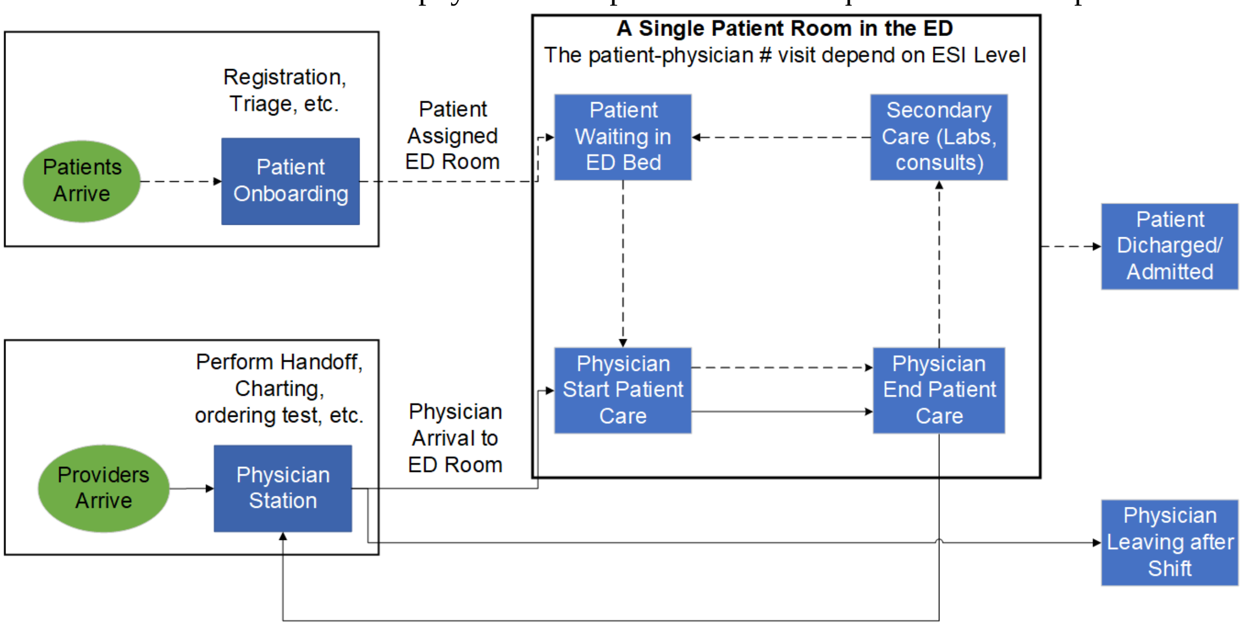


Figure 3.7: An agent-based approach for physician-patient interaction.

they are ending their shift and the patients are handed off to another physician. From a modeling standpoint, we have an array where each physician ID can handle multiple patient IDs, but each patient ID can link only to one physician at a time, thus replicating how an actual ED functions.

The number of subsequent visits and time spent with the patient during each visit again depends on the ESI level of the patient. Using the agent-based modeling approach for both patients and physicians overcomes the limitations of traditional DES, where physicians are modeled as resources and cannot make intelligent decisions. By modeling physicians as a resource, the decision-making capabilities are limited to patients where patients seize the physicians for a certain amount of time and release them, making the resource (physicians) available for the next patient. Further, a physician stays idle and doesn't flow in the ED and cannot make intelligent decisions limiting their ability to pick and choose patients based on their workload, time remaining in the shift, and pod limitations. However, modeling both patients and physicians as agents, as discussed in the research, allows both agents to make intelligent decisions based on rules replicating the actual ED activities, including charting, adding orders, handoffs, etc. Finally, our current modeling approach allows for the flexibility of continuous model development, especially when modeling secondary resources as they would act as independent activities.

Before beginning the model development, the first step was to create a process map of GMH ED to capture the day-to-day activities. Through observations and meetings with ED physicians, the research team developed a detailed process map using Microsoft Visio. Figure 3.8 below provides an overview of the ED activities for a single pod where patients arriving at ED are either triaged or sent directly to the trauma bay based on their medical condition. A majority of the patients are triaged, where they are assessed by a triage nurse and assigned an ESI level based on their medical condition. However, a few severe cases (e.g., car crashes, ST-Elevation Myocardial Infarction, etc.) might not be triaged and are provided care in the trauma bay. The triaged patients are then registered into the hospital's electronic health record and then directed to the waiting room, where they are prioritized based on their assigned severity level. When a bed is available in one of the ED pods, an ED nurse takes the patients from the waiting room based on their severity level and the capability of the ED pod. This is because certain pods in the ED do not have medical equipment and other capabilities to handle high-severity patients. In the figure, apart from the patient arrival activities (represented in the box) rest of the activities are specific for each pod, and each pod was modeled separately. The ED rooms and trauma bays are modeled as resources and divided into four pods where the capacities and capabilities of these resources are the same as in the GMH ED.

As seen in Figure 3.8, upon a physician's arrival for a shift in a specific pod, the physician who will be leaving the ED will transfer their patients to the arriving physician. As mentioned earlier, this process of transferring the care of a patient is defined as a handoff. In the absence of a physician in the station, the new physician will take a new patient and later meet the leaving physician for handoffs. These are usually rare because physicians leaving the ED do not tend to provide care during the last 15 minutes of the shift, as they would be focused on completing the patient charts. For the handoff process, we use a delay using a distribution based on the data collected from observations. In case no physician is leaving an ED pod, then there would be no handoffs, and the arriving physician would start taking new patients. Finally, in the case when a physician leaves the ED and a new physician is not arriving at the ED, which happens during night shifts, the leaving physician will handoff their remaining patients to the existing physician in the ED. Post handoff, the physician decides on taking a new patient depending on the current number of cases handled. The model ensures that the physicians working in the same pod simultaneously share the patient load equally. It should be noted that a physician's workload is considered balanced based on the number of patients they are providing care to and not necessarily based on the ESI level of the patient, as that is the practice followed in the ED.

As briefly discussed above, after accepting a new patient, the physician would then meet the patient in the ED room for the first evaluation and then returns to the physician station to document in the medical record, order test, labs, consults, and medicines as necessary. As the physician places the order, the nurse then completes the required documentation, the ordered tasks, medication administration, and runs bedside tests or ordered interventions. Additionally, patient care often includes diagnostic imaging that may require the patient to be moved out of the ED to the radiology suite or samples sent to the lab. These ancillary tasks are represented as a "black box" because these data are not inputted into the model. Following the drug administration, imaging, and diagnostic testing, the physician returns to the patient for the subsequent evaluation, and the physician provides care until the patient is clinically stabilized. After a subsequent patient visit, the physician might not necessarily return to the physician's station immediately. Hence, based on expert opinion, we used a 40% probability that the physician might visit another patient before returning to his or her station. After the final evaluation, the patient is either discharged or admitted as an inpatient to the hospital, and the physician may take on a new patient. During all the scenarios in the model, whenever a level 1 patient is presented in the ED, irrespective of all the policies and rules, the immediately available physician serves the patient.

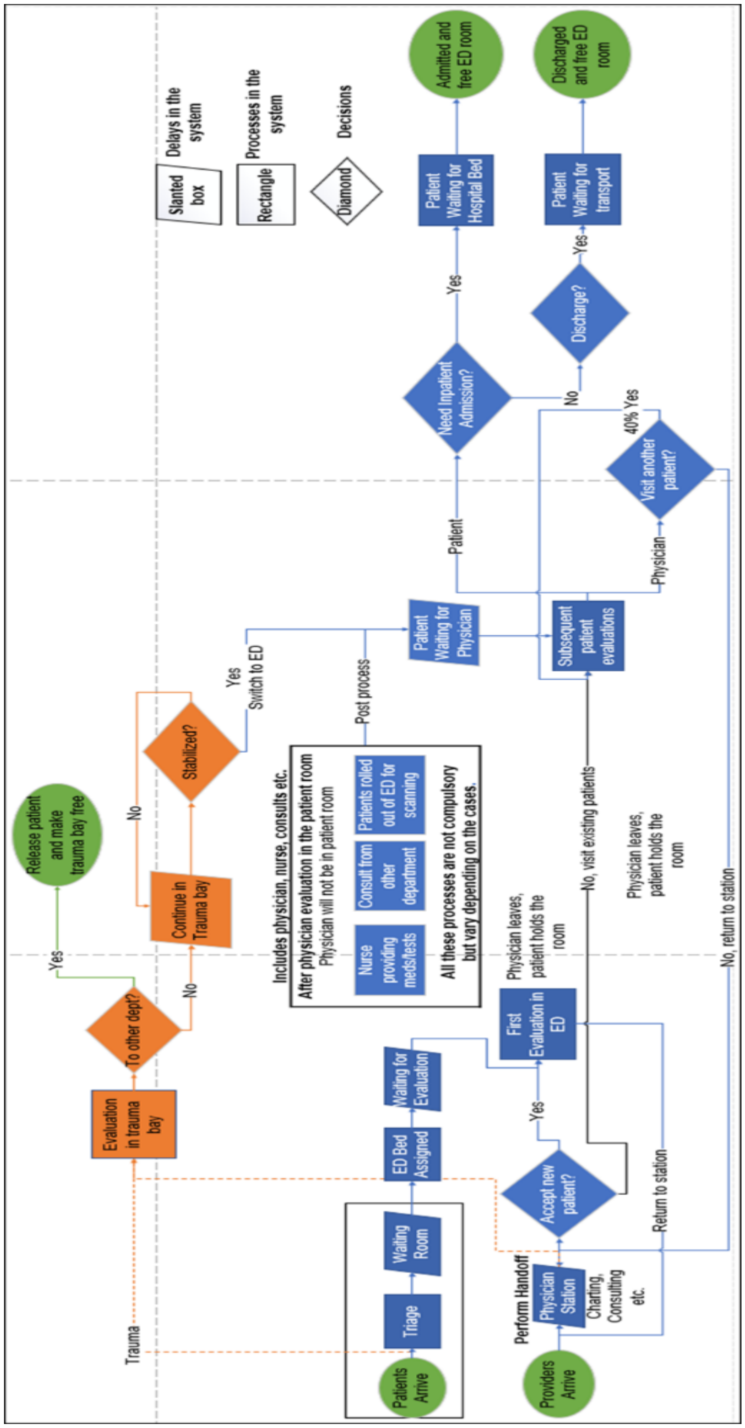


Figure 3.8: Partner ED patient flow process map.

While we did not consider nurses, consults, and ancillary resources as specific entities in the simulation model, because of interdependencies of different departments, the simulated time a patient spends in the system was validated against the actual time, which is discussed in detail later. Moreover, this approach allowed us to replicate physician activities and daily operations from an ED standpoint and investigate our primary aim of understanding the impact of overlapping shifts on the number of handoffs and patient flow without any restrictions.

After developing the model, the next step was to validate the model to ensure that the developed model replicates the partner ED. As mentioned in the prior sections, all the available data, including patient arrivals, ESI- level probability, physician shift schedule, patient visits based on severity, and the number of beds, were inputted into the model. However, for the "black box," which represents the ancillary activities (radiology, labs, etc.) and for the time between subsequent patient visits, the research team did not have detailed data and used a probability as discussed in the data section. Based on the expert opinion, which suggested a physician would spend between 15-30% of the bed to disposition with a patient and more time with high severity patients, the research team used different probabilities to calibrate it with actual data. Although the approach might not be intuitive, it helped in accounting for different delays in the ED without modeling all the different external processes. After calibrating the model, a physician's time with a patient for each visit was discussed with the ED physicians to ensure that these values were realistic. Based on the feedback, adjustments were made, and the model was recalibrated until the difference between the simulated time and actual time was less than 8%, as seen in

Table 3.5. Further, to make sure that the simulated time and actual group did not vary significantly, we performed a parametric and non-parametric statistical test. First, we conducted a t-test, and on comparing the simulated time to the actual time, we observed a p-value = 0.94, and on conducting the Mann-Whitney U-test, we observed a p-value = 0.92., suggesting that the simulated and actual time in the ED did not vary significantly. Further, the time a patient spends in the ED was also validated for each hour of the day, and we observed that the simulated time and actual time did not vary significantly (p-value = 0.87). Further, the simulated throughput and actual throughput values were also compared, and on performing a t-test, these values did not vary significantly (p-value - 0.90).

Severity	Actual Time in ED (mins)	Simulated Time in ED (mins)	Percentage Difference
ESI 1	236	218	-7.6%
ESI 2	272	281	3.3%
ESI 3	229	216	-5.7%
ESI 4	114	121	6.1%
ESI 5	122	122	0%

Table 3.5: Comparing actual time and simulated time.

3.3.1.3 Scenario Descriptions

Upon developing and validating the model, the next step was to test various physician shift scenarios to reduce the number of handoffs, improve patient safety and patient flow in the ED. However, to ensure that these new policies did not negatively affect other ED performance metrics, we utilized the following three metrics to evaluate the impact of each policy: number of handoffs, throughput, and patient time in ED. Here, the number of handoffs represents the total number of patients transferred between the physicians, throughput represents the number of patients leaving the ED, and time in the ED represents the time between first physician contact until patient discharge or admission to the hospital. A few of these measures were selected based on prior studies that used the same metrics to evaluate an ED's performance [112,113]. Next, we discuss two policies: non-overlapping and overlapping shifts, where four scenarios are tested under each policy, including the current policy used by the partner ED.

3.3.1.4 Non-Overlapping Shifts

3.3.1.4.1 Scenario 1 (Baseline)

This scenario represents the current physician staffing policy adopted by the partner ED for patient management. Table 3.6 and Figure 3.9. below represent the different shift slots currently used at the partner ED. It can be noted from the table that shifts are designed such that the end time of a shift is the same as the start time of the next shift. For example, a physician starting the shift at 7:00 am works for eight hours until 3:00 pm, and a new physician arrives at 3:00 pm. However, based on certain pods, there may be multiple physicians working at the same time. For example, while a physician starts their shift at 7:00 am, another physician can start their shift at 9:00 am. Under this scenario, a physician will not take any new patients during the last 15 minutes of their shift until it is an ESI-1 patient.

3.3.1.4.2 Scenario 2

This scenario is very similar to the first scenario but with an additional restriction. Here pods where multiple physicians are available, the physicians are restricted from signing up a new patient during the last 30 minutes of the shift. Additionally, if there is only a single physician working in the pod, then a new patient is not picked up by the physician during the last 15 minutes unless an ESI-1 patient. This approach was to investigate the impact of the restriction on handoffs.

3.3.1.4.3 Scenario 3

The difference between this scenario and the last scenario is that here pods where multiple physicians are available, the physicians are allowed to pick up only low severity patients (ESI-4,5) during the last 60 minutes of the shift. This specific restriction was placed as high acuity patients (ESI-1, 2 & 3) usually spend more time in the ED. Further, this restriction was placed as prior studies have proved that physicians' productivity decreases as the shift progress and increase the chances of errors [116,117].

3.3.1.4.4 Scenario 4

This scenario can be considered as a conservative version of scenario 2 where pods with multiple physicians are available; the physicians are restricted from signing up a new patient during the last 60 minutes of the shift. Here, a physician leaving the ED would focus on providing care to their existing patient and complete charting during the last hour.

Shift No	Non-overlapping shifts (8-hour)	Overlapping shifts (9-hour)
1 (baseline)	7:00 am – 3:00 pm	7:00 am – 4:00 pm
2	8:00 am – 4:00 pm	8:00 am – 5:00 pm
3	9:00 am – 5:00 pm	9:00 am – 6:00 pm
4	12:00 pm – 8:00 pm	12:00 pm – 9:00 pm
5	3:00 pm – 11:00 pm	3:00 pm – 12:00 pm
6	4:00 pm – 12:00 am	4:00 pm – 1:00 am
7	5:00 pm – 1:00 am	5:00 pm – 2:00 am
8	11:00 pm – 7:00 pm	11:00 pm – 8:00 pm

Table 3.6: Different physician shifts currently used in the partner ED.

3.3.1.5 Overlapping Shifts

Table 3.6 above represents the different shift slots available for an overlapping shift. It can be noted that the shift start times are the same as the non-overlapping shifts. However, the end time of the shift is increased by an hour, making this a 9-hour shift with a one-hour overlap. We

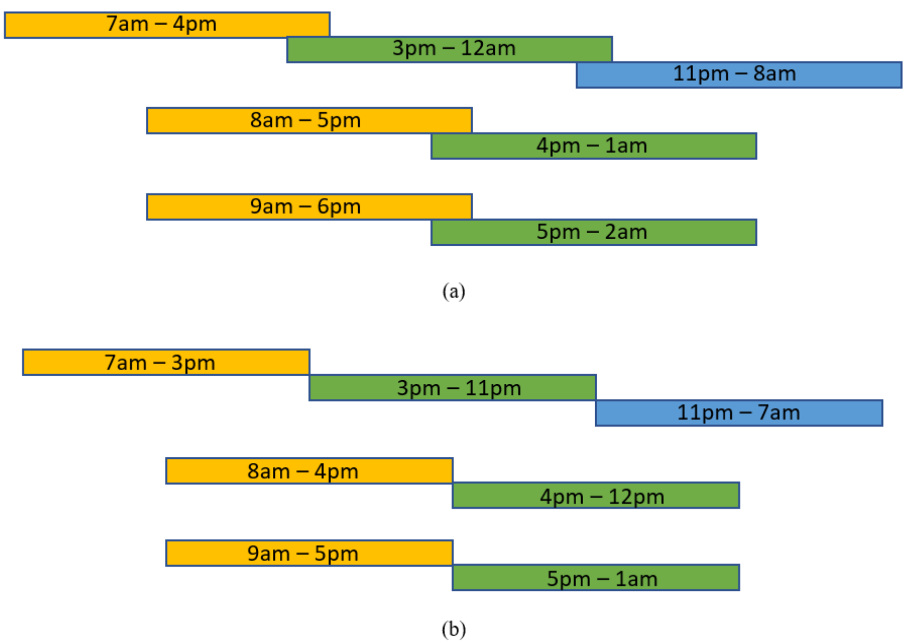


Figure 3.9: (a) One-hour overlapping shift schedules and (b) non-overlapping shift schedules.

utilized one-hour overlapping shifts after discussing with ED physicians and faculties as scheduling shifts with longer overlaps leads to a higher financial burden as a result of more physician hours. For example, a physician starting the shift at 7:00 am works for nine hours until 4:00 pm, and a new physician arrives at 3:00 pm, creating an hour of overlap.

3.3.1.5.1 Scenario 5

This scenario is the same as the current PRISMA health policy except for the fact that the shift length is increased to 9 hours with an hour of overlap. Here a physician will sign up a new patient until the last 15 minutes of their shift.

3.3.1.5.2 Scenario 6

Under this scenario, for pods where multiple physicians are available, the newly arriving physician assumes the primary role, and the leaving physician provides care for only ESI-4, 5 patients for the last one hour.

3.3.1.5.3 Scenario 7

This scenario is very similar to the prior scenario; however, irrespective of the number of physicians working in the pod, the newly arriving physician assumes the primary role, and the leaving physician provides care for only ESI-4, 5 patients for the last one hour.

3.3.1.5.4 Scenario 8

Under this scenario, for pods where multiple physicians are available, the newly arriving physician assumes the primary role, and the leaving physician is completely restricted from taking any new patients. Here the leaving physician will use the last hour of their shift to provide care for their existing patients and complete their charts.

3.3.2 Results

Each scenario discussed above was simulated over a three-week schedule with additional two days of warm-up period for ED to achieve an equilibrium. Further, the model was replicated 60 times to account for natural, random behavior such that the margin of error on time in the ED metric was ± 10 minutes (at $\alpha = 0.05$).

Table 3.7: Performance metrics for different scenarios.

Scenarios	Throughput (Patients per week)	Number of handoffs per day	Average Time in ED (mins)
Baseline, 1	1506±5.3	89±1.6	215±3.0
2	1508±5.8	85±1.8	217±3.1
3	1507±6.5	83±2.0* (p-value = 0.041)	216±2.6
4	1505±6.6	79±2.3* (p-value = 0.036)	222±3.9
5	1504±5.8	60±2.0** (p-value = 0.019)	178±1.5** (p-value = 0.009)
6	1505±5.3	57±1.9** (p-value = 0.013)	185±1.9** (p-value = 0.021)
7	1504±6.9	55±2.4** (p-value = 0.010	185±3.9** (p-value = 0.021)
8	1503±5.7	52±1.5** (p-value = 0.008)	185±1.8** (p-value = 0.021)

* = indicates significantly different from baseline
** = indicates significantly different from baseline and other non-overlapping scenarios.

3.3.2.1 Comparisons of restriction policies against baseline policy

The eight scenarios for physician scheduling were tested using the developed model, and the observed value for the performance metrics are recorded in Table 3.7 above. We first compare the scenarios under policy 1 (non-overlapping shifts), where scenario 1 (baseline) represents the current partner ED policy, and the other three represent new scenarios (2,3,4) that could be implemented in the ED. It can be noticed from the table that the throughput did not vary significantly between the three scenarios and the baseline scenario, and this was further verified by conducting a t-test (p-value > 0.05) where each scenario was compared against the baseline scenario. The throughput does not vary significantly for the different scenarios because we are modeling hourly patient arrival to the ED as a stationary Poisson process, so even if the patient flow improves, we have a limitation on the patient demand. Further, for these scenarios on comparing the time in ED metric to the baseline policy, it can be observed that these did not vary significantly (p-value > 0.05). These observations suggest that patient flow did not improve significantly with these scenarios of non-overlapping shifts. Finally, we compared the number of handoff metrics and observed that adding restrictions can reduce handoffs by as much as 11.2%. Although the number of handoffs reduced for all the non-overlapping scenarios compared to the baseline, on performing a t-test, we observed that only scenarios 3 and 4, where physicians were restricted from high severity patients or no patients during the last hour, showed a statistically significant (p-value < 0.05) reduction. Scenario 3 observed a 6.7% decrease in handoffs, and scenario 4 observed an 11.2% decrease in handoffs compared to the baseline scenario.

3.3.2.2 Overlapping vs. non-overlapping shifts

Next, we compare the overlapping policy to the baseline scenario and other non-overlapping scenarios. On comparing the throughput of overlapping scenarios to the non-overlapping scenarios, including the baseline scenarios, it can be observed that there is no significant (p-value > 0.05) difference. Again, this is because of the limited patient arrival data provided to the model. To investigate the impact of overlapping shifts on patient flow, we compared the time in ED metric of overlapping scenarios to the non-overlapping scenarios. On comparing, we observed that non-overlapping scenarios had a statistically significantly (p-value <0.05) lower time in the ED than the non-overlapping shifts. On average, overlapping scenarios reduced the time in ED by 15.7% compared to the non-overlapping scenarios. It is imperative that there would be a reduction in the patient's time in the ED with the overlapping shifts as the physician shift length is extended by 1-hour compared to the current practices. However, it can be observed that the reduction in patient time in the ED (15.7%) exceeds the expected reduction of 11.1% (1/9) by 4.6%, suggesting that this additional reduction could be because of the reduced number of handoffs. To verify this, we calculated the total time a physician spends performing handoffs under each policy. By reducing the handoffs under overlapping policies, we observed that the physicians spend less time performing handoffs during shifts, increasing their availability to provide care for patients by approximately 3.8%. Additionally, we observed that patients who require a handoff spent ten additional minutes on average in the system. These observations explain the additional gain achieved by the new policy and the potential to improve patient flow. Further, by comparing the number of handoffs that occurred during overlapping scenarios to the non-overlapping scenarios, we observed a statistically significant (p-value < 0.05) decrease in the number of handoffs.

3.3.2.3 Comparative results using queuing theory

To further evaluate if these reductions observed from the simulation model can be replicated using a mathematical model, we approach this problem using a queuing model. Since we aimed to get an approximate number of handoffs per day, we combined all types of patients (ESI 1-5), generated the Poisson distribution for the arrival process, and fit an exponential distribution to their time in the ED metric rather than segregating patients based on ESI level. Then using an $M/M/\infty$ queuing model, we calculated the number of handoffs for each physician shift to generate the number of handoffs per day. For the baseline scenario, the number of handoffs per day based on the queuing model was 91 compared to 89 from the simulation model output. Similarly, for the overlapping shift (scenario 4), which equated to the baseline policy except for the shift length, the number of handoffs using the queuing model was observed to be 63 compared to 60 from the simulation model output. These observations from the queuing model further validated the findings from the simulation model. Another critical observation was that the most impact of overlapping shifts on handoffs was observed during the peak hours of patient arrivals to the ED as more physicians were now available to provide patient care. On average, overlapping scenarios reduced the number of handoffs by 33.3% compared to the non-overlapping scenarios, suggesting improved patient safety. Figure 3.10 below represents the number of handoffs and time in ED for different scenarios.

3.3.2.4 Impact of varying patient arrivals

Finally, to comprehend the impact of non-overlapping and overlapping shifts on different patient arrivals to the ED, we performed a sensitivity analysis. We consider ten scenarios where the first scenario represents the current patient arrival to the GMH ED, and the other nine scenarios

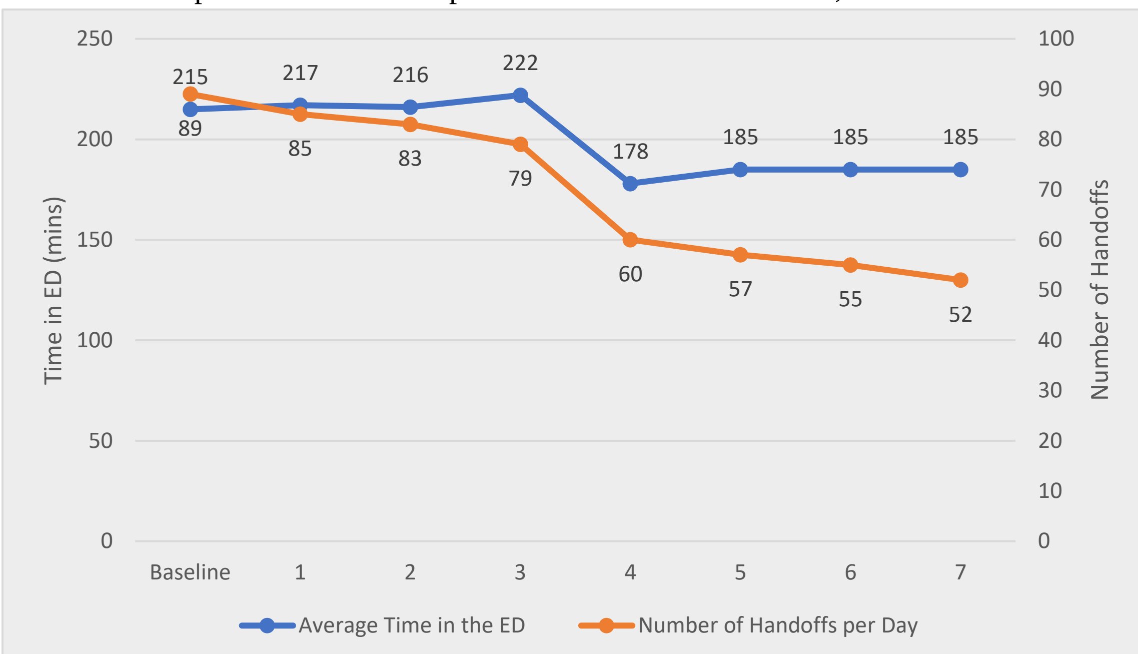


Figure 3.10: Number of handoffs and time in ED for different scenarios.

represent where the patient arrivals are increased by 5%. Our aim was to identify at what increment of patient arrivals the overlapping shift approximately equates to the current ED schedule performance and comprehend when the ED gets overloaded. Hence, we used the baseline policy for the non-overlapping policy and scenario 4 for the overlapping policy, which is the equivalent of the baseline policy except for the fact that it had an extra hour in the shift. Table 3.8 below represents the number of handoffs and time in ED for different levels of patient arrivals.

Table 3.8: Performance metrics for different patient arrivals.

Patient Arrival Scenarios	Number of handoffs per day		Average Time in ED (mins)		ED Wait Time (mins)	
	Non-overlapping	Overlapping	Non-overlapping	Overlapping	Non-overlapping	Overlapping
Baseline	89±1.6	60±2.0	215±3.0	178±1.5	29±1.1	17.1±0.9
5% increase	109±2.1	71±1.3	246±4.3	191±2.1	44±5.2	29±1.6
10% increase	130±0.9	83±1.6	283±4.6	205±2.6	66±9.2	43±2.3
15% increase	141±2.0	92±1.8	319±3.4	222±3.1	101±10.2	62±7.6
20% increase	151±1.0	106±1.3	331±1.7	234±2.2	226±53.4	72±3.8
25% increase	153±0.8	119±1.6	334±1.2	250±2.3	591±50.2	105±7.7
30% increase	153±1.7	122±2.0	335±1.4	263±2.4	944±53.8	203±40.5
35% increase	154±0.5	133±1.4	335±1.1	274±1.3	1351±51.2	371±50.2
40% increase	155±0.4	138±1.2	335±0.9	275±0.9	1642±46.2	670±51.4
45% increase	155±0.6	138±2.6	335±0.9	275±0.8	1885±41.6	925±71.2

To evaluate at what increment of patient arrivals the overlapping shift approximately equates to the current ED schedule performance, we analyze Table 3.8, Figure 3.11, and Figure 3.12. It can be observed that a 9-hour shift with a one-hour overlap can handle a 10-15% increase in arrivals and achieve the baseline performance based on the number of handoffs and Time in ED.

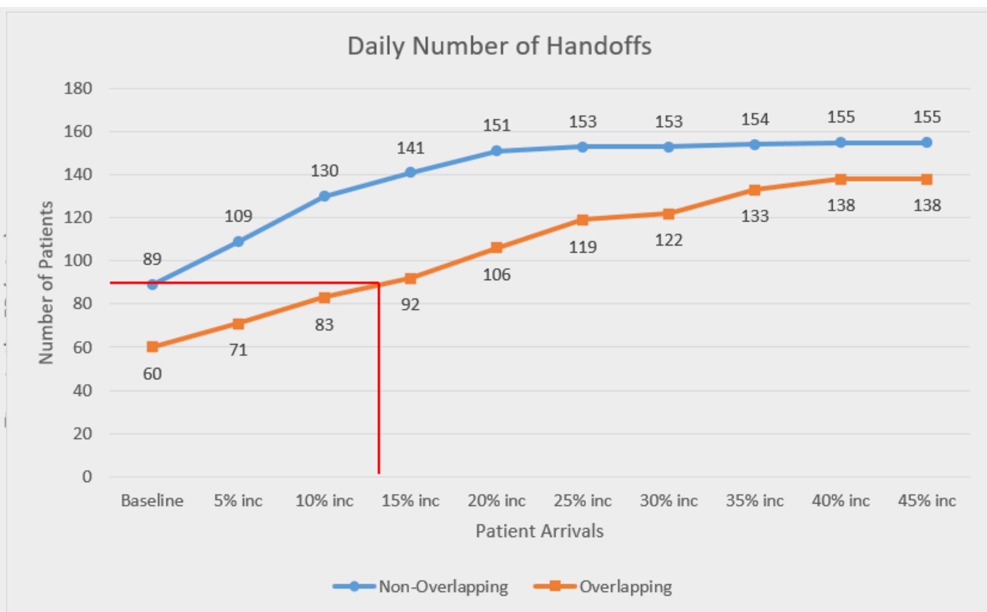


Figure 3.11: Number of handoffs for increasing patient arrivals.

Figure 3.12: Time in ED for increasing patient arrivals.

It can be observed from Figure 3.11, and Figure 3.12. that with both 8-hour non-overlapping shifts and 9-hour shifts with a one-hour overlap, the number of handoffs and time in the ED flattens after a certain period. However, it can be noticed that for non-overlapping shifts, the metrics start to flatten when the patient arrivals increase by 20%, whereas for the one-hour overlapping shifts, the flattening starts when patient arrivals increase by 35%. The flattening of the metrics suggests that after a specific increase in the patient arrivals, the ED gets overloaded, and irrespective of increasing patient arrivals, these patients have to wait to receive medical care because of resource constraints. However, this wait time will not be reflected in time in the ED metric or handoffs metric as they don't capture the wait time. Hence, we introduce Table 3.9, which represents the average patient wait times for a bed in the ED and weekly throughput, assuming that a patient arriving at the ED will not leave without getting served.

Table 3.9: Patient wait times and throughput for different patient arrivals.

Patient Arrival Scenarios	**Average Patient Wait Time (mins)**		**Weekly Throughput**	
	Non-overlapping	**Overlapping**	**Non-overlapping**	**Overlapping**
Baseline	29±1.6	16.1±0.9	1506±5.3	1504±5.8
5% increase	44±5.2	29±1.6	1577±9.1	1573±6.5
10% increase	68.9±9.2	43±2.3	1652±5.5	1651±7.2
15% increase	256±53.4	62±7.6	1684±11.4	1719±9.5
20% increase	591±50.2	72±3.8	1712±4.5	1778±6.9
25% increase	944±53.8	105±7.7	1715±5.1	1854±7.9
30% increase	1351±51.2	203±40.5	1722±15.2	1897±17.3
35% increase	1642±46.2	371±50.2	1733±4.5	1959±15.4
40% increase	1885±41.6	670±51.4	1739±4.5	1978±15.2
45% increase	2706±40.1	1055±71.2	1744±6.8	1980±20.1

From Table 3.9, it can be observed that as the patient arrival increases, the average patient wait time also starts increasing, which shows an overloading ED and the impact of resource constraints. Additionally, after a specific point, the weekly throughput begins to flatten, meaning that irrespective of increasing patient demands, the ED cannot keep up because of resource constraints. For the non-overlapping and overlapping scenario, the throughput doesn't change much after a 20% and 35% increase in patient arrivals suggesting that ED would get overloaded quickly under the non-overlapping scenario compared to the overlapping shift. The minor change in throughput after the inflection point can be attributed to the variation in ESI levels, i.e., with increasing patient arrivals, the number of patients in each ESI will increase, and physicians can dispose of patients who require less time in the ED who might be prioritized in the waiting room based on their ESI. It is evident from the analysis that a 9-hour shift with a one-hour overlap can significantly decrease the number of handoffs and time in ED and improve the ED's capability to handle more patients before overloading. However, there is an increase in full-time equivalents (FTEs) of physician staffing as a result of increased shift length.

3.3.2.5 Addressing the Secondary Resources

It is evident from the results that overlapping policies and restriction policies can improve patient safety and patient flow with a slight increase in FTEs. However, one of the key limitations is that secondary delays are not modeled separately in the simulation model. Although this does not affect the validation, it could be possible that the additional physician FTEs can inflate the ED performance as there are no secondary delays. From an ED standpoint, some of the most common secondary delays are radiology (medical imaging) and labs. Although there are other delays like consults, these are not frequent orders. This was further confirmed from the data analysis, where we observed that approximately 54% of patients arriving at the ED required imaging compared to only 8% requiring a consult order. Based on these observations and the lack of data regarding labs, we decided to model the radiology process into the simulation model. Using expert opinions from the ED physicians and available data points, we divide the radiology process into three steps: order to begin, begin to end, and end to read. The order to begin represents the time between the ED physician placing an order until the test is started, begin to end represents the time for performing the test as such, and finally, end to read represents the time between a test is completed and the results are read. The next step was to identify the population of patients that require radiology and those who do not need a radiology order. As patient severity level (ESI level) is one of the prominent factors influencing the radiology requirements, where it is a consensus that low severity patients (ESI 4 AND 5) would rarely require a radiology order, we decided to classify the data based on ESI levels. Table 3.10 represents the radiology requirements based on the ESI levels. It is evident from the table above that ESI levels have a significant impact on radiology requirements, with radiology orders increasing with patient severity. However, just looking at the radiology requirement is not sufficient for modeling because some patients require only a single imaging order, whereas a few others would require multiple imaging going as high as six orders.

Table 3.10: Radiology requirements based on ESI level.

ESI	**# Patients at ESI Level**	**Patient Distribution**	**% that Require Imaging**
1	877	3.22%	88%
2	6401	23.49%	64%
3	12588	46.20%	61%
4	5858	21.50%	36%
5	899	3.30%	5%

Moreover, on analyzing the data, we observed that although multiple orders are placed, if they are placed simultaneously, those were performed together and did not have a significant impact on the time required for the radiology process. In contrast, if orders are placed in a sequential order where the order time for the second order is after the end to read time of the previous, that has a significant impact on the process as the patient had to wait for the subsequent order to be completed. Further, from the modeling standpoint, if a patient required only an order where all radiology requests are placed at the same time, it meant that the radiology order happens after the first visit and before the second visit. If subsequent radiology orders are placed, it means that an additional visit was required with the physician. Thus, a patient with a single order would need two visits with the physician; two subsequent orders would need a total of 3 visits with the physician, and so on. However, based on discussion with physicians and for the sake of modeling, we decided to have a maximum of 4 visits as anything more than four orders were infrequent. Table 3.11 below represents the subsequent radiology order required based on the ESI level. After adding this information to the model, one final piece of distinction that was added to the model was classifying an order as Life or Death (LOD) or Routine. This was crucial as LOD orders that represented urgent orders required lesser order to begin time than routine orders.

Table 3.11: Subsequent radiology orders based on ESI level.

SI	**1 order**	**2 orders**	**3 orders or more**	**% Needing Imaging**
1	59%	22%	7%	88%
2	48%	13%	3%	64%
3	50%	10%	2%	61%
4	33%	3%	0%	36%
5	5%	0%	0%	5%

On investigating the LOD data, we observed that LOD orders were placed only for ESI 1 and 2 patients on analyzing the data. Table 3.12 below represents the time for various processes of radiology based on the ESI level. For LOD patients, the only difference was that they had only 10 mins of order to begin time.

Table 3.12: Radiology process time.

SI	**Order to Begin Time (min)**	**Begin to End Time (min)**	**End to Read (min)**
1	17	15	22
2	45	12	19
3	57	14	18
4	30	10	16
5	30	8	18

The next step was to validate the simulation model against the actual data. The patient's time with a physician and some other delays were adjusted from the previous model until the difference between the simulated data and actual data was less than 7%. Further, a t-test was performed to compare the simulated time to the actual time, and we observed a p-value = 0.96, suggesting that the simulated and actual time in the ED did not vary significantly. After validating the model, the next step was to investigate if the overlapping policies and restriction policies are effective in improving patient safety and patient flow in the ED. Figure 3.13 below represents the handoffs and time in the ED using the new simulation model. Further, Table 3.13 below represents compares the result from the updated model to the old model, which does not account for secondary delays.

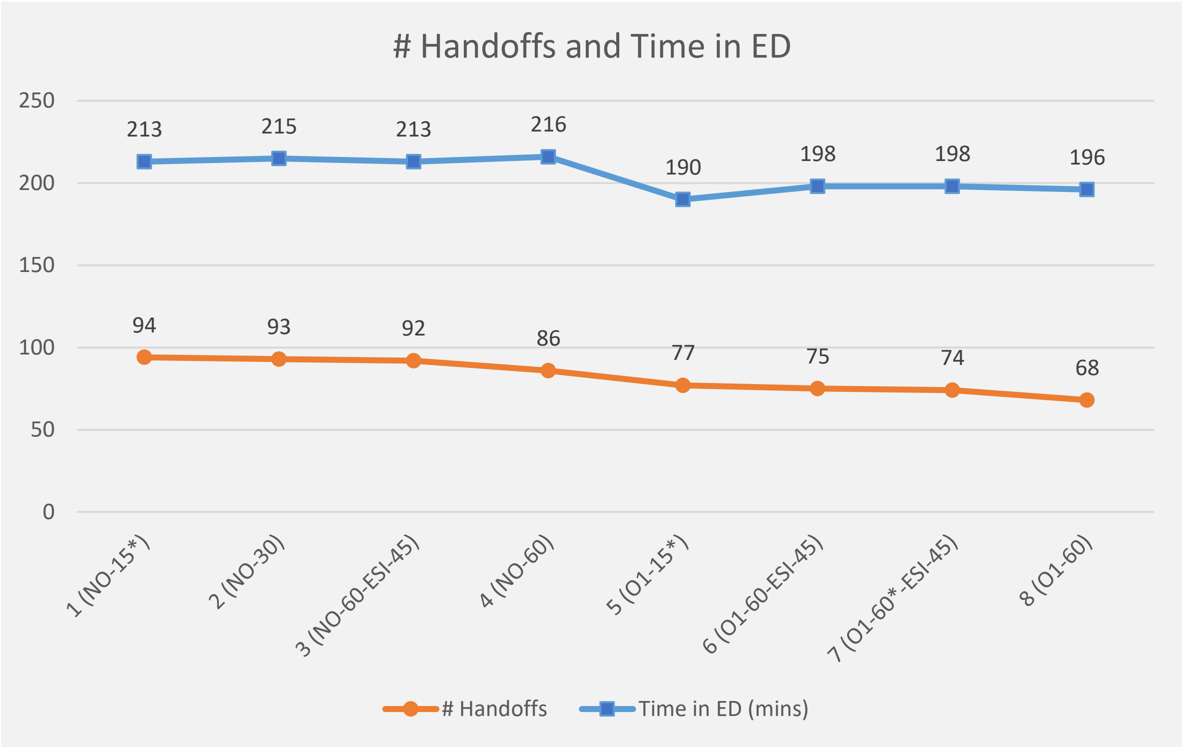


Figure 3.13: Number of handoffs and time in ED for different scenarios.

Table 3.13: Comparing overlapping to non-overlapping shifts.

Policy	Metrics	Model 1	Model 2
Baseline, NO-15	Handoffs	--	--

	Time in the ED	--	--
NO-30	Handoffs	-4%	-1%
	Time in the ED	1%	1%
NO-60, 45	Handoffs	-6%	-3%
	Time in the ED	0%	0%
N0-60	Handoffs	-11%	-8%
	Time in the ED	3%	2%
O-15	Handoffs	-33%	-18%
	Time in the ED	-17%	-11%
0-60, 45	Handoffs	-36%	-20%
	Time in the ED	-14%	-7%
0-60*, 45	Handoffs	-38%	-21%
	Time in the ED	-14%	-7%
0-60	Handoffs	-41%	-28%
	Time in the ED	-14%	-8%

It can be observed from the table above that compared to the non-overlapping baseline policy, overlapping policies and restriction policies improved the patient flow and patient safety in the ED. However, as suspected, adding the secondary delays reduced the effectiveness of the overlapping and restriction policies in improving the performance metrics. Another interesting observation was that the reduction associated with overlapping policies is higher compared to non-overlapping policies because, in the prior model, we were not modeling any secondary delays as a result of imaging consults, etc., and adding a resource (extra physician hours) had a direct inflated impact on the performance metric. However, in the updated model, the impact of additional resources and physician availability is limited as there are ED activities that are physician-independent which will still cause delays in the system.

3.3.3 Discussions and Conclusions

Transitions of patient care from one physician to another are major risk points that potentially lead to patient harm, especially in a care setting like the ED. Hence, it is critical that ED administrators consider the patient safety metric of handoffs along with other patient flow metrics while developing shift schedules to improve ED performance. This research focused on identifying ED physician shift schedules that reduce handoffs while not negatively affecting the patient flow and other CMS core performance metrics, including throughput, patient time in the ED, etc. To address this research objective, we developed and validated a simulation model that replicates the partner ED and tested different scenarios under non-overlapping and one-hour overlapping policies, including policies that restricted physicians from taking either all or high severity patients during the end of the shift [118,119].

We observed that restricting physicians from taking high-severity patients during the last hour of the shift could significantly reduce the number of handoffs without negatively affecting other performance metrics. Additionally, scenarios that restricted the physicians from taking both high and low-severity patients during the end of the shift were further able to reduce the number of handoffs without significantly increasing the patient time in the ED. These findings are similar to the findings from prior studies where restricting physicians during the end of the shift can significantly reduce the number of handoffs with a non-significant increase in time in ED [15,111]. However, the percentage decrease in handoffs varies across the studies, as it depends on patient volumes, ESI level, number of pods in the ED, current shift schedule, and hours for which the restriction is placed. On testing a nine-hour shift with one hour overlap, we observed that it could reduce the handoffs and reduce the time in ED significantly but with additional cost due to physician FTEs. Combining the overlapping policy with various restriction policies, we observed that handoffs and patient time in the ED could be reduced as much as 41.5% and 14%, showing potential for significant improvement in ED performance. From a different point of view, if current performance provides a level of patient care considered acceptable, the overlapping policy suggests that the ED can immediately accommodate a 10-15% increase in patient volume.

We observed that implementing policies that only restrict a physician from taking either all or high severity patients during the end of the shift can improve patient safety by reducing the number of handoffs. However, it should be noted that these restriction policies result in a slight (statistically insignificant) increase in the patient time in the ED as physicians are restricted from taking patients during the end of their shift. Hence, we recommend that smaller EDs with fewer patient arrivals or large EDs with higher staffing levels could improve patient safety by implementing restriction policies as they could potentially reduce handoffs. However, each ED should set the restriction time (30 minutes, 60 minutes, etc.) based on their ED demand requirements, as an extended physician restriction period could further increase the patient time in the ED. Unlike restriction policies, utilizing a 9-hour shift with a one-hour overlap can potentially reduce both patient handoffs and time in the ED. However, we observed the most significant improvement in ED performance from both patient safety and patient flow standpoint on combining the restriction policy with the one-hour overlapping. Hence for larger EDs and Level 1 trauma centers, although restriction policies could help in improving patient safety, we recommend a combination of restriction and overlapping shift policies to have the most impact. Although it is evident that overlapping policies and a combination of overlapping and restriction policies can significantly reduce the number of handoffs, the patient time in ED, and handle more patients, it incurs an additional cost due to extra physician hours. Hence, respective EDs must perform a risk-cost-benefit analysis to balance additional physician staff hours and expected patient arrivals such that the policies better fit their patient demands and resource availability. However, a long-term solution for EDs is to develop staffing schedules using a mathematical model capable of finding exact solutions based on key performance metrics (time in the ED, handoffs, etc.) and constraints (budget, physician preference on the length of shifts, and shift start times, etc.). This approach would help in physician shifts tailored for each ED based on their constraints and physician preferences, ensuring clinician and administration satisfaction.

While this modeling approach allows the testing of various policies on performance metrics of handoffs and ED time, it cannot identify the "optimal" number of staff required for a given week based on historical patient arrivals. A logical next step in our research plan is to identify a static staffing plan that is optimized based on these key metrics and then is later validated through testing in the detailed simulation model. While our model representation accounts for estimates of the total time required for processing items such as radiology and labs, our current research does not account for specific delays resulting from these ancillary departments to the ED. Future research and modeling will also include an examination of patient-level and physician-level impacts of these resources that are external to the ED.

4. Chapter 4

4.1 Optimal Staffing for Improving Patient Safety and Patient Flow in ED

4.1.1 Introduction

Emergency departments (EDs) act as one of the primary patient care access points for millions of people seeking medical care. The ever-increasing volume of patient arrivals and varying severity among cases makes ED one of the most complex healthcare settings and prone to crowding [3]. Crowding is well-recognized public health and patient safety issue that has been explored over the last few decades, which occurs when the patient demand for emergency care exceeds the resources available in the ED to provide care in an acceptable time period [4,120]. ED crowding has a negative impact on patients, providers, and health systems where it leads to reduced quality of care, poor patient outcomes, increased medical errors, higher patient mortality, increased stress and burnout among providers, and increased healthcare costs [9,10]. Despite the public awareness and significant efforts by researchers/government agencies, crowding still plagues EDs across the globe and has risen over the past several years [23].

Although various reasons contribute to ED crowding, one of the primary reasons leading to ED crowding within the US is the overwhelming increase in patient arrivals to the ED, which has increased by 24%, and the decrease in the number of EDs, which declined by 15% over the last decade [3,22]. This directly leads to a mismatch, and to worsen the scenario, most EDs are overwhelmed with a lack of provider availability and dynamic planning tools for maintaining adequate resources. Additionally, studies have attributed ED crowding to poor ED design and/or inefficient patient flow, often led by a lack of ED beds, inadequate staffing levels, and limited inpatient hospital beds [10,26]. Although lack of inpatient beds is a primary cause leading to bottlenecks in the ED, these are shared resources and are often affected by factors beyond the control of ED administrators and stakeholders. However, factors including inadequate staffing and bed shortages in the ED can be avoided through better planning. Although the easiest and quickest solution to address these issues would be adding extra resources, including beds, staff, and ancillary units, adding new hospital resources could be very expensive. Moreover, researchers have observed that rather than adding physical resources (e.g., bed, equipment, machines, etc.), temporarily adding or changing staff schedules are comparatively cheaper options. However, generating a new schedule is not trivial as factors including provider preference, hospital budget, resource restrictions, etc., should be considered carefully such that they can improve patient flow and patient safety in the ED without overstaffing.

Among prior studies that have focused on developing ED staffing schedules, most studies have accounted for patient flow factor and budget, but to our knowledge, none of these studies have accounted for a patient safety factor [121,122]. Patient safety is an integral aspect of the ED as it functions 24*7 for 365 days and interacts with multiple departments making it prone to errors. Recent studies have observed ED as one of the hospital departments with high error rates. Some of the common sources of ED errors are interruptions, miscommunications, and loss of information. Handoffs, transfer of a patient's care and responsibility from one physician to another, are fraught with miscommunications, omissions, errors, and information loss [85,86]. However, handoffs are unavoidable in EDs as they operate throughout the day, and a physician ending their shift is required to transfer their current patients to the newly arriving physician. Although unavoidable, handoffs should be minimized, as it is a significant patient safety concern.

A recent study where thirty-six ED physicians were shadowed for over 100 hours observed that a physician's likelihood of making an error while prescribing was significantly higher when interrupted [87]. Similarly, studies have observed that approximately 80% of serious medical errors involve miscommunication during the patient handoff [88]. Additionally, poor handoffs, which involve miscommunication, can lead to conflicting expectations for information and contribute to delayed patient onboarding and conditions that can pose safety threats [89]. Further, studies that specifically investigated ED shift-change handoffs observed that for approximately 75% of the patients, the vital signs were not communicated, and errors were observed in about 60% of cases [90]. Finally, insurance claims involving missed ED diagnoses that harmed patients reported that 24% of the cases involved inadequate handoffs [91].

As mentioned earlier, to our knowledge, none of the prior studies have considered patient safety metrics as a performance indicator of the ED while generating staffing schedules. This research developed a Mixed Integer Linear Programming (MILP) model to generate staffing schedules that can improve patient safety and patient flow while accounting for ED budgeting and not affecting other Centers for Medicare & Medicaid Services (CMS) core metrics.

4.1.2 Background and Literature

The contribution of operations research models and methodologies has had a significant impact on improving EDs throughout the world. A variety of approaches, including mathematical and optimization models, queuing theory, simulation modeling, and probabilistic models, have been used to address a variety of ED issues, including resource allocation, patient streaming, fast track ED, staffing, and scheduling, etc. Although various tools have been used to improve ED operations, researchers have endorsed simulation models as one of the best tools to model different phases of patient flow (arrival to departure) in the ED because of the complexity and nature of ED [92,93]. Specifically, researchers have identified discrete event simulation (DES) to be efficacious in representing and simulating ED activities [94,95]. Additionally, the ability of simulation tools to model different ED processes, phases of patient flow, and test "what-if" scenarios make it an essential tool to investigate staffing and scheduling, resource allocation, and overall process improvement before implementing changes.

Although simulation models allow for testing various staffing policies and scenarios to design and evaluate the ED physician shift schedules, these models are not capable of identifying the optimal staffing levels. A mathematical model can address this issue by formulating the problem with a specific objective, constraints, and parameters representing the system to generate an optimal solution. Over the last few years, various studies have used mathematical models to identify optimal staffing levels, generate schedules, optimal beds, other resource requirements, etc. [121,123–129]. Among these, queuing theory has been extensively used for determining staffing levels as it allows the evaluation of patient flow criteria such as the waiting time of patients, throughput, etc. [130,131]. However, using analytical formulas for generating optimal staffing schedules have several limitations, usually with complex systems, specifically in the case of ED replicating the multiple patient-physician interactions and accounting for the time-dependent stochastic arrival rates. However, numerical approximations can solve this problem where studies have used integrated queuing and optimization model to investigate the effect of time-varying arrival rates for staff scheduling [132]. Additionally, researchers have used heuristics for generating schedules in ED [125]. Although heuristic allows for generating a quicker solution, it cannot guarantee an optimal solution. In contrast, a mathematical programming approach guarantees an optimal solution. However, as the number of variables and constraints becomes large, the process of identifying the optimal solutions become time-consuming. Hence for large problems, researchers have usually integrated mathematical programming with other methods, including genetic algorithms [133].

Prior studies have used various mathematical programming, including integer programming and multi-objective goal programming, and achieved optimal solutions for staff scheduling problems in the ED [121,134–136]. However, except for one study which focused on physician scheduling in the ED, others focused on nurse scheduling. For the one study focusing on the physicians, the researchers used a CART analysis to estimate the various patient level parameters and then developed a mixed-integer linear programming (MILP) model to minimize understaffing with respect to patient volumes. In an effort to replicate the actual process in the ED, the researchers divided a patient visit to the ED into multiple patient-physician interactions. The findings from the study were implemented and resulted in significant improvements in different ED performance metrics, including median length of stay, door-to-provider time, and door-to-bed time [121]. Further, researchers have used a combination of simulation-optimization models to identify optimal solutions and test them in the simulation model for validating the optimal solutions [122,137]. However, all these studies have focused on identifying solutions that can improve patient flow or improve ED performance by reducing the waiting time and length of stay or improving the ED throughput. To our knowledge, none of the studies using mathematical or simulation model approaches have used patient safety as an ED performance metric. In this chapter, we develop a MILP model for identifying optimal shift schedules that minimize the combined cost of patient wait times, handoffs, and physician shifts, thus considering the patient flow, patient safety, and staffing budget to generate schedules. Additionally, these new staffing policies are tested in the validated simulation model to evaluate their effectiveness.

4.2 Phase One

4.2.1 Methods

4.2.1.1 Data

Input data for the model, including the number of beds, allowable physician shifts, patient arrivals, ESI level of the patients, patient time in the ED, and the number of interactions between physicians and patients, were gathered from the PRISMA Health Greenville Memorial Hospital (GMH), Greenville, SC. Additionally, observations were conducted in the GMH ED, and the research team included ED physicians working in GMH, SC for guidance and addressing any other physician-dependent activities in the ED to be included in the model. PRISMA Health is the largest healthcare provider in South Carolina and serves as a tertiary referral center for the entire Upstate region. The flagship GMH academic Department of Emergency Medicine is an Adult Level 1 Trauma Center seeing over 106,000 patients annually.

We first introduce Figure 4.1, which represents the patient arrivals to the GMH ED utilized in our model. As seen in the image, the patient arrivals are low during the early hours and slowly start picking up from 7:00 am until 12:00 pm when they reach the maximum and stay the same until 7:00 pm. This patient arrival trend is universal, and prior studies have reported the same [28,71].

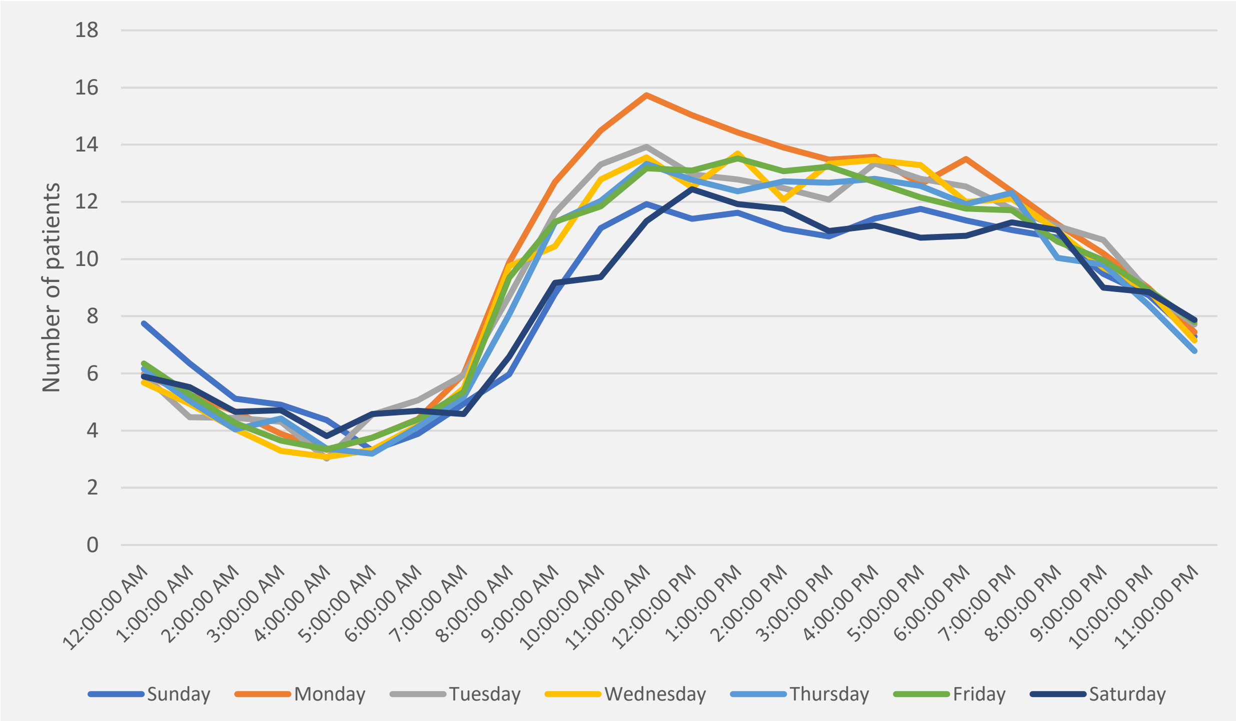


Figure 4.1: Patient arrivals to the GMH ED.

Moreover, it can be noted from Figure 4.1 that weekdays have higher patient arrivals compared to the weekends, and Mondays have the highest patient arrivals. Rather than using an entire year of patient arrival and using it for physician scheduling, we created clusters of 3 months and used the cluster with the highest patient arrivals for this research (July 2019 – September 2019). Additionally, based on expert opinions from the ED physicians, we wanted to use the pre-COVID-19 data as the patient arrivals varied significantly. Another reason for using this specific time period was to test the optimal schedule in our validated simulation model that used the same patient arrivals. However, the model was developed such that any patient arrivals can be used to generate a weekly schedule.

Next, we introduce Figure 4.2, which represents the time a patient spends in the ED based on their ESI levels. As seen below, we split the data into two parts: "Bed to Disposition" and "Disposition to ED Departure." Bed to disposition represents the time for which a patient occupies an ED bed and is provided care by physicians and other medical providers, including performing tests, providing medicines, blood draws, etc. Although for some time, patients will be waiting in their beds during this period without receiving direct care while waiting for test results, medicines, etc. However, all these delays are related to patients' medical care. In general, this represents the period a patient first occupies a bed in the ED until the physicians make a disposition decision (admit, discharge or transfer). The second part, "Disposition to Departure," is the period for which a patient occupies the ED bed from the time the physician makes a disposition decision until they are physically moved from the ED (discharged, admitted, or transferred). Hence, these are logistical delays where a patient can be either waiting until a bed is available in the hospital (admission) or waiting for transportation (discharged or transfer). As seen in the figure, the disposition to discharge time for ESI-1 patients, which represents the most urgent patients, is the highest and higher than their bed to disposition time because most ESI-1 patients are admitted to the hospital. Hence, they have to wait in the ED until a bed is available. However, as the severity reduces, the disposition to departure time also reduces as most of the low-severity patients are discharged, and the delay we observe here is usually a result of patients waiting for transportation from ED.

As mentioned earlier, the entire bed to disposition time of a patient is not spent with a physician as it includes other activities. Based on literature and discussions with ED physicians, we used between 15-30% of total time as the care time where a patient would be cared for by a physician [114]. The percentages were assigned based on severity, such that the total time spent with an ESI-1 patient was the highest and that with an ESI-5 patient was the lowest. This approach was used mainly used because of the lack of detailed visit-by-visit data available to support detailed modeling.

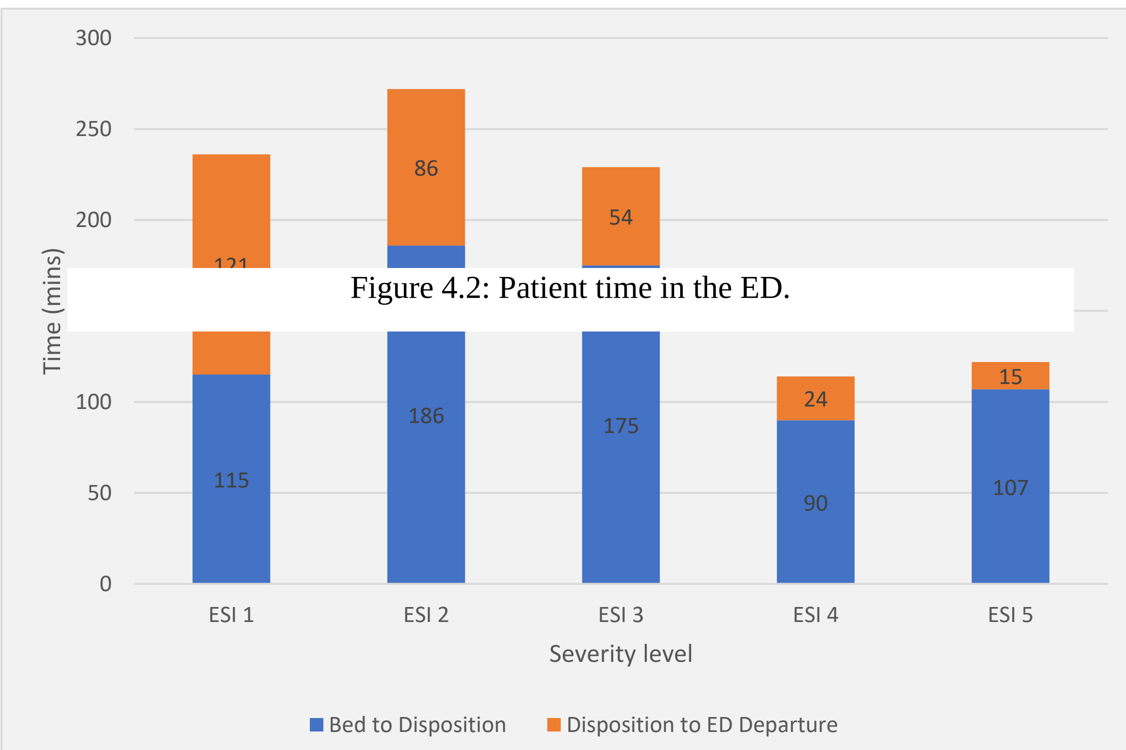


Figure 4.2: Patient time in the ED.

Further, to build a model representative of ED operations where a physician visits patients multiple times based on their severity (ESI- level), we split the care time into multiple smaller windows. Based on our past observational studies and discussion with ED faculties and physicians, on average, an ES1-1 patient was visited four times by a physician, ESI-2 and 3 were visited three times, and ES1-4 and 5 were visited two times. The physician's time with a patient for each visit was a constant time block of 15 minutes as the MILP modeling approach considers time as a discrete block of events.

4.2.1.2 Mathematical Model Development

We formulate the ED physician shift scheduling problem as a MILP problem. The primary goal is to identify the optimal staffing levels of ED physicians such that the patient onboarding time, waiting time after ED admission, and patient handoffs are minimized while considering the physician staffing cost to avoid overstaffing. To compare the impact of each factor, we identified dollar amount as the common scale. Based on the literature and expert opinions, costs pertinent to each factor were included.

Before formulating the problem, we first listed the key ED operational activities and processes that should be considered to develop an implementable MILP model. The first was accounting for the varying patient arrivals to the ED, including patient ESI levels. The second was modeling multiple patient-physician interactions based on the patient's ESI, accounting for minimum delays between patient-physician interactions to allow for secondary care (imaging, blood draws, etc.), ensuring that the same physician provides care for the patient unless the physician ends their shift (handoff), physician shift length is limited to 8 hours. Accounting for these operational activities, we next define the notation used in the MILP model.

4.2.1.2.1 Notations

We first introduce the sets and indices considered in this optimization model. The model included four sets and corresponding indices as follows:

- I represents the set of patient arrivals to the GMH ED indexed by i.
- K represents the set of possible physicians that can be staffed for a day indexed by k.
- T represents the set of time slots considered for staff scheduling indexed by t.
- M represents the set of physician visits required by a patient indexed by m.

Here, set I include all the unique patient arrivals to the GMH ED for a week, which totals more than 1500. Set K consists of the unique physician identification number that can start an ED shift for a day with an upper threshold of 25 physicians per day. Further, T represents timeslots for an entire week (which varies based on slot length). Finally, set M includes values from 1 through 4, representing the patient interaction with a physician. Next, we introduce the parameters considered in the model. Most of the parameters represent various patient characteristics, including severity, arrival time, physician visits, and fixed time slots that should be avoided for calculating patient wait time as these delays are inherent and one parameter defining the ED bed capacity.

- α_i represents the time slot of arrival for patient i.
- β_i represents the severity level of patient i.
- γ_i represents the total number of visits required by patient i.
- w_i represents the total time slots for patient i that should not be considered for waiting cost as this represents the minimum delay for secondary care.
- C represents the total bed capacity of the GMH ED.

Finally, we introduce the decision variables in the model:

- $U_{ik} = \begin{cases} 1, \text{ If patient i served by physician k} \\ 0, \text{ otherwise} \end{cases}$

- $Ystart_{kt} = \begin{cases} 1, \text{ If physician k starts their shift at time slot t} \\ 0, \text{ otherwise} \end{cases}$

- $Y_{kt} = \begin{cases} 1, \text{ If physician k is available for service at time slot t} \\ 0, \text{ otherwise} \end{cases}$

- $X_{iktm} = \begin{cases} 1, \text{ If patient i is served by physician k at time slot t for their visit m} \\ 0, \text{ otherwise} \end{cases}$

In our initial model formulation to best capture the ED activities discussed above, specifically, the multiple patient-physician interactions, we consider the time slot to be 10 mins meaning the t will be indexed over 144-time slots for a day. Additionally, in this stage, we also considered another notation, J, which represents the set of four pods in the GMH ED indexed by j. Although this allows for granularity in terms of operations, modeling the indices and iterations made it a hard problem to solve. In the next iteration, we increased the length of the timeslot to 30 mins which reduced the number of time slots for a day to 48. Although the indices for time slots decreased significantly, we were not able to generate a weekly schedule even after reformulations. After discussing with ED stakeholders and staff scheduling methods, we agreed on dropping the pod index as this only resulted in a drawback where physicians/schedulers had to manually assign the shift to different pods based on the desired coverages. For example, certain pods in GMH ED require double coverage (minimum of two physicians), whereas pods with smaller capacities can be managed with single coverage (single physician). However, this was not a huge drawback as physicians, and the other stakeholders had a clear understanding of the limitations of certain pods and the staffing requirements. Based on these inputs, we increased the time slots to 1 hour, meaning that a day would have only 24-time slots. This reduced the model computation time significantly compared to the previous scenarios. In our first formulation, which assumed a 10-minute block and other details, the model was not able to solve even after 20,000 secs (~6 hours), whereas the final formulation allowed the model to be solved in 393 secs. Next, we present the current formulation used for developing the ED physician staffing schedule.

In the formulation presented below, the objective function (1) minimizes the cost of staffing the ED physicians, handoffs, patient onboarding, and patient waiting time in the ED. The cost of staffing an ED physician (*SC*) was based using the national average rate for ED physicians, and the onboarding cost (*OC*) for patients based on their ESI level was derived from the literature [138,139]. However, because of the lack of data on the cost of patient waiting once admitted, we used a factor value (*F*) between 0 and 1 and multiplied it by the *OC* to calculate the waiting cost. Finally, for the handoff cost (*HC*), we used high values ($1,000) to avoid any possible handoffs.

Minimize:

$$\mathrm{SC}^* \sum_{kt} \mathrm{Ystart}_{kt} + \mathrm{OC}^* \sum_{ikt} t^*\mathrm{X}_{ikt1} - \alpha_i + \mathrm{OC}^*\mathrm{F}^* \sum_{ikt} (t^*\mathrm{X}_{ikt\gamma_i} - t^*\mathrm{X}_{ikt1} - w_i) + \mathrm{HC}^* \sum_{ik} \mathrm{U}_{ik}$$

Subject to:

$$\sum_{kt} t^*\mathrm{X}_{ikt1} \geq \alpha_i \qquad \forall i \in I$$

$$\sum_{ktm} \mathrm{X}_{iktm} = \gamma_i \qquad \forall i \in I$$

$$\sum_{km} \mathrm{X}_{iktm} \leq 2 \quad \forall i \in I, \forall t \in T$$

$$\sum_{kt} \mathrm{X}_{iktm} = 1 \quad \forall i \in I, \forall m \in M$$

$$\sum_{ikm} \mathrm{X}_{iktm} \leq C \qquad \forall t \in T$$

$$\sum_{kt} t^*\mathrm{X}_{iktm} \leq \sum_{kt} t^*\mathrm{X}_{iktm+1} \qquad \forall i \in I$$

$$\sum_{mt} \mathrm{X}_{iktm} \leq 4^*\mathrm{U}_{ik} \qquad \forall i \in I, \forall k \in K$$

$$\sum_{im} \mathrm{X}_{iktm} \leq 4^*\mathrm{Y}_{kt} \qquad \forall k \in K, \forall t \in T$$

$$\sum_{kt} \mathrm{Ystrt}_{kt} \leq 1 \qquad \forall k \in K$$

$$\sum_{kt} \mathrm{Ystrt}_{kt} \leq K$$

$$8^*\mathrm{Ystart}_{kt} \leq \sum_{q=t}^{\mathrm{Min}(168,\, t+7)} \mathrm{Y}_{kq} \quad \forall k \in K, \forall q \in T$$

$$\mathrm{U}_{ik}, \mathrm{Ystart}_{kt}, \mathrm{Y}_{kt}, \mathrm{X}_{iktm} \in \{0, 1\}$$

In the formulation, the first constraint ensures that a patient is served their first visit only after their arrival at the ED. The second constraint ensures that the patient is provided with all their required visits before discharging. As mentioned earlier, each hour represents a time slot, but from observations and discussions with physicians, we assume that a physician can visit four patients in an hour. However, the same patient cannot be visited four times during an hour as that is not realistic as patients wait to get their tests, imaging, radiology, etc., completed. The third constraint ensures that, at maximum, a patient can be visited only twice by a physician in an hour. The fourth constraint assures that each visit *m* for a patient cannot exceed 1, making sure that each visit is completed fully during a physician visit. The next constraint ensures that at any given time *t* the patients served cannot exceed the ED bed capacity. As patients have multiple interactions with physicians during an ED stay, these visits must be ordered such that a later visit follows the prior visit in terms of time slot, and our sixth constraint ensures the visits are ordered. The next two constraints assure that a patient can be visited a maximum of four times by a physician, and a physician can visit up to four patients during any given time slot (1-hour block). The next two constraints ensure that a physician starts their shift only once a day, and the total number of physicians staffed per day does not exceed the maximum possible physicians that can work for a day based on health system budget constraints. To ensure that a physician shift, once started lasts for eight hours, we use the second to the last constraint. Finally, the last constraint defines the variable types, which are all binary in this case.

Formulating the problem as discussed above allowed us to replicate an actual emergency department scenario where patients interact with physicians multiple times, wait for tests between visits, and, more importantly, account for patient care handoffs that impact patient safety.

4.2.1.3 Simulation Model

After developing the mathematical model to generate staffing schedules, the next step was to develop and validate a simulation model representative of the PRISMA Health ED. We utilized a novel hybrid modeling approach to develop the discrete event simulation where both patients and physicians are represented as agents with unique attributes. This approach allowed us to simulate the actual patient arrivals to the PRISMA health ED with specific features, including severity level, arrival time, etc. Moreover, the main reason to adopt this modeling methodology was to replicate the physician activities in the ED in a realistic manner, including starting a shift at a particular time, spending time in their workstation ordering tests, updating patient records, visiting patients multiple times, and finally handing off a patient to the next physician when their shift ends. These activities would have been challenging to include if we followed a traditional modeling approach where physicians are denoted simply as resources.

Figure 4.3 below provides a high-level overview of patient flow and physician activities in the ED for a single pod. The patient arrival to the ED and ESI assignments upon patient arrival during the triaging process was based on the historical data. After triage, the patient is assigned a bed if available, and in case no ED beds are available, the patient waits in the waiting room, where the patients are prioritized based on the assigned severity level. The second block of arrivals represents physicians arriving at a specific pod in the ED at their assigned shift starting time. A physician, upon arrival, goes to the physician station, and in case another physician is leaving the ED at the same time, the patients from the leaving physician are transferred to the arriving physician representing the handoffs as observed in the ED. After patient handoffs, the physician spends time in the station going through the patient charts and starts visiting the patients in their beds as necessary.

Further, whenever there are free beds in the ED, a physician, based on their workload, will sign up a patient from the waiting room and meet them in their bed (or room).To replicate the actual patient assignment process followed at PRISMA Health ED, physicians working in certain

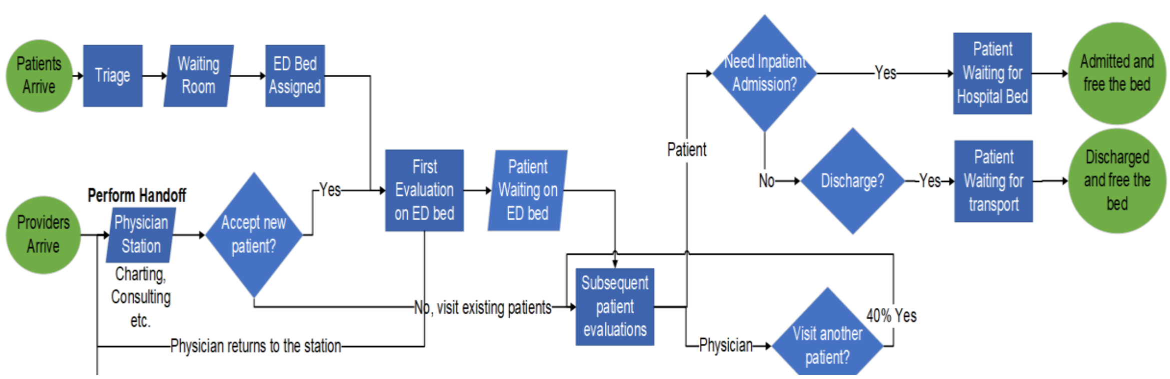


Figure 4.3: A high-level overview of patient and physician activities in a single ED pod.

pods were restricted from taking high severity patients as few pods do not have the equipment required to provide care for high severity patients. After visiting a patient for the first time, a physician always returns to the station to update charts and order tests. The patient will have subsequent visits by the same physician based on ESI level, as observed in the ED. If the particular physician is ending their shift, the patient is handed off to another active physician. Additionally, it can be noted from the figure that after a subsequent visit, there is a 40% chance that a physician visits another patient before returning to the station. Where historical data did not exist, expert opinions from ED physicians were used for modeling.

After developing the simulation model representative of PRISMA health ED, the next step was validating the model against the actual data. For this, we utilized the patient time in the ED for the ESI level as the validation metric to ensure that patient time spent in the simulation model did not vary significantly from the actual data for each ESI level. The model was simulated for a three-week schedule with an additional two-day warm-up period for the model to attain equilibrium. A total of 60 replications were performed, such that the margin of error on time in the ED metric was ± 10 minutes (at $\alpha=0.05$). Table 4.1 below represents the simulated data and actual data for each ESI level. It can be observed that the difference between these was less than 7% for each ESI level. Further, on conducting an independent t-test, there was no significant difference (p-value = 0.96) between the simulated data and actual data.

Severity	**Actual Time in ED (mins)**	**Simulated Time in ED (mins)**	**Percent Difference**
ESI 1	236±16.8	250±8.6	6%
ESI 2	272±12.4	272±6.6	0%
ESI 3	229±9.6	231±4.2	1%
ESI 4	114±8.9	117±5.4	3%
ESI 5	122±7.1	123±5.5	1%

Table 4.1: Simulation model validation (with secondary delays).

4.2.2

4.2.3 Results

To comprehend staffing schedules that can minimize patient handoffs, physician shifts, and patient wait times while considering the staffing budget, we specifically generated two policies.

- Policy 1: This policy aims to minimize the combined costs of handoff, patient waiting, and physician staffing using the MILP model based on the costs discussed in the formulation section.

- Policy 2: This policy also aims to minimize the combined costs of handoff, patient waiting, and physician staffing using the MILP model. However, here the handoff costs are penalized with a 3x multiplier of the original handoff cost with the central focus of eliminating handoffs as much as possible.

In the case of policy 2, handoff reduction might come at the cost of additional staffing as this policy aims to eliminate as much as handoffs as possible. However, we utilized an upper threshold on the number of physicians that can be staffed in the ED for a day. Weekly physician staffing schedules for both policies were generated such that a MipGap of < 3% was attained. Figure 4.4 below represents the average hourly patient arrivals and physician availability for the week under the two generated schedules and the baseline policy (current practices at the partner ED).

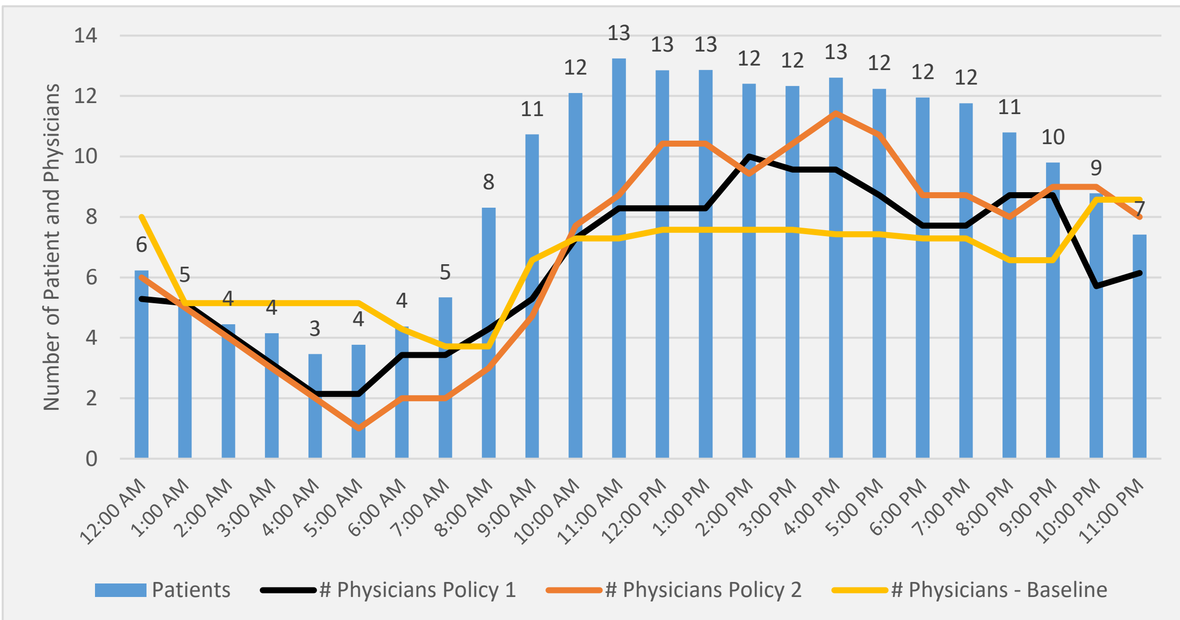


Figure 4.4: Hourly patient and physician availability in the ED.

From Figure 4.4 above, it can be clearly identified how the schedules generated using the MILP model staff the ED compared to the baseline policy. The baseline policy aims to maintain a steady physician availability throughout the day, with more physicians during the peak hours (8:00 am - 9:00 pm) and fewer physicians during the non-peak hours. However, both MILP models staff the ED in a dynamic manner considering the patient arrivals with a comparatively higher physician availability during the peak hours and lesser physicians during the non-peak hours. Table 4.2, represents the physician shift start times for the week, and it can be noted that compared to the baseline policy, the other two policies staff more physicians during peak hours of patient arrivals.

Table 4.2: Weekly physician shift start times.

Time	Baseline	Policy 1	Policy 2
12:00 AM	0	8	7
1:00 AM	0	0	0
2:00 AM	0	0	0
3:00 AM	0	0	0
4:00 AM	0	0	0
5:00 AM	0	0	0
6:00 AM	0	9	7
7:00 AM	22	7	0
8:00 AM	4	14	14
9:00 AM	20	7	12
10:00 AM	5	14	21
11:00 AM	0	7	7
12:00 PM	2	0	12
1:00 PM	0	0	0
2:00 PM	0	21	0
3:00 PM	21	4	7
4:00 PM	4	14	21
5:00 PM	20	1	7
6:00 PM	0	7	7
7:00 PM	0	7	7
8:00 PM	0	7	7
9:00 PM	0	0	7
10:00 PM	15	0	0
11:00 PM	21	7	0
Total Shifts	134	134	143
Weekly Hours	1128	1072	1144
Change in hours		-5%	1%

Additionally, it can be observed that the two new policies use an overlapping approach to start shifts rather than starting most of the shifts at the same time. For example, in the baseline policy, most physicians start their shift at 7:00 am, 9:00 am, 5:00 pm, and 11:00 pm, whereas the shifts are staggered for the other two policies. Further, it can be observed that policy 2 staffs more physicians as here handoffs are penalized significantly higher than the first policy. However, considering only the total number of shifts does not capture staffing cost, as some shifts are longer than 8 hours in the baseline policy. To ensure that handoffs are not minimized by overstaffing the ED, we compared the full-time equivalents (FTEs) under two new policies to the baseline policy. Although policy 1 reduced the FTE requirements by 5%, the FTE requirements increased by 1.4% under policy 2.

After generating the schedules, the next step was to test the two new policies along with the baseline policy in the validated simulation model. We used three ED performance metrics to compare the model performance: throughput, patient time in the ED, and the number of handoffs. The first two metrics evaluate the patient flow, and the third metric evaluates patient safety. All three policies were simulated in the model for a three-week schedule and replicated until the margin of error on time in the ED metric was ± 10 minutes (at α=0.05). From Table 4.3 below, it can be observed that both the new policies outperform the baseline policies. To comprehend if these differences were statistically significant, we conducted an independent ANOVA and observed that weekly throughput did not vary significantly among the three policies (p-value >0.05). It is imperative that the throughput will not vary significantly as the simulation model uses historical data and with a limited patient arrival. However, both handoffs per day and patient time in the ED were not the same (p-value = 0.03) for the three policies, suggesting a significant difference among at least one of the policies. To identify which groups varied significantly, we performed a Tukey posthoc test and observed that the number of handoffs in policy 2 varied significantly (p-value < 0.05) from baseline policy and policy 1. Additionally, patient time in the ED varied significantly between policy 2 and the baseline policy. Compared to the baseline policy, policy 1 reduced patient time in the ED by 2.5% and handoffs by 5.2%. Further, policy 2 reduced patient time in the ED by 6.4% and handoffs by 12.0%. Finally, in terms of FTE reduction, although we did not perform a statistical test, policy 1 reduced the FTE requirements by 56 hours (~1.5 FTEs) for a week, and FTE requirements increased for policy 2 by 16 hours (~.4 FTEs).

Table 4.3: Simulation model results.

Policy	Weekly Throughput	# handoffs per day	Time in the ED (mins)	Change in hours/ week
Baseline	1505	93	213	0
Policy 1	1503	88	207	-56
Policy 2	1506	81	199	+16

Finally, comparing policies 1 and 2, we observed that policy 2 improves patient safety and patient flow the best. However, it should be noted that the additional 4% reduction in patient time in the ED and 7% decrease in handoffs comes at the cost of ~2 additional FTE requirements for a week [140].

4.3 Phase Two

4.3.1 Methods

Although these initial findings are promising, the reduction in patient time in the ED and the number of handoffs add a cost burden on the system. Hence our next step was to identify areas of opportunities to improve these performance metrics without significantly increasing the FTEs. The results from our overlapping shift schedules suggest that staggering physician shifts with one-hour overlap can help in reducing the number of handoffs and patient time in the ED. However, this could also lead to additional FTEs, and more importantly, this would require ED physicians to work an additional hour. To understand the physicians' willingness to extend the shift, their perceptions of handoffs, and identify how overlapping could reduce handoffs: a) we decided to deploy a survey among all practicing attending physicians in the partner ED and b) analyze ED data to identify the pattern of handoffs.

After discussing with ED stakeholders, an online 7-question survey was developed and distributed to attending EM physicians via the email listserv using the Qualtrics survey tool. The survey was voluntary, anonymous, and gathered the physician perception of handoffs on patient safety, patient flow, patient satisfaction, preference for the number of handoffs, willingness to extend shift, and strategies to manage handoffs using multiple-choice, multiple-answer, and open-ended question styles. Survey questions were created by a senior attending physician along with a professor from the Dept. of Industrial Engineering and were vetted by other attending physicians. Participants completed the survey between the months of June - December of 2021. For the retrospective chart review, we utilized three years of data (Sep 2018 - Aug 2021) which included all patient arrivals for the respective years along with patient characteristics, including their severity (ESI) level, chief complaint, arrival time, admit time, disposition time, departure time, unique provider identifier, longest provider, etc.

4.3.2 Results

A total of 84 responses were collected with a 70% response rate from 120 attending ED physicians, and descriptive statistics were used to analyze multiple-choice and other responses to the survey response. Survey questions to understand the physician perception of handoffs on patient safety, patient flow, and patient satisfaction had a seven-point Likert scale which translated to positive, negative, and no impact. Figure 4.5 below represents the physician's perception of how handoffs impact various performance metrics, and it can be observed that 69% of physicians felt that handoffs have a negative impact on patient safety, and 67% felt that handoffs increased the patient length of stay (negative impact) and 56% felt that handoffs reduce patient satisfaction. For both patient length of stay and satisfaction, 25% and 31% of physicians felt that handoffs did not have any impact and 6% felt the same for patient safety. Consistent with the literature, a few physicians (25%) felt that handoffs actually improved patient safety, and 8% and 13% of physicians perceived that handoffs reduced patient length of stay and increased satisfaction.

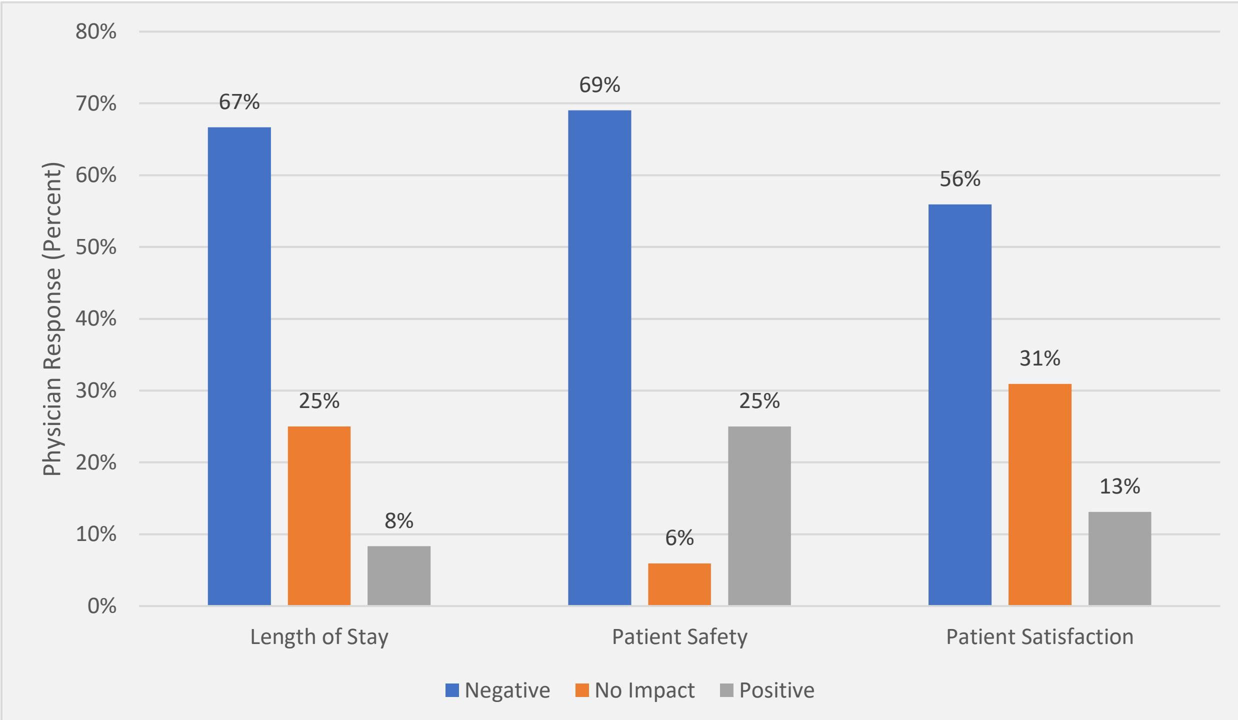


Figure 4.5: Physician perception of the impact of handoffs on length of stay, safety, and satisfaction.

To delve deeper and understand the physician's mental model for these perceptions, we analyzed the free-text response. A thematic content analysis was performed for the open-ended questions, which included narrative responses. The primary reason why physicians felt that handoffs affected patient safety negatively was the higher chances of confusion with change in the care team and missing information. In contrast, physicians perceiving handoffs to impact patient safety positively reported that change in the care team could improve patient safety (double-checking). Looking at the reasoning for increasing patient length of stay, physicians felt that handoff patients add a burden (workload) to oncoming physicians, and these patients may receive less attention as the oncoming physician might focus on new patients. The few physicians who felt that handoffs could reduce patient length of stay did not provide any particular reasoning.

However, those who responded that handoffs would not have any impact on patient length of stay reported that the oncoming physician would provide care as effective as them and would not have any impact on the length of stay. Finally, looking at patient satisfaction, physicians who perceived that handoffs decrease patient satisfaction reported that patients felt worried/concerned when physicians mentioned handoffs during the end of shift rounds. The physicians who perceived handoffs to have a positive impact on patient satisfaction reported that they discussed the care plan thoroughly with patients during the final round-up. Finally, those perceiving handoffs to have no impact on patient satisfaction reported that patients are responsive to the fact that handoffs are unavoidable in the ED. Although the physician perceptions of the impact of handoffs on patient safety, length of stay, and satisfaction were mixed, the majority of the physicians reported handoffs to have a negative impact on these performance metrics.

To understand physicians’ preference for the number of patients they handoff at the end of their shift and receive during the beginning of the shift, we analyzed the response to those two specific questions. Figure 4.6 below represents the physician's preference on the number of handed off and received during the beginning of the shift. It can be observed that, given the opportunity, 51% of physicians prefer not to hand off any patients at the end of their shift. Additionally, 37% and 10% of physicians responded that they prefer to hand off 1-2 and 3-4 patients during the end of their shift. Similarly, for the number of patients received during the beginning of their shift, 52% of physicians reported that they prefer not to receive any patients. Further, 31%, 13%, and 4% of physicians responded that they prefer to receive 1-2, 3-4, and 5-6% of patients during the end of their shift.

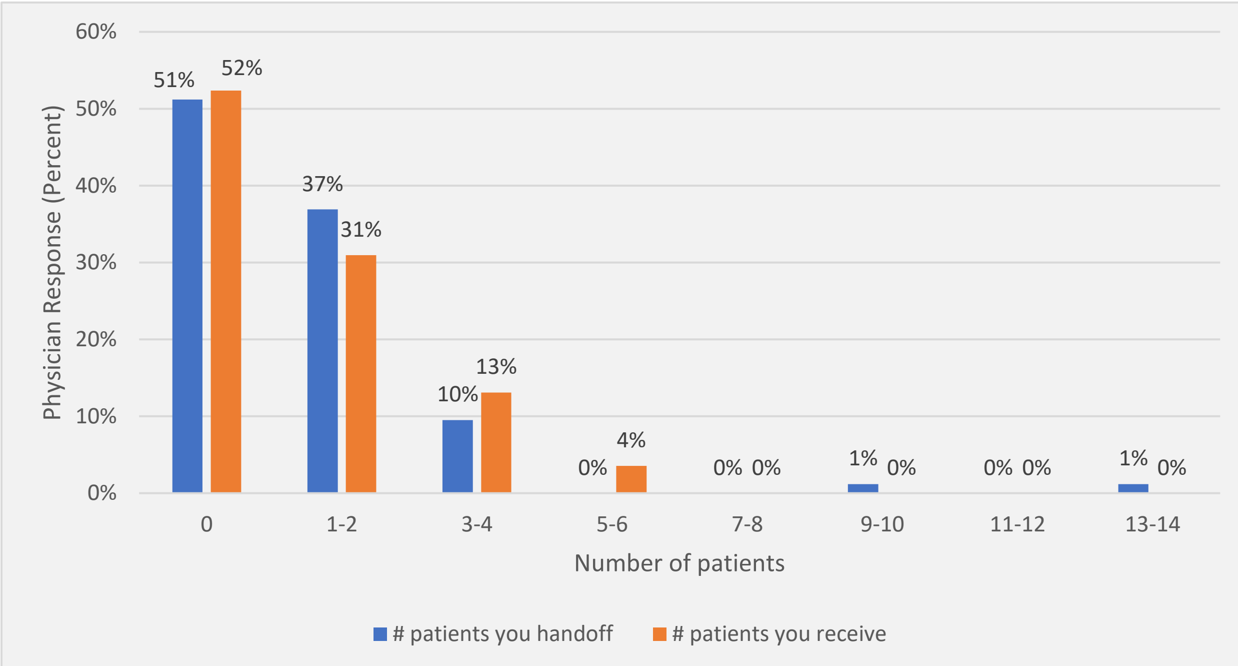


Figure 4.6: Physician preference on the number of patient handoffs.

These findings align with the majority response from the physician's perceptions of handoffs on performance metrics. However, even the physicians who perceived handoffs to have a positive impact on different performance metrics preferred to have fewer handoffs. Investigating the free-text response, most physicians preferred receiving and handing off fewer patients citing that handoffs burden the oncoming physician and affect patient safety. From these observations, it is evident that physicians prefer to avoid handoffs. However, to understand if physicians took any actions to avoid handoffs, we analyzed their response to the question, " How often do you make a conscious effort to minimize the number of handoffs?". The response was collected using a 5-point Likert scale, and Figure 4.7 below represents the physicians' responses. It can be observed that all the physicians reported making some conscious effort to avoid handoffs in the ED. Specifically, 77% of physicians reported that they *always* make a conscious effort to avoid handoffs. Further, 18% of physicians reported that they *frequently* take action to avoid handoffs and the rest 5%, reported that they take action *sometimes* to avoid handoffs. Analyzing the free-text response, the most common strategy used by physicians to avoid handoffs includes staying after shift, consolidating patient files during the end of shift, and signing up fewer patients at the end of shift.

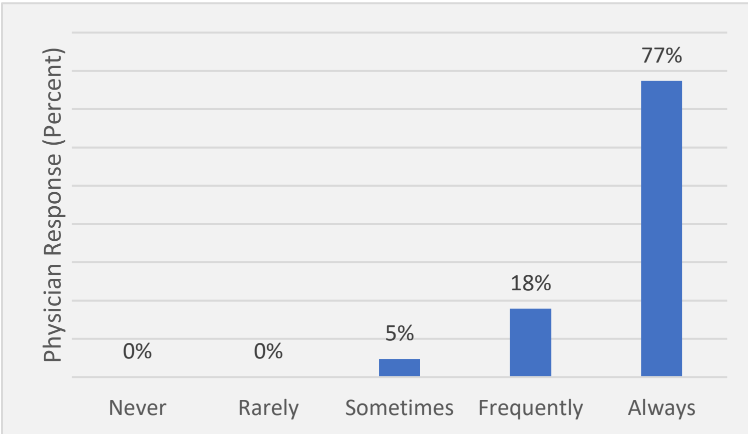


Figure 4.7: Frequency of physician's efforts to reduce the number of handoffs.

Finally, on analyzing the response to physicians' willingness to extend the shift if compensated, we observed that 82% of the physicians were willing to extend the shift, and 18% did not want to increase their shift length. Figure 4.8 below represents the physicians' willingness to extend the shift, and 61% of physicians reported that they were willing to extend the shift by an hour. Additionally, 4%, 8%, and 9% of physicians were willing to extend the shift by 90,30 and 120 minutes, respectively.

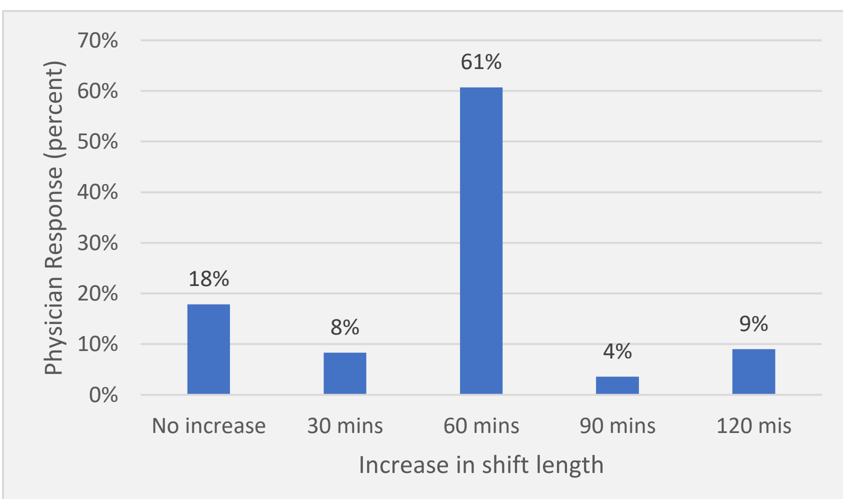


Figure 4.8: Physician willingness to extend shift.

To summarize the findings from the survey responses gathered, the majority of the physicians perceived handoffs reduced patient safety, patient satisfaction, and increased patient length of stay. Additionally, most physicians preferred not to hand off or receive patients during the end or beginning of the shift and reported taking actions to avoid handoffs. Finally, most physicians were willing to extend their shift by an hour which aligned with the idea of a one-hour overlapping shift.

From our simulation results and some sparse literature, it is evident that one-hour overlapping shifts can reduce the number of handoffs. However, to better understand the number of handoffs observed in the partner ED, we analyzed 3- years' worth of retrospective data from the partner ED. Since no reports directly track if a patient was handed off or not, we had to manipulate a few columns to get the number number of handoffs. Each patient visit in the ED is marked with three provider features, including the ID of the first physician providing care, the ID of the last physician providing care, and the ID of the longest physician providing care. If the first and the last care physician IDs match, it suggests there was no handoff, and if it doesn't match, it means the patient was handed off. If a patient was handed off, we compared the first physician ID to the longest physician ID, and the motivation for this was to identify if this handoff would have been avoidable. We define potentially avoidable handoffs as those where patients were signed up during the last hour of the shift and handed off. However, this is hard to capture as the time stamp of handoffs is not recorded. Hence we use time spent in the ED to identify these scenarios. For example, if a patient spends less than 121 minutes in the ED and the first physician is not the longest physician, then it suggests that the patient was signed up by the first physician during the last hour of the shift, which could have been potentially avoided. Similarly, we use various logic to isolate the avoidable handoff. However, it should be noted that if there are more than 2 physicians involved in care, then the handoffs cannot be captured using this approach, but those scenarios are rare as only a few patients spend over 10 hours in the ED. Table 4.4 below represents the percentage of patients handed off and the percent where the first physician is not the longest provider for three years.

Table 4.4: Percent patients handed off, and the first physician is not the longest.

Year	**Percent Patients Handed off**	**Percent Handoffs where the first provider is not the longest**
Sep 2018 – Aug 2019	30%	47%
Sep 2019 – Aug 2020	31%	48%
Sep 2020 – Aug 2021	30%	50%

From the table above, it can be observed that at least 30% of patients arriving at the ED are handed off, and of those handed off, about 47% of the cases had the first physician not be the longest caring physician. However, this does not suggest that 47% of handoffs are avoidable because some patients require complex care and spend extended time in the ED, which is unavoidable. However, we observed that 9-12% of cases of handoffs are avoidable based on the criteria defined above. This observation further suggested that one-hour overlapping shifts, if implemented, could practically reduce the number of handoffs.

After discussing the findings from the survey, retrospective data analysis, and the first two shift schedules with the physician stakeholders, our next step was to develop new shift schedules. According to the ED physicians, one of the main drawbacks of the first two shift schedules was their lack of practical application to the ED. Based on their feedback, a few additional constraints were added to the model to restrict shift start times during certain hours. Specifically, physicians did not want to start a shift after 5:00 pm in the evening as the end time of those shifts would be at odd hours. For example, a shift starting at 7:00 pm in the evening would end at 3:00 am, given it is an eight-hour shift. However, to ensure that ED is covered to meet patient demands, physicians suggested adding a potential time window at 11:00 pm that can be used as a shift stating time.

Physicians suggested using 11:00 pm as a potential shift start time as the shift end would be after 7:00 am. Further, any shift starting between 12:00 am, and 6:00 am was also restricted as these were operationally impossible. Further, we added a constraint to account for minimum ED coverage, which required at least three physicians to be present in the ED during any given hour of the day. With these new constraints, we generated an 8-hour shift schedule. Additionally, based on the survey feedback, we explored the idea of generating a 9-hour shift schedule as the majority of physicians were willing to extend their shift by an hour, and ED stakeholders suggested these were implementable in an ED. Finally, we also generated a weekly schedule with a combination of 8- and 9-hour schedules as some physicians preferred not to extend their shifts. Although a few physicians reported willingness to extend the shift by 30 mins, 90 mins, and 120 mins, we did not test these schedules as these were operationally impossible and added more confusion to shift scheduling start times and coverage. Thus, in total, we generated three new operational shift-shift schedules.

Table 4.5 below represents the new shift start times for the week under each policy. Here policy 3 represent an 8-hour physician shift, policy 4 represents a 9-hour physician shift, and policy 5 represents a combination of 8 and 9-hour shift. The first thing to notice here is how the start times are restricted to certain time frames that are very similar to the baseline policy, as these were added as the new constraints. Although the shift start windows are the same, one of the interesting factors to notice is how the schedule generated by the mathematical model recommends starting a shift in a staggering approach as opposed to starting shifts only at particular time frames (e.g., 7:00, 9:00, etc.) as observed in the baseline policy.

Table 4.5: Weekly physician shift start times.

Time	Baseline	Policy 3	Policy 4	Policy 5
12:00 AM	0	0	0	0
1:00 AM	0	0	0	0
2:00 AM	0	0	0	0
3:00 AM	0	0	0	0
4:00 AM	0	0	0	0
5:00 AM	0	0	0	0
6:00 AM	0	0	0	0
7:00 AM	22	21	14	21
8:00 AM	4	14	14	14
9:00 AM	20	14	14	14
10:00 AM	5	7	7	7
11:00 AM	0	7	7	7
12:00 PM	2	7	0	0
1:00 PM	0	0	7	7
2:00 PM	0	7	7	7
3:00 PM	21	14	14	14
4:00 PM	4	14	21	14
5:00 PM	20	14	0	7
6:00 PM	0	0	0	0
7:00 PM	0	0	0	0
8:00 PM	0	0	0	0
9:00 PM	0	0	0	0
10:00 PM	15	0	0	0
11:00 PM	21	21	21	21
Total Shifts	134	140	126	133
Weekly Hours	1128	1120	1134	1127
Change in hours		-1%	1%	0%

Figure 4.9 below represents the hourly average patient arrival and ED physician availability for a day. It can be noticed that the staffing schedule generated by the staffing schedule staffs ED in a dynamic manner based on patient arrivals to the ED, where more physicians are staffed during the peak time of patient arrivals to the ED. Especially between 10:00 AM to 6:00 PM, the dynamic staffing policies track the same pattern of patient arrivals, while the baseline policy aims to maintain a steady level of physician availability. Additionally, it can be noticed how the new

staffing policies generated using the mathematical model maintain the minimum required levels of physician staffing during the non-peak hours (12:00 am – 7:00 am). Further, for only a 9-hour policy (policy 4), it can be observed that staffing availability increases during 3:00 and 4:00 PM because a 9-hour shift cannot be started at 5:00 pm because of operational infeasibility.

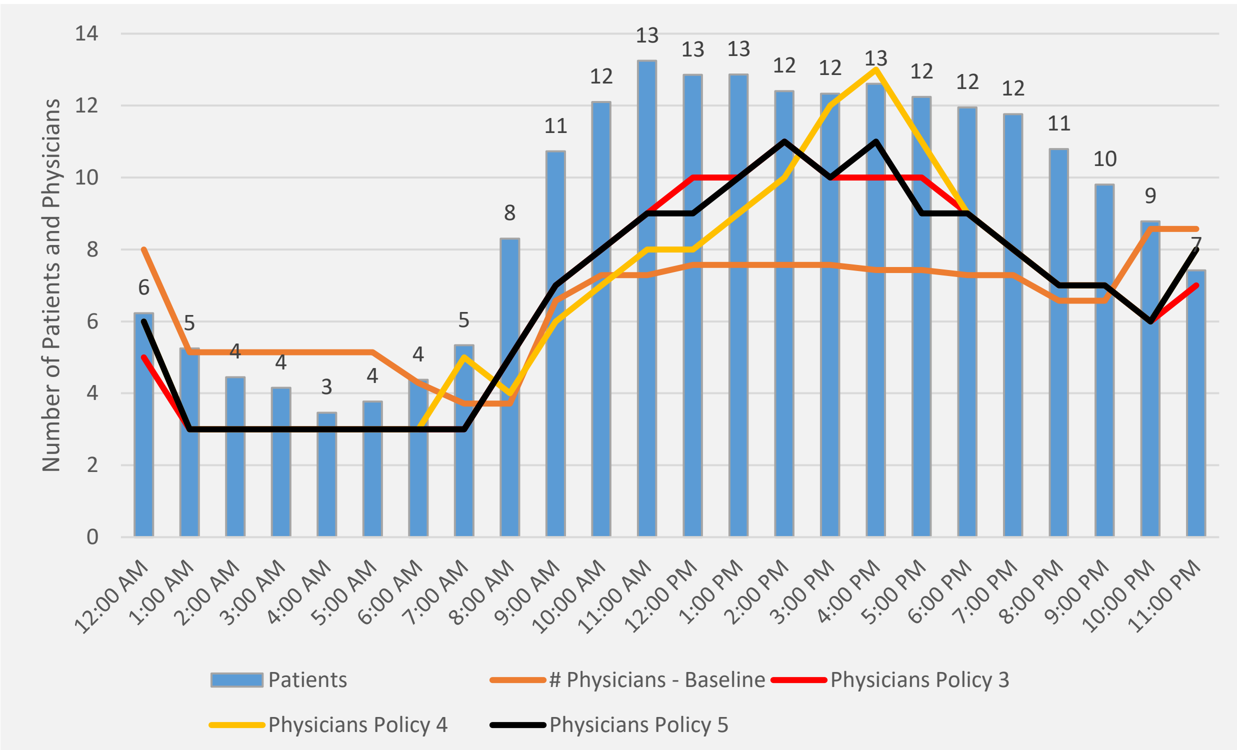


Figure 4.9: Hourly patient arrivals and physician availability in the ED.

After generating the new schedules, we tested these three policies in the validated simulation model. The simulation runs and all the parameters were the same as the prior runs. Table 4.6 below represents the performance metrics under each policy, including the non-operational policies. It can be noticed here that with these additional constraints, the amount of variation in FTEs compared to the baseline policy is not significantly different. However, we did observe a statistically significant difference in the time in the ED and the number of handoffs compared to the baseline policy. Further, in Table 4.7, we present these results in percentage differences for easier comparison and identifying the best policies.

Table 4.6: Simulation model results.

Policy	Weekly Throughput	# handoffs per day	Time in the ED (mins)	Change in hours/week (FTEs)
Baseline	1505	93	213±4.6	0
Policy 1*	1503	88	207±7.2	-56 (-1.4FTEs)
Policy 2*	1506	81	199±6.1	+16 (+0.4 FTEs)
Policy 3	1509	87	206±8.3	-8 (-0.2 FTEs)
Policy 4	1503	84	201±5.9	+6 (+0.2 FTEs)
Policy 5	1501	85	201±6.0	-1 (+0.1 FTEs)

*Non-operation policies

Table 4.7: Percentage difference in metrics compared to the baseline policy.

Policy	Handoffs	Time in ED	Change in hours /Week
Policy 1*	-5.2%	-2.5%	-56 (-1.4FTEs)
Policy 2*	-12.0%	-6.4%	+16 (+0.4 FTEs)
Policy 3	-6.1%	-3.3%	-8 (-0.2 FTEs)
Policy 4	-9.2%	-5.6%	+6 (+0.2 FTEs)
Policy 5	-8.7%	-5.4%	-1 (+0.1 FTEs)

*Non-operation policies

From Table 4.7, it is evident that policy 1 appears to be the best in terms of FTE reduction and performance improvement. But, policy 2 has the most significant reduction in the number of handoffs and time in the ED for a slight increase in FTE needs. However, based on feedback from the ED physicians, these two policies are operationally infeasible. Hence we move to the operational policies (3, 4, 5), which were developed based on the feedback from survey research and retrospective data analysis. Among these three policies, 8- and 9-hour shifts add the most value in terms of a decrease in the number of handoffs, time in ED, and a slight reduction in FTEs. Additionally, these policies align with the subjective feedback provided by the physicians, where some preferred to have a 9-hour shift whereas a few others preferred an 8-hour shift. However, policy 3, and 4, which maintains the same shift length (8 or 9 hours), might be preferred in some EDs as it ensures fairness and less confusion among ED physicians. A modified version of the shift schedule generated by the mathematical model is currently implemented at the partner health system.

4.3.3 Discussions and Conclusions

Optimal staffing of an ED is a crucial factor in ensuring smooth patient flow and improving patient safety by avoiding overcrowding. Researchers have used various operations research methods to improve patient flow by minimizing the wait times, patient time in the ED and improving ED throughput. However, very few studies have considered patient safety as a performance metric along with the patient flow metrics to evaluate ED performance. Additionally, one of the primary issues with some results generated from operations research methods is the lack of implementation in the actual testing environment. Hence it is critical to involve stakeholders and end users while developing the solution. This research focused on developing optimal ED physician shift schedules that improve patient safety and patient flow while considering the staffing budget. To address this research objective, we developed a MILP model and used survey research along with retrospective data from partner ED to inform the modeling. Further, the shift schedules generated from the MILP model were tested in the validated simulation model representative of the partner ED.

From the survey responses, we identified that 69% of physicians felt that handoffs have a negative impact on patient safety, 67% felt that handoffs increased the patient length of stay (negative impact), and 56% felt that handoffs reduce patient satisfaction. Although prior studies have reported similar findings with respect to patient safety, it is interesting to notice that handoffs can also act as a surrogate for patient satisfaction metrics. Additionally, the survey responses also helped shed light on physician preferences on the length of their shift and willingness to extend their shift, along with the preferred number of patients they would like to hand off or receive. Finally, the retrospective analysis further helped us understand the opportunity for reducing handoffs by using the overlapping shift approach discussed in the prior chapter.

Although handoffs, the transfer of patient care from one physician to another, are unavoidable in the ED because of the continuous patient flow, we developed a MILP model that minimizes the combined cost of handoffs as the (patient safety metric), time in the ED and waiting times (patient flow metric), along with the staffing cost to represent the ED staffing budget. After generating schedules based on physician preferences and feedback, we observed that a combination of 8- and 9-hour staffing schedule was the most effective, where it reduced the patient time in the ED by 5.4% and handoffs by 8.7% without affecting the ED budget. These findings could be scaled to other large Level 1 trauma centers depending on the patient census. However, smaller EDs with fewer patient arrivals (<35,000 patient arrivals/year) that have different operation constraints should be cautious while adopting similar policies.

5. Chapter 5

5.1 Understanding and Detecting Physician Stress in Emergency Departments

5.1.1 Background and Literature

The crisis of physician burnout, stress, and clinical errors has been a topic of discussion and research for the past two decades [141,142]. Burnout is defined as a condition of high emotional exhaustion, depersonalization, and low personal accomplishment. The burnout rates among physicians are increasing irrespective of the research and preventive measures that are adopted. A 2019 survey investigating the burnout rate among 15,000 physicians from 29 different specializations reported that 44% of the physicians were burned out [11]. Moreover, Emergency Medicine was one of the top five specialties reporting higher levels of burnout [11]. A comparative study that examined the burnout rates of physicians in the US to other general working adults reported that physicians had a 10% higher chance of burnout [143]. Additionally, a two-stage study investigating burnout among ED physicians reported that they had high emotional exhaustion and depersonalization [144].

The main reasons attributed to physician burnout are bureaucratic tasks (e.g., charting, paperwork) resulting in increased time spent on the EHR, long working hours, and stress [11,145–149]. Frequently, burnouts are preceded or accompanied by periods of prolonged stress [142]. ED staff are often exposed to high levels of stress due to the diverse nature of patient conditions, which include life-threatening emergencies, injuries, and chronic ailments. ED overcrowding, another potential source of stress, has also become more common. Prior studies have identified emergency department overcrowding, high patient inflow, and long work hours as significant contributors to higher levels of stress and frustration among ED physicians [9,10,150].

Stress and fatigue are important factors that contribute to medical errors and fatalities. According to the Institute of Medicine's (IOM) seminal article, "To Err is Human," between 44,000-98,000 deaths per year result from medical errors [151]. Furthermore, there were higher chances of errors when the ED was overloaded and did not have adequate resources or equipment [152]. Overall, human errors, primarily cognitive and incorrect clinical assessments, were reported as the leading causes of errors in ED [152].

Studies that investigated the role of experience on stress and medical errors have observed that experience helps physicians to develop internal control mechanisms to cope with various treatment conditions in the ED. In a study analyzing over 7,000 hours of endocrine stress response from 112 nurses and 27 physicians working in critical care, the mean raw cortisol levels were lower among the experienced team members, suggesting the role of experience in stress management [153]. The capability of experienced physicians to manage stress could be attributed to the coping mechanism they develop [154]. A study investigating medication errors in the pediatric ED setting reported that less experienced resident physicians made a higher number of errors compared to experienced physicians [155]. Another study focusing on the extent of supervision required for the ED residents reported that out of 480 patients reviewed by the residents, 37% required a change in proposed care [156].

Prior studies have extensively investigated the causes of errors, the number of errors, and their association with physician experience in the ED. Additionally, a few studies have compared the stress and burnout among the attending and resident physicians, but these studies used only qualitative methods and did not include an attending and resident physician working together on a shift [157–160]. A recent study that investigated the change in heart rate (HR) and salivary cortisol levels of pediatric ED attending physicians following resuscitation observed that both HR and salivary cortisol significantly increased after these events [161]. Similar to other studies, this study did not consider a resident physician working on the same shift and did not monitor the physiological measures for the entire shift. However, the study by Joseph et al., (2016) compared the stress of attending and resident physicians during trauma and emergency surgery using qualitative and quantitative measures of heart rate variability (HRV) [162]. Our research furthered the existing research by comparing the physician's stress during an entire shift, which included both trauma and non-trauma events, using qualitative measures and quantitative measures of HRV and electrodermal activity (EDA). Moreover, we considered the various domains of HRV metrics, including time domain and frequency domain metrics. Finally, we also investigated the physician burnout levels. To our knowledge, this will be the first study that uses subjective and objective data to compare stress and burnout among attending and resident physicians, stress changes during an entire ED shift, and specific trauma and non-trauma events. Specifically, we investigated the following questions in this chapter:

- Is the stress level the same in attending and resident physicians for an entire ED shift?
- Is there a change in HRV of attending and resident physicians during trauma and non-trauma events?
- Is the burnout rate similar between attending and resident physicians?
- Does experience impact the perceived workload of an ED shift?
- Is there a correlation between the subjective and objective measures?

5.2 Phase One

5.2.1 Methods

5.2.1.1 Participants

Participants for this study included 12 emergency physicians (8 male, 4 female) working a 3:00 pm - 11:00 pm shift at Greenville Memorial Hospital (GMH) in Greenville, SC. The Greenville Health System (GHS), now called PRISMA Health, is the largest healthcare provider in South Carolina and serves as a tertiary referral center for the entire Upstate region. The flagship GMH academic Department of Emergency Medicine is integral to GHS patient care services as the Adult Level 1 and Pediatric Level 2 Trauma Center, Stroke and ST-Elevation Myocardial Infarction (STEMI) Comprehensive Center seeing over 106,000 patients annually.

Meetings with ED physicians and faculty were organized prior to the start of data collection to discuss the purpose and methodology. Six participants (mean age = 26.8 ± 1.5 years, 4 male, 2 female) were first-year resident physicians, and the other six (mean age = 42.66 ± 2.8 years, 4 male, 2 female) were attending physicians with an average experience of 8 years of practice. An attending was paired with only one resident physician during the shift, and all the attendings had a minimum of five years of experience working in the ED. For analysis, we considered 42 events (21 trauma and 21 non-trauma) for each group. The number of events was based on a meta-analysis study that evaluated 297 studies that used HRV to compare two groups [163]. Further, by power analyses, when the significance level was set at 0.05, a sample size of 21 events had 80.0% power to detect an effect size of .9 between two groups. Consent was obtained from physicians before the shift, and the study was approved by GHS IRB Pro00058516.

5.2.1.2 Apparatus

The Empatica E4 watch is a wearable device that collects real-time physiological data. This wrist band is equipped with four types of sensors: two metallic electrodes for measuring EDA, a three-axis accelerometer, an optical thermometer, and photoplethysmogram sensors for computing HRV. This device can measure skin temperature, HRV, EDA, and acceleration. Prior research has validated the effectiveness of this device, and one study that specifically compared it to the medical devices used in the hospital reported that Empatica E4 echoed the data collected from the medical devices [164]. Additionally, multiple research studies have used Empatica E4 for computing stress, emotional arousal, epileptic seizures, sleep quality, and arterial fibrillation [165–170].

Empatica EDA data captures the skin conductance response (SCR), i.e., the phasic response, and skin conductance level (SCL), i.e., tonic response. Variations in the phasic component are observed as GSR peaks, and it is sensitive to specific emotionally arousing stimulus events. In this research, we used only the tonic component, SCR, to examine the stress levels during an event.

HRV measures the change in the time between successive heartbeats. The time between beats is measured in milliseconds (ms) and is called an “R-R interval” or “inter-beat interval (IBI).” In stressful situations, HRV is a product of a change in the autonomic nervous system, which is composed of the sympathetic and parasympathetic nervous system. Examining the relationship between the sympathetic and parasympathetic nervous system provides insight into the stress state of an individual [171–173]. Empatica E4 collects the EDA data at a sampling rate of 4 Hz and HRV data at 64 Hz.

5.2.1.3 Procedure

First, we identified the pair of residents and attending physicians working together for an eight-hour shift. Prior to the shift, each physician was handed the consent form. The physicians were then asked to put on the Empatica E4 wristwatch at least five minutes prior to the start of their shift to obtain the baseline data. As mentioned above, Empatica E4 collects various physiological measures, including EDA, HR, and HRV. Finally, before starting the shift, physicians were handed the Maslach Burnout Inventory-Human Service Survey (MBI-HSS), which measures emotional exhaustion, depersonalization, and personal accomplishment. Collectively, these measures provide a surrogate for the burnout rate. The MBI-HSS survey was administered prior to the shift to control for the effect of the shift on their response. They were also given the NASA-TLX survey at the beginning of the shift to assess the current perceived workload.

During the shift, an observation sheet was used by the person shadowing the physician to note the physician's activities. The main activities noted were computer interaction, patient interaction, discussion, and trauma, as described in Table 5.1. The researcher shadowing the physician time-stamped the beginning and end of each activity on the observation sheet. The physician activities were classified and coded as represented in Table 5.1 to maintain the consistency of classifying the physician activities by different researchers shadowing the physician. Further, any events or incidents that a researcher had concerns about were noted on the observation sheet and discussed as a team to address the issue. These time stamps were later used during the analysis to comprehend the change in physiological measures during certain events. Any events that could have a confounding effect on the data were noted on the observation sheet and not considered for analysis.

Table 5.1: Description of different shift activities.

Activity	Description	Average Number of events per shift
Computer Interaction	Physicians use the EHR, Charting, and reading reports.	21 ± 2.1
Patient Interaction	Physicians interact with patients regarding their health issues.	16 ± 1.8
Discussion	Attending physicians discuss/teach the resident physicians. Physicians talk / engage with other physicians.	10.3 ± 2.3
Trauma	Level 1, 2, and 3 trauma cases include car crashes, gunshot wounds, etc. This includes complex procedures such as intubation, etc.	4 ± 0.9

Post-shift, the physicians were again given the NASA-TLX to assess the shift workload and obtain a workload index score. NASA-TLX measures include mental demand, physical demand, temporal demand, effort, performance, and frustration. NASA-TLX has been used and validated in the healthcare setting to measure stress, and the workload during a particular event, including surgeries, ICU shifts, etc. [174–177].

5.2.1.4 Analysis

The data collected from the Empatica E4 was first preprocessed before analysis. The data did not require downsampling as the maximum sampling rate was 4Hz. Hence, the first step was mapping the data points to the events recorded in the observation sheet based on the time stamp. This was performed as the first step to avoid the mix-up of data and events during data preprocessing. To remove the artifacts from the HRV data, we first visualized the RR Intervals (the time between two successive R waves), and any ectopic beats and motion artifacts were manually identified and marked [178]. Additionally, any RR Intervals of more than 1300 ms and less than 400 ms were also marked. The marked data and any missing values were interpolated based on preceding and successive beats using the cubic spline interpolation technique [179]. Similarly, any confounding spikes resulting from hand motions, etc., recorded on the observation sheet were identified and interpolated from the EDA data. Only events that lasted for at least three minutes were considered for analysis because of the lack of inference that could be drawn from short events. The three minutes duration of events was based on observations from prior studies investigating the time and frequency domain metrics of HRV. Shaffer & Ginsberg (2017) observed that at least two minutes of HRV data is required for interpreting and inferring accurate conclusions. However, a recent study that compared the 2 minutes vs. 3 minutes data to >5 minutes data observed that 3 minutes data had a strong correlation to >5 minutes data [181]. Hence, in this study, we decided to use three minutes as the minimum duration of an event to be considered for analysis. Finally, any events noted in the observation sheet as potentially having confounding effects were not included in the analysis.

For event-based comparison, we first sampled 21 random trauma events and 21 non-trauma events (7 patient interactions, 7 discussions, and 7 computer interactions) from various ED shifts. The 42 events of interest were now split into individual events from the dataset. Further, the physiological measures (EDA & HRV) for the activities of interest were compared between the attending and resident physician pairs.

EDA refers to the variation of electrical properties of the skin in response to sudomotor activity [182]. The nerve fibers trigger the sudomotor activity, and concurrence of the firing of these fibers results in a quick burst, which can be recorded as a skin conductance response (SCR) [183]. SCRs are also referred to as the rapidly occurring phasic component of EDA. Variations in the phasic component are observed as EDA peaks, and it is sensitive to specific emotionally arousing stimulus events [184–186]. Prior studies investigating the change in EDA during stressful stimuli have observed an increase in the SCR activity and amplitude during the stressful stimulus [187–189]. The EDA data were compared by the number of skin conductance response (SCR) peaks for the events.

For this, data for each event were imported to MATLAB and prepared for analysis using the open-source MATLAB package, Ledalab. Further, the data was analyzed by continuous decomposition analysis and optimized to correct motion artifacts. The peak threshold for the SCR was set as 0.01 micro siemens based on prior research recommendations [190,191]. Hence a burst will only be considered an SCR if its rise phase exceeds the threshold value. Finally, the number of peaks for the specific events was recorded. EDA data for the entire shift was not compared, as the high sensitivity of the data makes it undesirable to make accurate inferences [191].

Similar to the EDA analysis, the HRV analysis of the attending and resident physician pair was compared for 21 trauma events and 21 non-trauma events (7 patient interactions, 7 discussions, and 7 computer interactions) from various ED shifts. Additionally, to understand the long-term effect of stress, we conducted an HRV analysis on the full shift data.

HRV, which is the change in the time between successive heartbeats, is a reliable reflection of many physiological factors [192]. It has been used as a quantitative marker to understand the interplay between the sympathetic and parasympathetic nervous systems. The sympathetic nervous system is our fight and flight response, whereas the parasympathetic is the rest and digest response. The former activates during high stress, anxiety, or fear, while the latter helps the body to maintain homeostasis [193]. Although HRV can be analyzed and interpreted using a variety of methods, the most common and reliable methods used are time domain and frequency domain metrics [178]. Hence for analysis, we considered both time domain and frequency domain metrics of the HRV.

In the time-domain metrics, the root means square of successive RR differences (RMSSD) is one of the most used metrics to interpret stress, and it is closely related to parasympathetic activity [194,195]. Similarly, in the frequency domain metrics, the low frequency (LF) 0.04-0.15Hz and high frequency (HF) 0.15-0.4Hz bands and their ratio, i.e., LF/HF, are the validated metrics for interpreting stress [196–198]. The LF/HF ratio is interpreted as the ratio of the activity of the sympathetic nervous system to that of the parasympathetic nervous system (sympathovagal balance) [178]. For analysis, we considered RMSSD as the time domain metric, LF/HF ratio as the frequency domain metric, and RR Interval as a metric to measure the overall HRV. A higher RMSSD value, RR Intervals, and a low LF/HF ratio suggests low stress [180]. Description and interpretation of the HRV metrics are provided in Table 5.2. These metrics were selected based on previous research studies and the task force report [178,180].

The preprocessed RR Interval data were imported to a validated HRV analysis software, Kubios [199]. First, we calculated the RMSSD from RR Intervals data. Further, to obtain the frequency domain metrics, we used a fast Fourier transform (FFT) on RR Intervals. FFT was preferred over the autoregressive transformation because the latter tends to smooth the frequency curves, leading to misinterpretation of results [178].

Table 5.2: Description and interpretation of HRV.

Measure (units)	**Interpretation**
RMSSD (ms)	The root mean square of differences of successive RR intervals describes short-term variations. A low value indicates high stress.
RR Interval (ms)	The time elapsed between two successive R waves of the QRS signal on the ECG. A higher value indicates higher variability/low stress.
Power LF and HF (n.u.)	High LF indicates high stress, and high HF indicates low stress. The ratio of LF and HF frequency band powers. A low LF/HF ratio indicates lower stress.

5.2.2 Results

In the following sections, we present results based on three classifications: stress, burnout, and workload. A series of two-sample t-tests were conducted to analyze the difference in physiological measures, NASA-TLX scores, and MBI-HSS scores of attending and resident physicians. A 0.05 alpha level (α) level of significance was maintained for all the t-tests.

5.2.2.1 Stress

The stress levels of the attending and the resident physicians were assessed via the different functions of the HRV and EDA from the data. Only the HRV functions were used to calculate the stress for the entire shift because the high sensitivity of EDA data makes it undesirable to use for longer durations. The stress levels during trauma and non-trauma events were interpreted with both HRV and EDA responses, and a total of 21 trauma and 21 non-trauma events were used for the analysis.

For the 21 trauma events, attending physicians had a lower level of stress compared to the residents. This was supported by the time domain components and the frequency domain components of the HRV, as seen in Table 3. Figure 5.1 represents the box plot of RMSSD for attending and resident physicians. From the box plot, it is evident that the residents had a low overall RMSSD compared to the attendings. RMSSD reflects the beat-to-beat variance in the HR and estimates the parasympathetic changes in the HRV. The higher value among attendings represents a higher parasympathetic activity among this population during trauma events, indicating lower stress.

Table 5.3: t-test result from trauma HRV.

Function	**Physician**	**N**	**Mean**	**P-value**
RMSSD	Attending	21	47.0 ± 7.7	0.001
	Resident	21	35.2 ± 12.4	
RR Interval	Attending	21	845.4 ± 49.31	<0.001
	Resident	21	774.6 ± 49.83	
LF/HF	Attending	21	1.7 ± 0.5	0.001
	Resident	21	2.5 ± 0.8	

Similarly, Figure 5.2 represents the box plot of RR Intervals for both the attending and resident physicians for the trauma events. The RR Interval, which is the period between successive heartbeats, is an estimate of the overall HRV. The attendings had higher RR Intervals compared to the resident physicians during the trauma events, suggesting lower levels of stress. Figure 5.3 represents the box plot of the LF/HF ratio, which represents the sympathovagal balance (the ratio of sympathetic activity to parasympathetic activity). Attending physicians had a lower LF/HF ratio compared to the resident population during the trauma events, suggesting higher parasympathetic activity (i.e., lower stress). Although the GSR metric of the mean number of SCRs was 25.3% higher among the resident physicians compared to the attendings, indicating elevated arousal levels suggesting higher stress levels during the trauma events, the results were not statistically significant. More samples of trauma events would be needed to reach a conclusion regarding EDA.

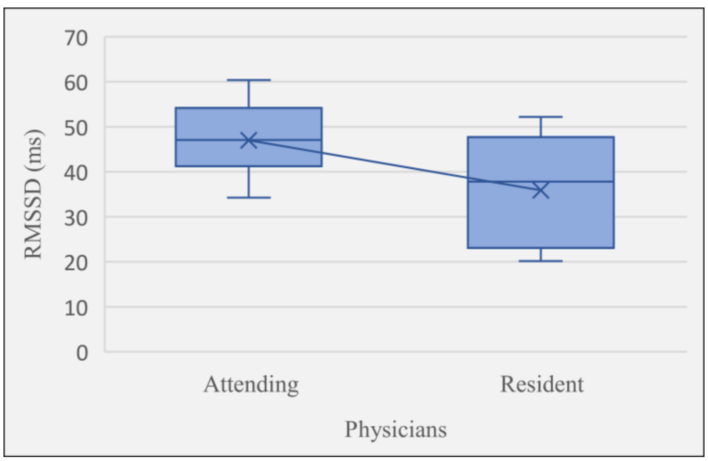


Figure 5.2: Box plot of RMSSD during trauma.

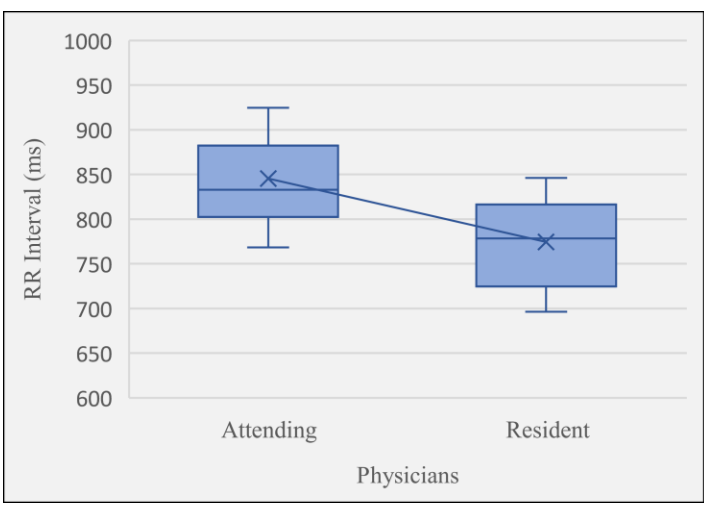


Figure 5.1: Box plot of RR Interval during trauma.

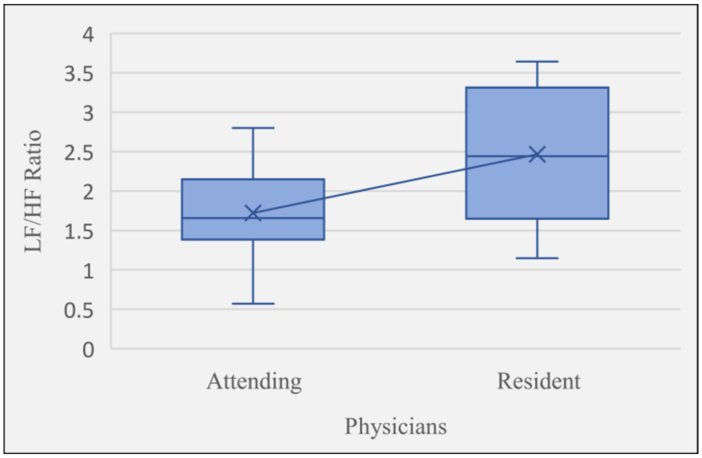


Figure 5.3: Box plot of LF/HF ratio during trauma.

Table 5.4 represents the t-test results for the full shift HRV analysis. It can be observed that the difference in RMSSD values was significant, suggesting that attending physicians had a better coping mechanism for the entire shift. However, unlike in trauma events, the frequency domain metric differences were not significant.

Table 5.4: t-test result from full shift HRV.

Function	Physician	N	Mean	P-value
RMSSD	Attending	6	44.2 ± 7.5	0.033
	Resident	6	35.4 ± 4.5	
RR Interval	Attending	6	829.4 ± 66.5	0.199
	Resident	6	780.1 ± 57.1	
LF/HF	Attending	6	2.3 ± 0.5	0.106
	Resident	6	2.7 ± 0.4	

Figure 5.4 represents the box plot of the average RMSSD value for the attending and resident physicians for the full shift. Similar to the trauma events, physicians had higher RMSSD compared to the resident physicians. It can be noticed that the maximum value of the residents' RMSSD is almost the same as the median value of the attending physician. The higher variability in the HRV among the attending physicians represents lower stress. Although the attending physicians had a low LF/HF ratio and higher RR Intervals compared to the residents, the data was not statistically significant.

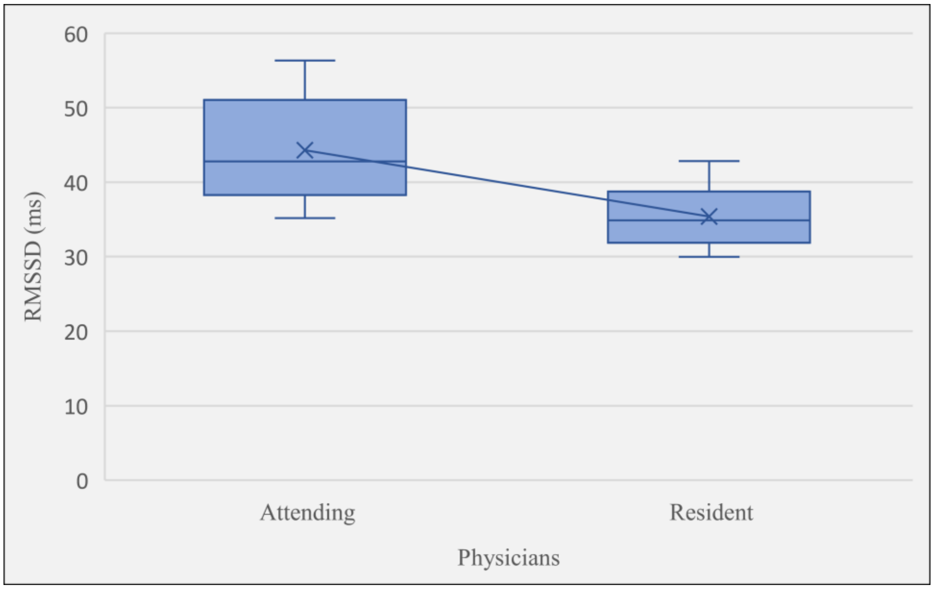


Figure 5.4: Box plot of RMSSD for the full shift.

Finally, from the 21 non-trauma events considered, both the attending and resident physicians demonstrated higher RR Intervals, RMSSD, and lower LF/HF ratio, and the differences between groups were not statistically significant. This observation suggests that during non-trauma events, there was no significant difference in the levels of stress demonstrated by the attending and resident physicians.

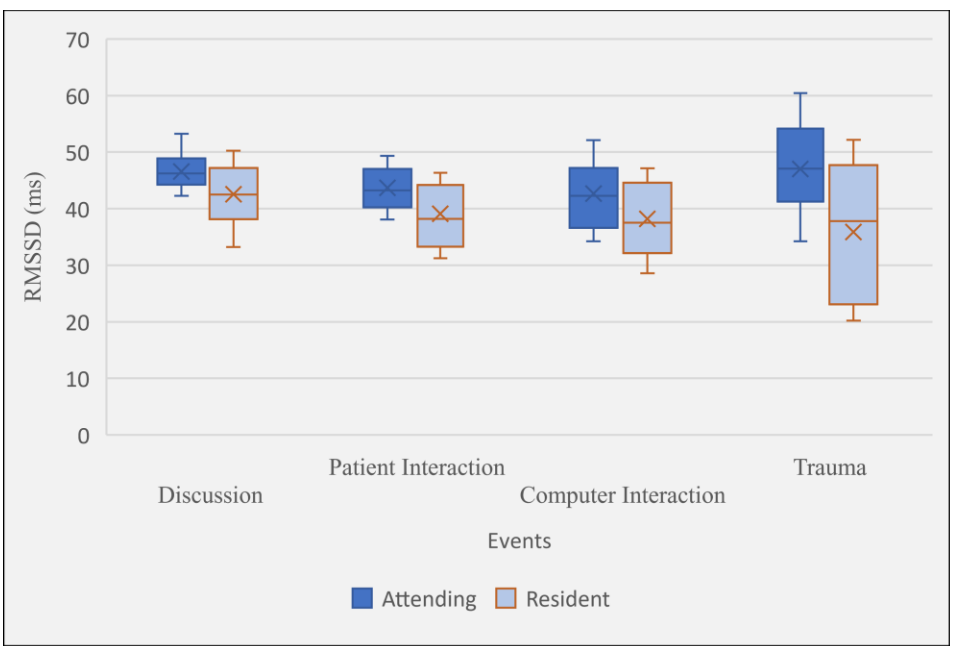


Figure 5.5: Box plot of RMSSD for all events.

Figure 5.5 above represents the box plot of RMSSD for attending and resident physicians for all events considered in the study. It can be observed that there was a high variability during the trauma events compared to the other three events. Additionally, although statistically insignificant, it can be seen that for all non-trauma events, the attending physicians had a higher RMSSD, which was also observed throughout the shift, suggesting lower stress among the attending physicians.

5.2.2.2 Burnout

Burnout was measured using MBI-HSS, which includes emotional exhaustion, depersonalization, and personal accomplishment. The survey noted that personal accomplishment was high for both attending and resident physicians (100% n=12), while depersonalization was higher in resident physicians (high=50%, average=33.3% and low=16%, n=6) compared to attending physicians (high=16.6%, average=16.6% and low =66%, n=6). For emotional exhaustion both groups reported low or average levels (resident: high=0%, average= 66%, low =34%, n=6; attending: high=0%, average =50% low =50%, n=6).

A t-test was conducted to analyze the differences in the MBI-HSS score of attending and resident physicians, and not significant differences were observed (p-value = 0.12). This could be because of the small sample size, and additional attending and resident physicians' participation would be needed to understand the burnout levels among the two groups.

5.2.2.3 Workload

The physician workload for the shift was calculated with the NASA-TLX response. The unweighted NASA-TLX was used for an accurate response. Table 5.5 represents the t-test results to understand the difference in the perceived workload. The results show that the attending physician had a lower NASA-TLX score. Additionally, investigating the individual effects, we observed that attending physicians had lower mental demand, physical demand, and effort. The results for the temporal demand, performance, and frustration were not significant for the t-test as seen in Table 5.5 below. Although we did not aim to investigate the physiological responses from a workload standpoint, it is interesting to notice that the attending physicians reported less perceived workload, which aligned with the calmer (less stressful) physiological responses and vice versa for the resident physicians. These findings provide valuable insight into how attending and resident physicians perceive their workload and its relationship with physiological responses.

Table 5.5: t-test results for NASA-TLX.

Function	Physician	N	Mean	P-value
Mental	Attending	6	47.8 ± 6.4	0.002*
	Resident	6	69.7 ± 11.4	
Physical	Attending	6	23.2 ± 10	0.01*
	Resident	6	47 ± 15.3	
Temporal	Attending	6	37.7 ± 11.3	0.137
	Resident	6	49.2 ± 13.2	
Performance	Attending	6	14.7 ± 12.5	0.356
	Resident	6	25 ± 23	
Effort	Attending	6	52.7 ± 13.8	0.042*
	Resident	6	70.2 ± 12.1	
Frustration	Attending	6	34.7 ± 22	0.95
	Resident	6	34 ± 12.8	
TLX-SCORE	Attending	6	35.1 ± 7.3	0.004*
	Resident	6	49.2 ± 5.7	

Figure 5.6 below represents the NASA-TLX score of the attending and resident physicians collected after the shift. It can be observed that the highest score observed among the attending physicians is less than the first quartile score of the resident physicians. Additionally, Figure 5.7 below represents the mental demand score of the attending and the resident physicians. On average, the resident physician's mental demand score was 46% higher than the attending physician. However, the temporal load, frustration, and performance did not show a significant difference.

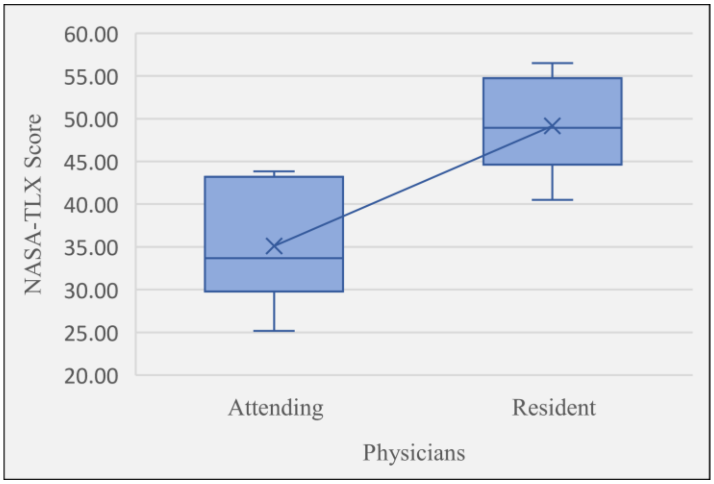


Figure 5.6: Box plot of NASA-TLX SCORE.

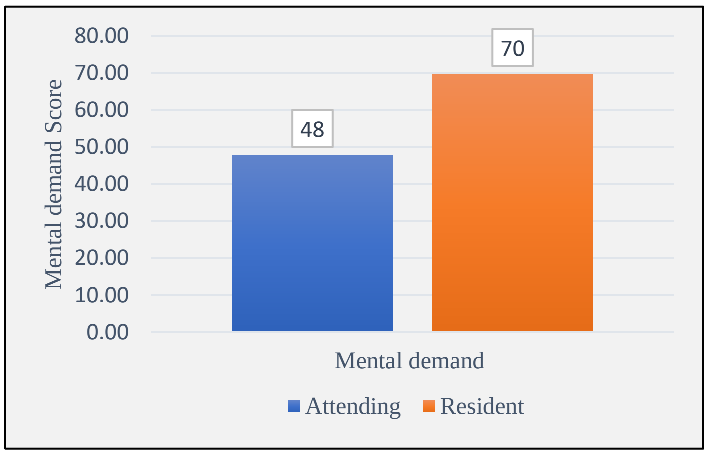


Figure 5.7: Mental demand score of physicians.

Finally, to comprehend the relationship between the subjective and objective measures, we performed correlation tests. Results from correlation tests of NASA-TLX vs. HRV metrics, we observed that NASA-TLX scores and RMSSD ($r(10) = -0.41$, $p = 0.18$) and RRI ($r(10) = -0.1$, $p = 0.76$) were negatively correlated and LF/HF ($r(10) = 0.28$, $p = 0.37$) ratio was positively correlated. These observations suggest that NSASA-TLX scores were high for physicians with high stress; however, these observations were statistically insignificant. For the correlation tests between MBI-HSS and HRV, the results did not replicate the pattern as observed in the NASA-TLX score, and all the results were statistically insignificant, which could be because of the small sample size.

Further, to investigate the relationship between individual components and objective metrics, a series of correlation tests were conducted. A strong negative correlation was observed between the mental demand component of NASA-TLX and RMSSD, $r(10) = -0.70$, $p= 0.01$. None of the other individual components and objective metrics were statistically significant.

5.2.3 Discussions and Conclusions

The first research question sought to investigate if experience played a role in managing the stress of physicians' work in the ED. To answer this, the stress levels for the entire shift were analyzed. Results from the HRV analysis showed that the experienced physicians had higher variability in the time domain HRV metrics, suggesting lower stress levels as compared to the residents. These observations are similar to previous studies that used subjective measures to compare attending and residents working in inpatient and pediatric medicine [157], junior and experienced attending endoscopists [158], and junior and experienced attending physicians [159]. For the frequency domain components of HRV, the attendings had a lower LF/HF ratio, suggesting a higher parasympathetic activity. However, the results were not statistically significant. One reason for this observation could be the frequent change in the user's position and movements in the ED, which could have potentially resulted in motion artifacts. Moreover, it could be because of the lack of LF/HF ratio to systematically represent the sympathetic and parasympathetic activity as prior studies have reported this ambiguity while considering the long-term measurement of the LF/HF ratio [180].

The second question focused on investigating if the HRV of attending and resident physicians were the same during specific events. Our results showed that differences in HRV between the attending and resident physicians were highly significant during trauma events, with attending showing high HRV, suggesting lower levels of stress. Additionally, we observed that the resident physicians had a higher number of SCRs during the trauma events compared to the attending physicians. These observations of stress differences during stressful events are similar to previous studies that evaluated the salivary cortisol levels of experienced and novice physicians and nurses after stressful events [153], novice and experienced physical therapists [200], HRV of attending and resident physicians during trauma activation and emergency surgery [162] suggesting that experience plays an essential role in stress management during complex situations. We also observed that both the attending and resident physicians demonstrated low levels of stress during the non-trauma events, and there were no significant differences in their physiological measures.

The third research question investigated the burnout levels among the attending and resident physicians. We hypothesized that the resident physicians would have a higher burnout rate compared to the attending physicians. However, the results were not significant to support this hypothesis as only the depersonalization items supported our view. This observation was similar to a past study that compared the burnout levels of attending and resident emergency physicians [160]. Although the results are statistically insignificant, it is worth noting that of the six residents, five reported moderate or high depersonalization compared to only two attendings.

The fourth question investigated the perceived workload of physicians during an ED shift. NASA- TLX measured the mental, physical, and temporal demand along with the performance, effort, and frustration. We hypothesized that the experienced physicians would have a lower NASA-TLX score as they are more accustomed to the environment, and the results supported this. We did observe a stark contrast in the mental demand, where the average resident's score was 46% higher than the attending physician's score, again supporting our hypothesis. This observation was similar to a previous study that used NASA-TLX to compare the attending and resident physicians' mental strain during trauma activation and emergency surgery [162]. The two possible reasons for these results could be a) the transition from academia to practice, especially in the case of first-year residents, and b) the coping mechanism adopted.

Finally, our last research question examined the correlation between subjective and objective measures. Although statistically insignificant, we observed that stress levels measured using HRV metrics correlated with the NASA-TLX scores. Specifically, we observed a strong negative correlation between the mental demand component of the NASA-TLX score and RMSSD. This observation is in line with prior studies that have observed RMSSD decreases during stressful situations [201–203].

There are a few limitations associated with this first phase of the study. One of the drawbacks is the low number of participants. However, we collected over 100 hours of data which was sampled at 64 Hz, providing a large data set for analysis. Moreover, we considered 42 unique events for event-based analysis, which is good sample size for a pilot study. Further, collecting data over an entire shift, accounting for many movements and actions, contributes to more noisy data and results in a complex dataset to analyze. In particular, a sensitive response like EDA records irrelevant peaks or bursts, limiting researchers from using this metric for long-term analysis. However, by using multiple metrics, HRV and HR allowed us to draw conclusions from the entire ED shift. Moreover, our analysis was performed by considering all of these extraneous factors, including motion artifacts and ectopic beats. All recorded shifts were also from the same 3:00 pm - 11:00 pm time period and in the same pod, which included many trauma events. Therefore, we cannot generalize these results to other ED shifts with different parameters. Finally, we considered only first-year residents, which does not include the whole resident population.

5.3 Phase Two

The observations and findings show that attending and resident physicians working in the ED are exposed to stressful events, and their physiological parameters change during these events. Our next step was to develop machine learning models to detect the early onset of stress among ED physicians. Stress can be estimated using various techniques, including subjective and objective measures. Few validated subjective measures include the Perceived Stress Scale (PSS) and other questionnaires, and objective measures include heart rate (HR), heart rate variability (HRV), electrodermal activity (EDA), and cortisol levels in endocrine stress response [204–209]. Although PSS and other questionnaires are validated methods to estimate stress levels, these are subjective responses that could be biased., To monitor the stress levels continuously without interrupting the user to capture the involuntary changes in physiological features, objective measures are primarily adopted.

The current developments in machine learning methods provide a great opportunity to use these stress response data to predict future stress levels and prevent risks. Deep learning neural networks has been applied in various research, including image detection in healthcare, natural language processing, detection of health conditions from electronic medical records, etc., with high accuracy and outperforming the current practices [47,210–213]. Deep learning is a type of machine learning where a model is trained to predict outputs based on the inputs with the help of multiple hidden layers. Deep learning is highly efficient compared to other traditional methods because the multiple hidden layers enhance the model performance by calculating the probability of each output and updating the weights. A recent study implemented deep learning to predict in-hospital cardiac arrest, and this model significantly outperformed other methods, including the random forest algorithm and logistic regression [214].

Deep learning can be incorporated with different techniques depending on the type of input data. One of the most common approaches used in predicting temporal sequence data is Recurrent Neural Network (RNN). The RNNs, unlike a typical feedforward NN, use their internal memory to hold the temporal behavior of the input data to predict the output. A few common RNN architectures currently used for speech recognition and time-series data predictions are Long Short-term Memory networks (LSTMs) and Gated Recurrent Units (GRUs). LSTM is a type of RNN that can keep track of the temporal behavior of the sequence without losing the long-term dependencies. The main advantage of LSTM over a traditional RNN is its ability to address the vanishing gradient problem. Vanishing gradients occur in stochastic gradient descent or any gradient-based learning methods where the NN weights are not updated as the gradient values diminish. The gradient value decreases during the backpropagation through time as the gradient values are computed by the chain rule during the backpropagation. In a few cases, the vanishing gradients stop a NN from further training. Most of the time, the NN keeps training slowly but may leave out critical information from the previous sequences resulting in developing an incorrect model for prediction. LSTMs address this issue with the help of a memory cell with gates that regulate the flow of information. Figure 5.8 below shows the fundamental design of an LSTM cell without focusing on the underlying activations and mathematical complexities.

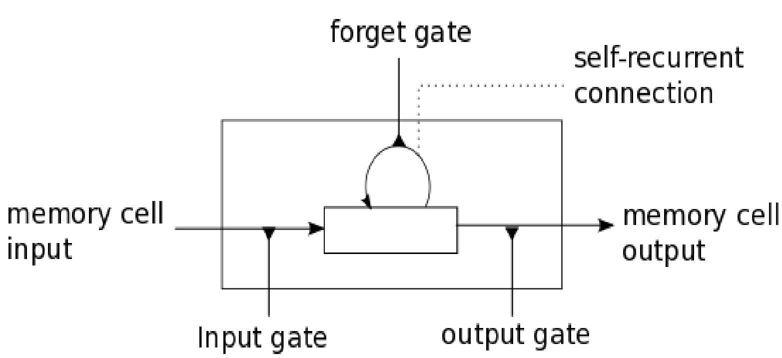


Figure 5.8: The fundamental architecture of an LSTM cell.

An LSTM has multiple gates and cell states which manage to pass the critical information without loss. Gates can be considered a passage that controls the flow of information (data) passing through them. There are three gates in an LSTM cell, including the input gate, output gate, and forget gate. The first gate in the LSTM cell is the forget gate, as this gate decides how much information from the past and new input should be allowed to the input gate. The input gate is used to update the cell state where the data from the previous hidden state and new input are transferred. The cell state, which is multiplied by the forget vector, forgets values close to zero, and the remaining values are added to the data from the input gate. The last gate in an LSTM cell is the output gate, which passes the new hidden state to the next LSTM cell, where this process is repeated. An LSTM cell has a self-recurrent connection, as seen in Figure 5.8 above. This research developed deep learning supervised LSTM to predict the physician HR and EDA based on their current HR and EDA to help them better manage an ED shift.

5.3.1 Methods

5.3.1.1 Participants

Participants for this study included 12 emergency physicians (8 male, 4 female) working a 3:00 pm - 11:00 pm shift at PRISMA Health - Greenville Memorial Hospital (GMH) in Greenville, SC. The PRISMA Health, is the largest healthcare provider in South Carolina and serves as a tertiary referral center for the entire Upstate region. This ED serves as the Adult Level 1 and Pediatric Level 2 Trauma Center, Stroke, and ST-Elevation Myocardial Infarction (STEMI) Comprehensive Center seeing over 106,000 patients annually. Six participants (mean age = 26.8 ± 1.5 years, 4 male, 2 female) were first-year resident physicians, and the other six (mean age = 42.66 ± 2.8 years, 4 male, 2 female) were attending physicians with an average experience of 8 years of practice. This particular sample set was selected to represent the diverse population of ED physicians. Consent was obtained from physicians before the shift, and the study was approved by GHS IRB Pro00058516.

5.3.1.2 Apparatus

For collecting the physiological responses (HR, HRV, and EDA), we used the Empatica E4 wearable research device that allows real-time physiological data acquisition. This wrist band is equipped with four types of sensors that record EDA, hand motions, body temperature, HR, and HRV. Multiple studies have validated the efficacy of this device, and one study specifically compared it to the gold standard Holter Monitor and observed that Empatica E4 echoed the data collected from the medical devices [164]. Additionally, multiple research studies have used this research device for computing stress, emotional arousal, sleep quality, and arterial fibrillation [167–170]. Empatica E4 collects the EDA data at a sampling rate of 4 Hz and HR data at 64 Hz.

5.3.1.3 Procedure and Data Processing

For the data collection, we first identified a resident-attending pair working an eight-hour shift in the ED. Both attending and resident physicians were familiarized with the study and asked to sign a consent form. After this, each physician was outfitted with the Empatica E4 at least five minutes prior to the beginning of the shift to collect baseline data. As detailed above, Empatica E4 collects various physiological data, including HR, HRV, and EDA. Data collected using Empatica were first preprocessed for each physician separately. Initially, the data was visualized to remove the outliers and incorrect data points. Further, the missing values were interpolated using cubic spline interpolation. [179]. Following the initial data preparation, the data was standardized to address the variations in the HR and EDA data. Later, each dataset was split into an 80:20 ratio for training and testing purposes, which roughly converts to 23,040 data points for training and 5,760 data points for testing for each physician. This split was adopted based on observations from prior studies that have reported that a physician's productivity decreases as the shift progress and increases the chances of errors and develop a robust training model [116,117]. We aimed at predicting the last 1.5 hours of the shift for each physician, which can thus help in managing the stress/fatigue experienced during the end of the shift.

Finally, after training and testing each dataset individually, the hyperparameters were further tuned to improve the accuracy of the model. To evaluate if a general model with data from multiple physicians could improve the model, the individual datasets were merged, resulting in a dataset with 345,600 data points. As each data point represented HR and EDA values for a second, two consecutive data points were averaged to reduce the dataset by half. Further, to validate and test the new model, the data were randomly split into a train, validation, and test set with a 60:20:20 split. A validation set approach was adopted to address the model overfitting issue. Following the training, the model was initially fit on the validation set, and hyperparameters were further tuned and tested on the random validation set. Finally, the model was evaluated on the test set.

5.3.1.4 Model Architecture

A deep learning neural network with a single input layer, three hidden layers, and a single output layer was developed. The input was a multi-unit LSTM with an input channel shape similar to the training data shape, i.e., the LSTM can hold t-n steps of data in the input layer, where t denotes a data point at time t and n denotes the look_back (n) function. It equips the model to learn from the past n data points as input variables to predict the output variable. The output layer was designed to hold one output value. Between the input and output layer, there were three bidirectional multi-unit LSTMs. Each layer contains 50 units (25 in each direction). A dropout rate of 0.2 was applied to the final layer, and a tanh (hyperbolic tangent function) activation was used, which resulted in the outputs ranging from -1 to 1. The output was later inverse transformed for deriving the HR and EDA. These values were selected from multiple model iterations and testing on the validation set, testing set, and prior research, which built an LSTM model for similar input data [215]. Figure 5.9 below shows the underlying architecture of the final model, which was used to predict the physician's HR and EDA. The model was implemented using the pen-source python program.

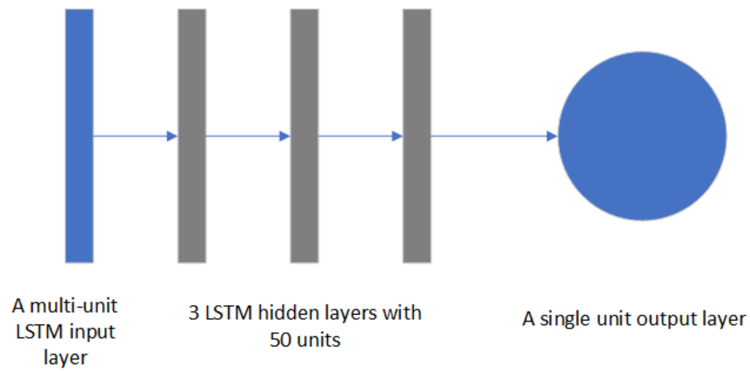


Figure 5.9: Model Architecture.

Initially, we used a multi-layer single unit LSTM; however, this resulted in underfitting where the model did not capture the temporal dependencies. To address this, the hidden layers were stacked with multiple LSTM units. Although this architecture resulted in more computational time, the multiple connections between the units assured consideration of all dependencies and improved the robustness of the model resulting in a better model fit. Figure 5.10 below shows the difference between a single unit LSTM and multi-unit LSTM cells and their computational differences. In this research, we used a 50-unit multi-unit LSTM with three hidden layers.

Additionally, in this model, we used a mean squared error method from the Keras library to compute the loss and a stochastic gradient descent algorithm: Adam. Adam is a combination of the Adaptive Gradient Algorithm (AdaGrad), which maintains a per-parameter learning rate that improves performance on problems with sparse gradients, and Root Mean Square Propagation (RMSProp), which uses the same learning rate technique that is adapted based on the average of recent magnitudes of the gradients for the weight. Adam uses the benefits of both methods to result in a better algorithm. Further, a dropout with a probability of 0.2 was used to prevent overfitting. This rate was derived from multiple model iterations and prior studies that used HR to predict cardiovascular risk [215]. Lastly, return sequences were used in this model so that the hidden state output for each input time was used for developing the model.

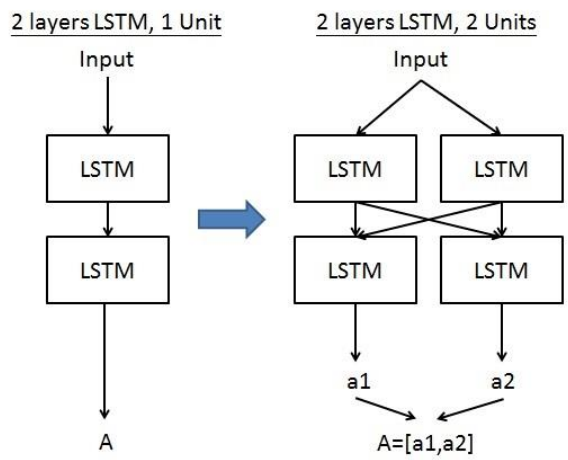


Figure 5.10: A single unit v/s multi-unit LSTM.

5.3.2 Results

This research focused on developing a multi-unit deep learning LSTM model to predict physicians' HR and EDA during an ED shift. To test the model, initially, we used n = 60 data points to predict the next data point for each physician. The model was run for ten epochs, using both return sequences and a dropout rate of 0.2. The computation time was around 15 minutes per dataset, and R-squared, Root Mean Square Error (RMSE), and loss were computed post-run. The average observed R-squared, RMSE, and loss for the HR data of twelve physicians were 0.90, 0.97, 0.004, and 0.89, 1.04, and 0.003 for the EDA data. Further, to develop and evaluate a general model, all 12 datasets were merged. Following the training, the model was validated against the validation set and evaluated on the test set. A validation set approach was adopted to address the issue of model overfitting commonly observed in the machine learning model. On the validation set, the model achieved average values of 0.97, 0.31, and 0.004 for the R-squared, RMSE, and loss for EDA, and values of 0.99, 0.44, and 0.002 for HR. On the test set, the model achieved values of 0.98, 0.17, and 0.005 for the R-squared, RMSE, and loss for EDA, and values of 0.99, 0.41, and .002 for HR. Finally, the predicted HR and EDA values were plotted against the real HR and EDA values, as represented in Figure 11 and Figure 12 below. The model was able to predict with high accuracy, as seen in Figure 11 and Figure 12 below, on the test data because of the model validation and hyperparameters tuning.

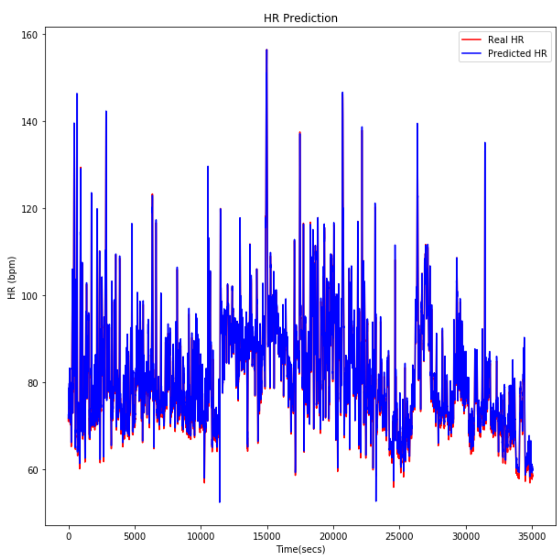


Figure 5.11: Predicted HR v/s Real HR.

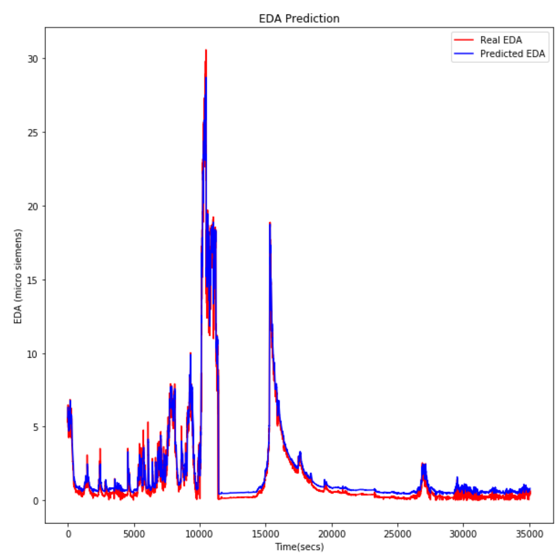


Figure 5.12: Predicted EDA v/s Real EDA.

5.3.3 Discussions and Conclusions

This research observed that a multi-unit deep learning LSTM could be used to develop general models for predicting heart rate and electrodermal activity [216]. Although the HR and EDA raw values do not add value, converting these into stress levels can help physicians better manage their shifts. We observed that training the model with more participants could develop a much more generalizable model that can better estimate the HR and EDA values. Our next step is to predict heart rate variability (HRV), which provides more precise information regarding stress and can be used to represent a stress score. Further, we plan to utilize questionnaires and other subjective feedback from physicians to gather the physician's perceived stress. This will allow for developing multi-modal datasets with subjective and objective feedback (physiological measures) to develop single-point stress scores that can be used for tracking and monitoring physician well-being. This stress score can be used to inform physicians and help them manage their shifts by taking short breaks or signing up less severe patients when stress levels are high, etc.

6. Chapter 6

6.1 Contributions and Future Work

In this chapter, we outline the main contributions of this dissertation and potential research directions.

6.1.1 Contributions

This research focused on improving patient safety, patient flow, and physician well-being in an academic ED that serves as an adult level 1 trauma center. EDs act as the healthcare safety net and is the primary entry point for millions of people seeking care in the US. Over the past several years researchers have focused on developing various strategies and tools to thwart the public health issues of ED crowding, medical errors in the ED, and burnout among ED physicians. Our work has contributed to each of these issues in the following ways:

- *Forecasting patient arrivals to the ED*: We developed long-term and short-term forecasting models that can estimate daily and hourly patient arrivals to the ED along with their ESI levels. This is the first study that considers the two-forecasting time frames and provides insights on patient severity, which can be used for planning to avoid the issue of ED crowding. Additionally, this model uses only two simple input variables, which can be accessed directly from the hospital EHR database. Although we have not reconciled the results from the long and short-term forecasts in our current approach, our next step is to improve the model further using a hierarchical reconciliation model.
- *ED physician shift design and scheduling*: This research developed a new ED shift design which by staggering (overlapping) physician shifts during the peak hours of patient arrivals to the ED. Our results show an improvement in patient flow and patient safety in the ED

evaluated using various validated metrics. These findings can be scaled and implemented in other EDs to potentially reduce ED crowding and the likelihood of medical errors. From a modeling perspective to our knowledge, our work is the first in mathematical modeling approaches (simulation and MILP) to consider patient safety as a performance metric and model individual patient-physician interactions to replicate actual ED operations.

- *ED physician burnout and well-being*: Our research makes a two-fold contribution to this area of research where we first inform how stress, well-being, and burnout varied between attending and resident physicians using multi-modal data sources. The findings from this observation can help academic/teaching EDs to better plan their shift length for attendings and residents to that they are not overloaded. Second, we developed early-stage machine learning algorithms to detect the early onset of stress using physiological measures. This observation is critical to provide ED physicians with real-time interventions and feedback on their stress levels and suggest breaks.

6.1.2 Future Work

Based on the observations from this dissertation, there are various extensions of this work within the ED. Additionally, there are other possible areas of research within healthcare where some of these methodologies, approaches, and models can be applied. The immediate potential extension of work is presented in Figure 6.1 below. Here we propose an end-to-end solution for improving patient flow, patient safety, and physician's well-being in the ED by using output from various models discussed in this research along with updated models. Using the output from the patient arrival forecasting model, we will inform the ED administrators of long-term and short-term planning. Additionally, the patient flow coordinator will use the short-term forecasts to better assign a patient pod depending on future needs and resource availability (pod parameters). Further, the physician stress detection model will inform the patient flow coordinator about the physician stress levels and workload, which they can use to match the arriving patient to a physician. The overall objective of this system is to reduce the chances of crowding and medical errors by improving the patient-to-physician assignment while accounting for future arrivals and considering current levels of physician stress, time left in the shift, and current level of crowding measured using the National Emergency Department Overcrowding Scale (NEDOCS).

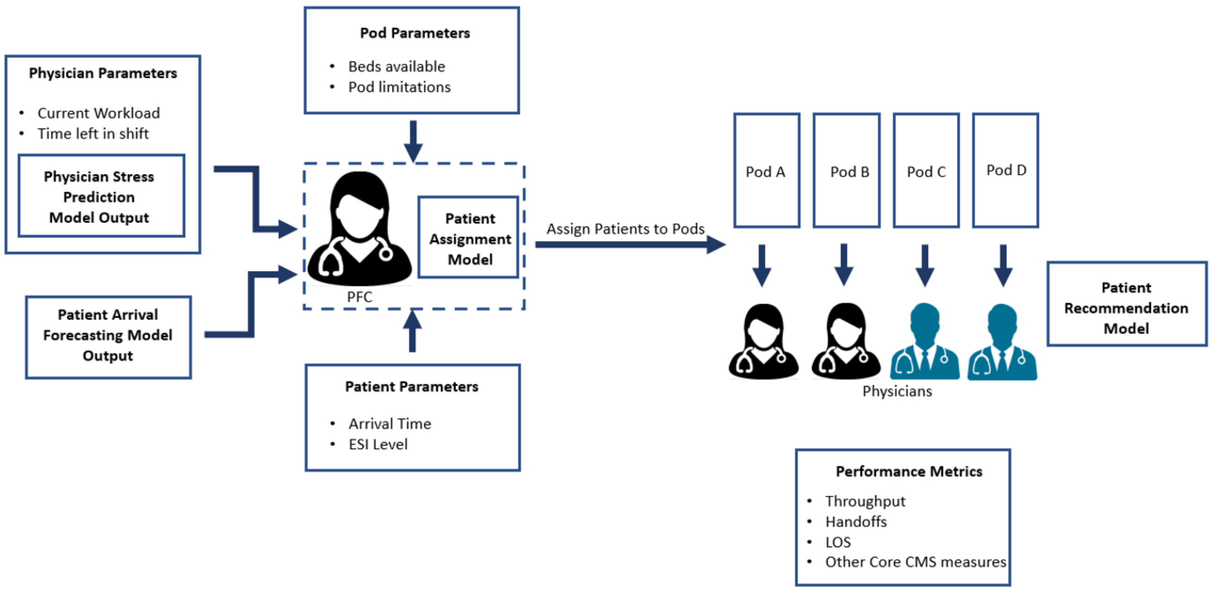


Figure 6.1: An end-to-end ED system for managing patient flow in the ED.

Another potential extension of work is incorporating the nursing team shift scheduling to get an overall perspective of the system. Additionally, consider redesigning the ED nurse and physician schedules by incorporating and analyzing factors including shift preferences and physiological parameters that influence clinician stress, burnout, and chances of medical errors. Along the same lines, physician nurse teaming and matching shift start times, etc. can be explored to identify how that impacts the overall patient flow and the likelihood of committing medical errors.

Beyond ED, the idea of incorporating physician well-being parameters measured using a combination of subjective and objective data can be used for long-term shift scheduling and real-time shift interventions. Stress scores developed by combining physiological measures and subjective data can inform on factors such as taking breaks or providing specific interventions for real-time interventions. For long-term planning, the change in physiological parameters as the shift progresses and physician preferences can help better define the shift design (length, breaks, etc.) Moreover, these well-being parameters can be considered with patient demands to define various shift lengths to avoid over and under-staffing. Finally, identifying the involuntary physiological parameters and developing models to learn patterns and detect variations can be used along with a grounded theory framework in various areas of research to detect the onset of events of interest (stress, pain, medical condition, etc.) [217,218].

References

1. Committee on the Future of Emergency Care in the United States Health System, Institute of Medicine. *Hospital-Based Emergency Care: At the Breaking Point*. National Academies Press; 2006. doi:10.17226/11621

2. Centers for Medicare & Medicaid Services (CMS). *Emergency Medical Treatment and Labor Act*. Department of Health and Human Services; 1986:262. Accessed July 8, 2019. http://www.access.gpo.gov/nara__docs/,

3. Centers for Disease Control and Prevention. FastStats - Emergency Department Visits. Published 2016. Accessed October 29, 2019. https://www.cdc.gov/nchs/fastats/emergency-department.htm

4. American College of Emergency Physicians. *Crowding*.; 2019.

5. Trzeciak S, Rivers EP. Emergency Department Overcrowding in the United States: an Emerging Threat to Patient Safety and Public Health. *Emerg Med J*. 2003;20(5):402-405. doi:10.1136/emj.20.5.402

6. Schneider SM, Gallery ME, Schafermeyer R, Zwemer FL. Emergency department crowding: A point in time. *Ann Emerg Med*. 2003;42(2):167-172. doi:10.1067/mem.2003.258

7. Hoot NR, Aronsky D. Systematic Review of Emergency Department Crowding: Causes, Effects, and Solutions. *Ann Emerg Med*. 2008;52(2):126-136. doi:10.1016/j.annemergmed.2008.03.014

8. George F, Evridiki K. The Effect of Emergency Department Crowding on Patient Outcomes

Results. *Heal Sci J*. 2015;9(1):1-6.

9. Derlet RW, Richards JR. Overcrowding in the Nation's Emergency Departments: Complex Causes and Disturbing Effects. *Ann Emerg Med*. 2000;35(1):63-68. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed5&NEWS=N&AN=2000304429

10. Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L. Emergency department crowding: A systematic review of causes, consequences and solutions. *PLoS One*. 2018;13(8):1-42. doi:10.1371/journal.pone.0203316

11. Kane L. Medscape National Physician Burnout, Depression & Suicide Report. Published 2019. Accessed July 8, 2019. https://www.medscape.com/slideshow/2019-lifestyle-burnout-depression-6011056

12. Amir Hossein Jafari-Rouhi, Sara Sardashti, Ali Taghizadieh HS and MB. Emergency Severity Index (ESI): A Triage Tool for Emergency Department | Agency for Healthcare Research & Quality (AHRQ). International journal of emergency medicine. Published 2013. Accessed July 11, 2022. https://www.ahrq.gov/patient-safety/settings/emergency-dept/esi.html

13. Swan B, Ozaltin O, Hilburn S, Gignac E, McCammon G. Evaluating an Emergency Department Care Redesign: A Simulation Approach. In: *Proceedings - Winter Simulation Conference*. Vol 2019-December. Institute of Electrical and Electronics Engineers Inc.; 2019:1137-1147. doi:10.1109/WSC40007.2019.9004947

14. Corlu CG, Maleyeff J, Wang J, Yip K, Farris J. Real-Time Nurse Dispatching Using Dynamic Priority Decision Framework. *Proc - Winter Simul Conf*. 2020;2020-

December:782-793. doi:10.1109/WSC48552.2020.9384076

15. Prabhu VG, Taaffe K, Pirrallo R. Patient Care Management for Physicians: Reducing Handoffs in the ED. In: *Proceedings - Winter Simulation Conference*. Vol 2019-Decem. ; 2019:1126-1136. doi:10.1109/WSC40007.2019.9004784

16. Prabhu VG, Taaffe K, Pirrallo R, Jackson W, Ramsay M. Physician Shift Scheduling to Improve Patient Safety and Patient Flow in the Emergency Department. *Proc - Winter Simul Conf*. 2021;2021-December. doi:10.1109/WSC52266.2021.9715398

17. Goldman J, Knappenberger HA, Eller JC. Evaluating bed allocation policy with computer simulation. *Health Serv Res*. 1968;3(2):119-129. Accessed July 20, 2020. http://www.ncbi.nlm.nih.gov/pubmed/5686350

18. Ahsan KB, Alam MR, Morel DG, Karim MA. Emergency department resource optimisation for improved performance: a review. *J Ind Eng Int*. 2019;15(1):253-266. doi:10.1007/S40092-019-00335-X/TABLES/3

19. Yousefi M, Yousefi M, Fogliatto FS. Simulation-based optimization methods applied in hospital emergency departments: A systematic review: *https://doi.org/101177/0037549720944483*. 2020;96(10):791-806. doi:10.1177/0037549720944483

20. Srinivas S, Nazareth RP, Shoriat Ullah M. Modeling and analysis of business process reengineering strategies for improving emergency department efficiency: *https://doi.org/101177/0037549720957722*. 2020;97(1):3-18. doi:10.1177/0037549720957722

21. Centers for Disease Control and Prevention. NCHS Pressroom - Fact Sheet - Emergency Department Visits. Published 2010. Accessed April 29, 2022. https://www.cdc.gov/nchs/pressroom/04facts/emergencydept.htm

22. Hsia RY, Kellermann AL, Shen YC. Factors associated with closures of emergency departments in the United States. *JAMA - J Am Med Assoc*. 2011;305(19):1978-1985. doi:10.1001/jama.2011.620

23. Di Somma S, Paladino L, Vaughan L, Lalle I, Magrini L, Magnanti M. Overcrowding in Emergency Eepartment: an International Issue. *Intern Emerg Med*. 2015;10(2):171-175. doi:10.1007/s11739-014-1154-8

24. Kelen GD, Wolfe R, D'onofrio G, et al. Emergency Department Crowding: The Canary in the Health Care System. Published online 2021. doi:10.1056/CAT.21.0217

25. Kulstad EB, Sikka R, Sweis RT, Kelley KM, Rzechula KH. ED overcrowding is associated with an increased frequency of medication errors. *Am J Emerg Med*. 2010;28(3):304-309. doi:10.1016/j.ajem.2008.12.014

26. Moskop JC, Sklar DP, Geiderman JM, Schears RM, Bookman KJ. Emergency Department Crowding, Part 1—Concept, Causes, and Moral Consequences. *Ann Emerg Med*. 2009;53(5):605-611. doi:10.1016/J.ANNEMERGMED.2008.09.019

27. Derlet RW, Richards JR. Ten Solutions for Emergency Department Crowding. *West J Emerg Med*. 2008;9(1):24. Accessed July 8, 2022. /pmc/articles/PMC2672221/

28. Alvarez R, Sandoval G a, Quijada S, Brown AD. A Simulation Study to Analyze the Impact of Different Emergency Physician Shift Structures in an Emergency Department. *Proc 35th*

Int Conf Oper Res Appl to Heal Serv ORAHS July 1217 2009 Leuven Belgium ISBN 9789081409902. 2009;(January). http://www.econ.kuleuven.be/eng/tew/academic/prodbel/ORAHS2009//1b.pdf

29. Chatfield C. Time-Series Forecasting. Published online October 25, 2000. doi:10.1201/9781420036206

30. Liu H, Tian H qi, Li Y fei. Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction. *Appl Energy*. 2012;98:415-424. doi:10.1016/J.APENERGY.2012.04.001

31. Qiu M, Song Y. Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model. *PLoS One*. 2016;11(5):e0155133. doi:10.1371/JOURNAL.PONE.0155133

32. Zhang X, Pang Y, Cui M, Stallones L, Xiang H. Forecasting mortality of road traffic injuries in China using seasonal autoregressive integrated moving average model. *Ann Epidemiol*. 2015;25(2):101-106. doi:10.1016/J.ANNEPIDEM.2014.10.015

33. Agostino IRS, da Silva WV, Pereira da Veiga C, Souza AM. Forecasting models in the manufacturing processes and operations management: Systematic literature review. *J Forecast*. 2020;39(7):1043-1056. doi:10.1002/FOR.2674

34. Kaushik S, Choudhury A, Sheron PK, et al. AI in Healthcare: Time-Series Forecasting Using Statistical, Neural, and Ensemble Architectures. *Front Big Data*. 2020;3:4. doi:10.3389/FDATA.2020.00004/BIBTEX

35. Zinouri N, Taaffe KM, Neyens DM. Modelling and forecasting daily surgical case volume

using time series analysis. *https://doi.org/101080/2047696520171390185*. 2018;7(2):111-119. doi:10.1080/20476965.2017.1390185

36. Lipton ZC, Kale DC, Elkan C, Wetzel R. Learning to Diagnose with LSTM Recurrent Neural Networks. *4th Int Conf Learn Represent ICLR 2016 - Conf Track Proc*. Published online November 11, 2015. doi:10.48550/arxiv.1511.03677

37. Pham T, Tran T, Phung D, Venkatesh S. DeepCare: A deep dynamic memory model for predictive medicine. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*. 2016;9652 LNAI:30-41. doi:10.1007/978-3-319-31750-2_3/COVER/

38. Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Sci Rep*. 2018;8(1):1-12. doi:10.1038/s41598-018-24271-9

39. Permanasari AE, Hidayah I, Bustoni IA. SARIMA (Seasonal ARIMA) implementation on time series to forecast the number of Malaria incidence. In: *Proceedings - 2013 International Conference on Information Technology and Electrical Engineering: "Intelligent and Green Technologies for Sustainable Development", ICITEE 2013*. IEEE Computer Society; 2013:203-207. doi:10.1109/ICITEED.2013.6676239

40. Zhang PG. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*. 2003;50:159-175. doi:10.1016/S0925-2312(01)00702-0

41. Yan W. Toward automatic time-series forecasting using neural networks. *IEEE Trans Neural Networks Learn Syst*. 2012;23(7):1028-1039. doi:10.1109/TNNLS.2012.2198074

42. Lu CJ, Lee TS, Chiu CC. Financial time series forecasting using independent component analysis and support vector regression. *Decis Support Syst*. 2009;47(2):115-125. doi:10.1016/J.DSS.2009.02.001

43. He W, Wang Z, Jiang H. Model optimizing and feature selecting for support vector regression in time series forecasting. *Neurocomputing*. 2008;72(1-3):600-611. doi:10.1016/J.NEUCOM.2007.11.010

44. Peng HW, Wu SF, Wei CC, Lee SJ. Time series forecasting with a neuro-fuzzy modeling scheme. *Appl Soft Comput*. 2015;32:481-493. doi:10.1016/J.ASOC.2015.03.059

45. Abdollahzade M, Miranian A, Hassani H, Iranmanesh H. A new hybrid enhanced local linear neuro-fuzzy model based on the optimized singular spectrum analysis and its application for nonlinear and chaotic time series forecasting. *Inf Sci (Ny)*. 2015;295:107-125. doi:10.1016/J.INS.2014.09.002

46. Ma F, Chiia R, Zhou J, You anzeng, Sun T, Gao J. Dipole: Diagnosis Prediction in Healthcare via Attention-based Bidirectional Recurrent Neural Networks. *Proc 23rd ACM SIGKDD Int Conf Knowl Discov Data Min*. 17. doi:10.1145/3097983

47. Che Z, Kale D, Li W, Bahadori MT, Liu Y. Deep Computational Phenotyping. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15*. ACM Press; 2015:507-516. doi:10.1145/2783258.2783365

48. Qiu X, Zhang L, Ren Y, Suganthan P, Amaratunga G. Ensemble deep learning for regression and time series forecasting. *IEEE SSCI 2014 - 2014 IEEE Symp Ser Comput Intell - CIEL 2014 2014 IEEE Symp Comput Intell Ensemble Learn Proc*. Published online January 20, 2014. doi:10.1109/CIEL.2014.7015739

49. Oliveira M, Torgo L. Ensembles for time series forecasting. In: *Journal of Machine Learning Research*. Vol 39. ; 2014:360-370. Accessed July 11, 2022. https://proceedings.mlr.press/v39/oliveira14.html

50. Ying X. An Overview of Overfitting and its Solutions. In: *Journal of Physics: Conference Series*. Vol 1168. IOP Publishing; 2019:022022. doi:10.1088/1742-6596/1168/2/022022

51. Batal H, Tench J, McMillan S, Adams J, Mehler PS. Predicting patient visits to an urgent care clinic using calendar variables. *Acad Emerg Med*. 2001;8(1):48-53. doi:10.1111/J.1553-2712.2001.TB00550.X

52. Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ, Snow GL. Forecasting Daily Patient Volumes in the Emergency Department. *Acad Emerg Med*. 2008;15(2):159-170. doi:10.1111/J.1553-2712.2007.00032.X

53. Zhang Y, Zhang J, Tao M, Shu J, Zhu D. Forecasting patient arrivals at emergency department using calendar and meteorological information. *Appl Intell*. 2022;52(10):11232-11243. doi:10.1007/S10489-021-03085-9/TABLES/5

54. Khaldi R, Afia A El, Chiheb R. Forecasting of weekly patient visits to emergency department: real case study. *Procedia Comput Sci*. 2019;148:532-541. doi:10.1016/J.PROCS.2019.01.026

55. Aboagye-Sarfo P, Mai Q, Sanfilippo FM, Preen DB, Stewart LM, Fatovich DM. A comparison of multivariate and univariate time series approaches to modelling and forecasting emergency department demand in Western Australia. *J Biomed Inform*. 2015;57:62-73. doi:10.1016/J.JBI.2015.06.022

56. Sun Y, Heng BH, Seow YT, Seow E. Forecasting daily attendances at an emergency department to aid resource planning. *BMC Emerg Med*. 2009;9(1):1-9. doi:10.1186/1471-227X-9-1/FIGURES/5

57. Côté MJ, Smith MA, Eitel DR, Akçali E. Forecasting emergency department arrivals: a tutorial for emergency department directors. *Hosp Top*. 2013;91(1):9-19. doi:10.1080/00185868.2013.757962

58. Xu M, Wong TC, Chin KS. Modeling daily patient arrivals at Emergency Department and quantifying the relative importance of contributing variables using artificial neural network. *Decis Support Syst*. 2013;54(3):1488-1498. doi:10.1016/J.DSS.2012.12.019

59. Kadri F, Harrou F, Chaabane S, Tahon C. Time series modelling and forecasting of emergency department overcrowding. *J Med Syst*. 2014;38(9):1-20. doi:10.1007/S10916-014-0107-0/TABLES/6

60. Choudhury A, Urena E. Forecasting hourly emergency department arrival using time series analysis. *Br J Heal Care Manag*. 2020;26(1):34-43. doi:10.12968/BJHC.2019.0067/ASSET/IMAGES/LARGE/BJHC.2019.0067_F04.JPEG

61. Hertzum M. Forecasting Hourly Patient Visits in the Emergency Department to Counteract Crowding. *Ergon Open J*. 2017;10(1):1-13. doi:10.2174/1875934301710010001

62. Whitt W, Zhang X. Forecasting arrivals and occupancy levels in an emergency department. *Oper Res Heal Care*. 2019;21:1-18. doi:10.1016/J.ORHC.2019.01.002

63. Carvalho-Silva M, Monteiro MTT, Sá-Soares F de, Dória-Nóbrega S. Assessment of forecasting models for patients arrival at Emergency Department. *Oper Res Heal Care*.

2018;18:112-118. doi:10.1016/J.ORHC.2017.05.001

64. Rosychuk RJ, Youngson E, Rowe BH. Presentations to Alberta Emergency Departments for Asthma: A Time Series Analysis. *Acad Emerg Med*. 2015;22(8):942-949. doi:10.1111/acem.12725

65. Becerra M, Jerez A, Aballay B, Garcés HO, Fuentes A. Forecasting emergency admissions due to respiratory diseases in high variability scenarios using time series: A case study in Chile. *Sci Total Environ*. 2020;706:134978. doi:10.1016/j.scitotenv.2019.134978

66. Siami-Namini S, Tavakoli N, Siami Namin A. A Comparison of ARIMA and LSTM in Forecasting Time Series. In: *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*. Institute of Electrical and Electronics Engineers Inc.; 2019:1394-1401. doi:10.1109/ICMLA.2018.00227

67. Cai M, Pipattanasomporn M, Rahman S. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. *Appl Energy*. 2019;236:1078-1088. doi:10.1016/j.apenergy.2018.12.042

68. Hua Y, Zhao Z, Li R, Chen X, Liu Z, Zhang H. Deep Learning with Long Short-Term Memory for Time Series Prediction. *IEEE Commun Mag*. 2019;57(6):114-119. doi:10.1109/MCOM.2019.1800155

69. Gers FA, Eck D, Schmidhuber J. Applying LSTM to Time Series Predictable Through Time-Window Approaches. Published online 2002:193-200. doi:10.1007/978-1-4471-0219-9_20

70. Li Y, Zhu Z, Kong D, Han H, Zhao Y. EA-LSTM: Evolutionary attention-based LSTM for

time series prediction. *Knowledge-Based Syst*. 2019;181:104785. doi:10.1016/J.KNOSYS.2019.05.028

71. Whitt W, Zhang X. Forecasting arrivals and occupancy levels in an emergency department. *Oper Res Heal Care*. 2019;21:1-18. doi:10.1016/j.orhc.2019.01.002

72. Hertzum M. Patterns in Emergency-Department Arrivals and Length of Stay: Input for Visualizations of Crowding. *Ergon Open J*. 2016;9(1):1-14. doi:10.2174/1875934301609010001

73. George E. P. Box GMJ. Time Series Analysis: Forecasting and Control - George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung - Google Books. Published 2015. Accessed July 11, 2022. https://books.google.com/books?hl=en&lr=&id=rNt5CgAAQBAJ&oi=fnd&pg=PR7&ots=DK66yOjZSB&sig=whYpEju58AW58KSTOYfvVJpMM68#v=onepage&q&f=false

74. Kwiatkowski D, Phillips PCB, Schmidt P, Shin Y. Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root? *J Econom*. 1992;54(1-3):159-178. doi:10.1016/0304-4076(92)90104-Y

75. Holt CC. Forecasting seasonals and trends by exponentially weighted moving averages. *Int J Forecast*. 2004;20(1):5-10. doi:10.1016/j.ijforecast.2003.09.015

76. Winters PR. Forecasting Sales by Exponentially Weighted Moving Averages. *Manage Sci*. 1960;6(3):324-342. doi:10.1287/mnsc.6.3.324

77. Hyndmann RJ, Athanasopoulos G. 7.3 Holt-Winters' seasonal method | Forecasting: Principles and Practice (2nd ed). Forecasting: Principles and Practice. Published 2018.

Accessed July 11, 2022. https://otexts.com/fpp2/holt-winters.html

78. Carmona P, Dwekat A, Mardawi Z. No more black boxes! Explaining the predictions of a machine learning XGBoost classifier algorithm in business failure. *Res Int Bus Financ*. 2022;61:101649. doi:10.1016/j.ribaf.2022.101649

79. Understanding XGBoost | Towards Data Science. Accessed July 24, 2022. https://towardsdatascience.com/de-mystifying-xgboost-part-ii-175252dcdbc5

80. What is a good MAPE score and how do I calculate it? Accessed July 25, 2022. https://stephenallwright.com/good-mape-score/

81. Center for Disease Control and Prevention. FastStats - Emergency Department Visits. National Hospital Ambulatory Medical Care Survey 2011. Published 2015. Accessed March 1, 2021. https://www.cdc.gov/nchs/fastats/emergency-department.htm

82. Behr JG, Diaz R. Emergency Department Frequent Utilization for Non-Emergent Presentments: Results from a Regional Urban Trauma Center Study. *PLoS One*. 2016;11(1). doi:10.1371/journal.pone.0147116

83. Singer AJ, Thode HC, Pines JM. US Emergency Department Visits and Hospital Discharges among Uninsured Patients before and after Implementation of the Affordable Care Act. *JAMA Netw Open*. 2019;2(4):192662. doi:10.1001/jamanetworkopen.2019.2662

84. Augustine J. . Latest Data Reveal the ED's Role as Hospital Admission Gatekeeper - ACEP Now. *ACEPNow*. Published online December 20, 2019:1-3. Accessed May 25, 2021. https://www.acepnow.com/article/latest-data-reveal-the-eds-role-as-hospital-admission-gatekeeper/

85. Maughan BC, Lei L, Cydulka RK. ED handoffs: Observed practices and communication errors. In: *American Journal of Emergency Medicine*. Vol 29. W.B. Saunders; 2011:502-511. doi:10.1016/j.ajem.2009.12.004

86. Venkatesh AK, Curley D, Chang Y, Liu SW. Communication of vital signs at emergency department handoff: Opportunities for improvement. In: *Annals of Emergency Medicine*. Vol 66. Mosby Inc.; 2015:125-130. doi:10.1016/j.annemergmed.2015.02.025

87. Westbrook JI, Raban MZ, Walter SR, Douglas H. Task errors by emergency physicians are associated with interruptions, multitasking, fatigue and working memory capacity: A prospective, direct observation study. *BMJ Qual Saf*. 2018;27(8):655-663. doi:10.1136/bmjqs-2017-007333

88. Joint Commission on Accreditation of Healthcare Organizations. Joint Commission Center for Transforming Healthcare releases targeted solutions tool for hand-off communications - PubMed. *Jt Comm Perspect*. 2012;32(8):1-2. Accessed May 26, 2021. https://pubmed.ncbi.nlm.nih.gov/22928243/

89. Lee S-H, Phan PH, Dorman T, Weaver SJ, Pronovost PJ. Handoffs, safety culture, and practices: evidence from the hospital survey on patient safety culture. *BMC Health Serv Res*. 2016;16(1):254. doi:10.1186/s12913-016-1502-7

90. Venkatesh AK, Curley D, Chang Y, Liu SW. Communication of vital signs at emergency department handoff: Opportunities for improvement. In: *Annals of Emergency Medicine*. Vol 66. Mosby Inc.; 2015:125-130. doi:10.1016/j.annemergmed.2015.02.025

91. Kachalia A, Gandhi TK, Puopolo AL, et al. Missed and Delayed Diagnoses in the Emergency Department: A Study of Closed Malpractice Claims From 4 Liability Insurers.

Ann Emerg Med. 2007;49(2):196-205. doi:10.1016/j.annemergmed.2006.06.035

92. Kolker A. Process modeling of emergency department patient flow: Effect of patient length of stay on ED diversion. *J Med Syst*. 2008;32(5):389-401. doi:10.1007/s10916-008-9144-x

93. Saghafian S, Austin G, Traub SJ. Operations research/management contributions to emergency department patient flow optimization: Review and research prospects. *IIE Trans Healthc Syst Eng*. 2015;5(2):101-123. doi:10.1080/19488300.2015.1017676

94. Komashie A, Mousavi A. Modeling Emergency Departments Using Discrete Event Simulation Techniques. In: *Proceedings - Winter Simulation Conference*. Vol 2005. IEEE; 2005:2681-2685. doi:10.1109/WSC.2005.1574570

95. Connelly LG, Bair AE. Discrete Event Simulation of Emergency Department Activity: A Platform for System-level Operations Research. *Acad Emerg Med*. 2004;11(11):1177-1185. doi:10.1197/j.aem.2004.08.021

96. Prabhu VG, Taaffe K, Caglayan C, Isik T, Song Y, Hand W. Team Based, Risk Adjusted Staffing during a Pandemic: An Agent Based Approach. In: *Proceedings - Winter Simulation Conference*. Vol 2020-Decem. Institute of Electrical and Electronics Engineers Inc.; 2020:747-758. doi:10.1109/WSC48552.2020.9384125

97. Elbeyli S, Krishnan P. *In-Patient Flow Analysis Using Pro- Model Simulation Package*.; 2000. Accessed July 20, 2020. https://www.semanticscholar.org/paper/IN-PATIENT-FLOW-ANALYSIS-USING-PROMODEL-(TM)-Elbeyli-Krishnan/6ee477db750f00d6f20d1c40f8bb24d36f0222e2

98. Harper PR, Shahani AK. Modelling for the planning and management of bed capacities in

hospitals. *J Oper Res Soc*. 2002;53(1):11-18. doi:10.1057/palgrave/jors/2601278

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

100. Oh C, Novotny AM, Carter PL, Ready RK, Campbell DD, Leckie MC. Use of a simulation-based decision support tool to improve emergency department throughput. *Oper Res Heal Care*. 2016;9:29-39. doi:10.1016/j.orhc.2016.03.002

101. Ahmed MA, Alkhamis TM. Simulation optimization for an emergency department healthcare unit in Kuwait. *Eur J Oper Res*. 2009;198(3):936-942. doi:10.1016/j.ejor.2008.10.025

102. Weng SJ, Cheng BC, Kwong ST, Wang LM, Chang CY. Simulation optimization for emergency department resources allocation. In: *Proceedings - Winter Simulation Conference*. ; 2011:1231-1238. doi:10.1109/WSC.2011.6147845

103. Keshtkar L, Salimifard K, Faghih N. A simulation optimization approach for resource allocation in an emergency department. *QScience Connect*. 2015;2015(1):8. doi:10.5339/connect.2015.8

104. Sinreich D, Jabali O. Staggered work shifts: A way to downsize and restructure an emergency department workforce yet maintain current operational performance. *Health Care Manag Sci*. 2007;10(3):293-308. doi:10.1007/s10729-007-9021-z

105. Jones SS, Evans RS. An agent based simulation tool for scheduling emergency department physicians. *AMIA Annu Symp Proc*. 2008;2008:338-342. Accessed February 28, 2021.

/pmc/articles/PMC2656074/

106. Dingley C, Daugherty K, Derieg MK, Persing R. Improving Patient Safety Through Provider Communication Strategy Enhancements. In: *Advances in Patient Safety: New Directions and Alternative Approaches (AHRQ)*. ; 2008:18. doi:NBK43663 [bookaccession]

107. Dahlquist RT, Reyner K, Robinson RD, et al. Standardized Reporting System Use During Handoffs Reduces Patient Length of Stay in the Emergency Department. *J Clin Med Res*. 2018;10(5):445-451. doi:10.14740/jocmr3375w

108. Mullan PC, Macias CG, Hsu D, Alam S, Patel B. A novel briefing checklist at shift handoff in an emergency department improves situational awareness and safety event identification. *Pediatr Emerg Care*. 2015;31(4):231-238. doi:10.1097/PEC.0000000000000194

109. Campbell D, Dontje K. Implementing Bedside Handoff in the Emergency Department: A Practice Improvement Project. *J Emerg Nurs*. 2019;45(2):149-154. doi:10.1016/j.jen.2018.09.007

110. Cheung DS, Kelly JJ, Beach C, et al. Improving Handoffs in the Emergency Department. *Ann Emerg Med*. 2010;55(2):171-180. doi:10.1016/j.annemergmed.2009.07.016

111. Yoshida H, Rutman LE, Chen J, et al. Waterfalls and Handoffs: A Novel Physician Staffing Model to Decrease Handoffs in a Pediatric Emergency Department. *Ann Emerg Med*. 2019;73(3):248-254. doi:10.1016/j.annemergmed.2018.08.424

112. Welch S, Augustine J, Camargo CAJ, Reese C. Emergency department performance measures and benchmarking summit. *Acad Emerg Med*. 2006;13(0):1074-1080.

113. Sørup CM, Jacobsen P, Forberg JL. Evaluation of emergency department performance - a systematic review on recommended performance and quality-in-care measures. *Scand J Trauma Resusc Emerg Med*. 2013;21(1):62-76. doi:10.1186/1757-7241-21-62

114. Füchtbauer LM, Nørgaard B, Mogensen CB. Emergency department physicians spend only 25% of their working time on direct patient care. *Dan Med J*. 2013;60(1). Accessed March 1, 2021. https://pubmed.ncbi.nlm.nih.gov/23340186/

115. Girishan Prabhu V, Taaffe K, Pirrallo R, Shvorin D. Stress and burnout among attending and resident physicians in the ED: a comparative study. *IISE Trans Healthc Syst Eng*. Published online September 8, 2020:1-19. doi:10.1080/24725579.2020.1814456

116. Jeanmonod R, Jeanmonod D, Ngiam R. Resident productivity: does shift length matter? *Am J Emerg Med*. 2008;26(7):789-791. doi:10.1016/j.ajem.2007.10.037

117. Silverman M. Shifting Shift Lengths: Is Shorter Better? -. Emergency Physicians Monthly.

118. Girishan Prabhu V, Taaffe K, Pirrallo RG, Jackson W, Ramsay M. Overlapping shifts to improve patient safety and patient flow in emergency departments. *Simulation*. Published online June 1, 2022:003754972210995. doi:10.1177/00375497221099547

119. Girishan Prabhu V, Taaffe K, Pirrallo R, Jackson W, Ramsay M. Reducing Handoffs and Improving Patient Flow in the Ed. *Winter Simul Conf*. Published online 2020:3-6. doi:10.21203/rs.3.rs-220758/v1

120. Hooten WM, St Sauver JL, McGree ME, Jacobson DJ, Warner DO. Incidence and Risk Factors for Progression From Short-term to Episodic or Long-term Opioid Prescribing. *Mayo Clin Proc*. 2015;90(7):850-856. doi:10.1016/j.mayocp.2015.04.012

121. Sir MY, Nestler D, Hellmich T, et al. Optimization of multidisciplinary staffing improves patient experiences at the mayo clinic. *INFORMS J Appl Anal*. 2017;47(5):425-441. doi:10.1287/inte.2017.0912

122. Ghanes K, Jouini O, Diakogiannis A, et al. Simulation-based optimization of staffing levels in an emergency department. *Simulation*. 2015;91(10):942-953. doi:10.1177/0037549715606808

123. Luscombe R, Kozan E. Dynamic resource allocation to improve emergency department efficiency in real time. *Eur J Oper Res*. 2016;255(2):593-603. doi:10.1016/J.EJOR.2016.05.039

124. Feng YY, Wu IC, Chen TL. Stochastic resource allocation in emergency departments with a multi-objective simulation optimization algorithm. *Health Care Manag Sci*. 2017;20(1):55-75. doi:10.1007/S10729-015-9335-1/TABLES/12

125. Wong TC, Xu M, Chin KS. A two-stage heuristic approach for nurse scheduling problem: A case study in an emergency department. *Comput Oper Res*. 2014;51:99-110. doi:10.1016/J.COR.2014.05.018

126. Svirsko AC, Norman BA, Rausch D, Woodring J. Using Mathematical Modeling to Improve the Emergency Department Nurse-Scheduling Process. *J Emerg Nurs*. 2019;45(4):425-432. doi:10.1016/J.JEN.2019.01.013

127. EL-Rifai O, Garaix T, Augusto V, Xie X. A stochastic optimization model for shift scheduling in emergency departments. *Health Care Manag Sci*. 2015;18(3):289-302. doi:10.1007/S10729-014-9300-4/FIGURES/9

128. Ferrand Y, Magazine M, Rao US, Glass TF. Building Cyclic Schedules for Emergency Department Physicians. *https://doi.org/101287/inte11100563*. 2011;41(6):521-533. doi:10.1287/INTE.1110.0563

129. Al-Najjar SM, Ali SH. Staffing and Scheduling Emergency Rooms in Two Public Hospitals: A Case Study. *Int J Bus Adm*. 2011;2(2):137. doi:10.5430/ijba.v2n2p137

130. Defraeye M, Van Nieuwenhuyse I. Staffing and scheduling under nonstationary demand for service: A literature review. *Omega (United Kingdom)*. 2016;58:4-25. doi:10.1016/j.omega.2015.04.002

131. Green L V., Soares J, Giglio JF, Green RA. Using queueing theory to increase the effectiveness of emergency department provider staffing. *Acad Emerg Med*. 2006;13(1):61-68. doi:10.1197/J.AEM.2005.07.034

132. Ingolfsson A, Haque MA, Umnikov A. Accounting for time-varying queueing effects in workforce scheduling. *Eur J Oper Res*. 2002;139(3):585-597. doi:10.1016/S0377-2217(01)00169-2

133. Duenas A, Tütüncü GY, Chilcott JB. A genetic algorithm approach to the nurse scheduling problem with fuzzy preferences. *IMA J Manag Math*. 2009;20(4):369-383. doi:10.1093/IMAMAN/DPN033

134. Azaiez MN, Al Sharif SS. A 0-1 goal programming model for nurse scheduling. *Comput Oper Res*. 2005;32(3):491-507. doi:10.1016/S0305-0548(03)00249-1

135. Parr D, Thompson JM. Solving the multi-objective nurse scheduling problem with a weighted cost function. *Ann Oper Res 2007 1551*. 2007;155(1):279-288.

doi:10.1007/S10479-007-0202-4

136. Burke EK, Li J, Qu R. A hybrid model of integer programming and variable neighbourhood search for highly-constrained nurse rostering problems. *Eur J Oper Res*. 2010;203(2):484-493. doi:10.1016/J.EJOR.2009.07.036

137. Ghanes K, Jouini O, Jemai Z, et al. A comprehensive simulation modeling of an emergency department: A case study for simulation optimization of staffing levels. In: *Proceedings - Winter Simulation Conference*. Vol 2015-Janua. Institute of Electrical and Electronics Engineers Inc.; 2015:1421-1432. doi:10.1109/WSC.2014.7019996

138. Salary.com. Physician - Emergency Room Salary. Published 2021. Accessed May 11, 2021. https://www.salary.com/research/salary/benchmark/er-doctor-salary

139. Woodworth L, Holmes JF. Just a Minute: The Effect of Emergency Department Wait Time on the Cost of Care. *Econ Inq*. 2020;58(2):698-716. doi:10.1111/ecin.12849

140. Prabhu VG, Taaffe K, Pirrallo R, Jackson W, Ramsay M. Physician Shift Scheduling to Improve Patient Safety and Patient Flow in the Emergency Department. In: *Proceedings - Winter Simulation Conference*. Vol 2021-Decem. Institute of Electrical and Electronics Engineers Inc.; 2021. doi:10.1109/WSC52266.2021.9715398

141. Shanafelt TD, Balch CM, Bechamps G, et al. Burnout and Medical Errors Among American Surgeons. *Ann Surg*. 2010;251(6):995-1000. doi:10.1097/SLA.0b013e3181bfdab3

142. Pruessner J, Hellhammer D, Kirschbaum C. Burnout, Perceived Stress and Cortisol Response to Awakening. *Psychosom Med*. 1999;61(2):197-204. Accessed July 9, 2019. https://oce.ovid.com/article/00006842-199903000-00012/HTML

143. Shanafelt TD, Boone S, Tan L, et al. Burnout and Satisfaction With Work-Life Balance Among US Physicians Relative to the General US Population. *Arch Intern Med*. 2012;172(18):1377. doi:10.1001/archinternmed.2012.3199

144. Popa F, Arafat R, Purcărea VL, Lală A, Popa-Velea O, Bobirnac G. Occupational burnout levels in emergency medicine--a stage 2 nationwide study and analysis. *J Med Life*. 2010;3(4):449-453. Accessed July 9, 2019. http://www.ncbi.nlm.nih.gov/pubmed/20945809

145. Patel RS, Bachu R, Adikey A, Malik M, Shah M. Factors Related to Physician Burnout and Its Consequences: A Review. *Behav Sci (Basel)*. 2018;8(11):1-7. doi:10.3390/bs8110098

146. Bragard I, Dupuis G, Fleet R. Quality of work life, burnout, and stress in emergency department physicians. *Eur J Emerg Med*. 2015;22(4):227-234. doi:10.1097/MEJ.0000000000000194

147. Babbott S, Manwell LB, Brown R, et al. Electronic medical records and physician stress in primary care: Results from the MEMO Study. *J Am Med Informatics Assoc*. 2014;21(E2):100-106. doi:10.1136/amiajnl-2013-001875

148. Friedberg MW, Chen PG, Van Busum KR, et al. Factors Affecting Physician Professional Satisfaction and Their Implications for Patient Care, Health Systems, and Health Policy. *Rand Heal Q*. 2014;3(4):1. Accessed March 30, 2020. http://www.ncbi.nlm.nih.gov/pubmed/28083306%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5051918

149. Shanafelt TD, Dyrbye LN, Sinsky C, et al. Relationship Between Clerical Burden and Characteristics of the Electronic Environment With Physician Burnout and Professional

Satisfaction. *Mayo Clin Proc*. 2016;91(7):836-848. doi:10.1016/j.mayocp.2016.05.007

150. Phipps L. Stress among doctors and nurses in the emergency department of a general hospital. *Can Med Assoc J*. 1988;139:375-376. Accessed July 9, 2019. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1268150/pdf/cmaj00174-0017.pdf

151. Institute of Medicine, Committee on Quality of Health Care in America. *To Err Is Human: Building a Safer Health System*. Vol 1. (Kohn LT, Corrigan JM, Donaldson MS, eds.). National Academic Press; 2000. doi:10.1016/s1051-0443(01)70072-3

152. Källberg A-S, Göransson KE, Florin J, Östergren J, Brixey JJ, Ehrenberg A. Contributing factors to errors in Swedish emergency departments. *Int Emerg Nurs*. 2015;23(2):156-161. doi:10.1016/J.IENJ.2014.10.002

153. Fischer JE, Calame A, Dettling AC, Zeier H, Fanconi S. Experience and endocrine stress responses in neonatal and pediatric critical care nurses and physicians. *Crit Care Med*. 2000;28(9):3281-3288. doi:10.1097/00003246-200009000-00027

154. Zwack J, Schweitzer J. If every fifth physician is affected by burnout, what about the other four? Resilience strategies of experienced physicians. *Acad Med*. 2013;88(3):382-389. doi:10.1097/ACM.0b013e318281696b

155. Vilà-de-Muga M, Colom-Ferrer L, Gonzàlez-Herrero M, Luaces-Cubells C. Factors Associated With Medication Errors in the Pediatric Emergency Department. *Pediatr Emerg Care*. 2011;27(4):290-294. doi:10.1097/PEC.0b013e31821313c2

156. Sacchetti A, Carraccio C, Harris RH. Resident management of emergency department patients: Is closer attending supervision needed? *Ann Emerg Med*. 1992;21(6):749-752.

doi:10.1016/S0196-0644(05)82797-0

157. Stucky ER, Dresselhaus TR, Dollarhide A, et al. Intern to attending: Assessing stress among physicians. *Acad Med*. 2009;84(2):251-257. doi:10.1097/ACM.0b013e3181938aad

158. Keswani RN, Taft TH, Coté GA, Keefer L. Increased Levels of Stress and Burnout Are Related to Decreased Physician Experience and to Interventional Gastroenterology Career Choice: Findings From a US Survey of Endoscopists. *Am J Gastroenterol*. 2011;106(10):1734-1740. doi:10.1038/ajg.2011.148

159. Kawamura Y, Takayashiki A, Ito M, Maeno T, Seo E, Maeno T. Stress Factors Associated With Burnout Among Attending Physicians: A Cross-Sectional Study. *J Clin Med Res*. 2018;10(3):226-232. doi:10.14740/jocmr3299w

160. Lu DW, Dresden S, McCloskey C, Branzetti J, Gisondi MA. Impact of burnout on self-reported patient care among emergency physicians. *West J Emerg Med*. 2015;16(7):996-1001. doi:10.5811/westjem.2015.9.27945

161. Chang TP, Azen C, Sherman JM. Physiological stress markers following resuscitations remain elevated throughout physician shift hours. *Acad Emerg Med*. Published online April 13, 2020. doi:10.1111/acem.13982

162. Joseph B, Parvaneh S, Swartz T, et al. Stress among surgical attending physicians and trainees: A quantitative assessment during trauma activation and emergency surgeries. *J Trauma Acute Care Surg*. 2016;81(4):723-728. doi:10.1097/TA.0000000000001162

163. Quintana DS. Statistical considerations for reporting and planning heart rate variability case-control studies. *Psychophysiology*. 2017;54(3):344-349. doi:10.1111/psyp.12798

164. McCarthy C, Pradhan N, Redpath C, Adler A. Validation of the Empatica E4 wristband. In: *Proceedings of IEEE EMBS International Student Conference (ISC)*. IEEE; 2016:1-4. doi:10.1109/EMBSISC.2016.7508621

165. Kikhia B, Stavropoulos T, Andreadis S, et al. Utilizing a Wristband Sensor to Measure the Stress Level for People with Dementia. *Sensors*. 2016;16(12):1989. doi:10.3390/s16121989

166. Cogan D, Birjandtalab J, Nourani M, Harvey J, Nagaraddi V. Multi-Biosignal Analysis for Epileptic Seizure Monitoring. *Int J Neural Syst*. 2017;27(1):165-171. doi:10.1142/S0129065716500313

167. Corino VDA, Laureanti R, Ferranti L, Scarpini G, Lombardi F, Mainardi LT. Detection of atrial fibrillation episodes using a wristband device. *Physiol Meas*. 2017;38(5):787-799. doi:10.1088/1361-6579/aa5dd7

168. Onton JA, Kang DY, Coleman TP. Visualization of whole-night sleep EEG from 2-channel mobile recording device reveals distinct deep sleep stages with differential electrodermal activity. *Front Hum Neurosci*. 2016;10:605-621. doi:10.3389/fnhum.2016.00605

169. Lutscher D. The relationship between skin conductance and self-reported stress : does the relationship exist and, if so, does it differ across different types of stressors? 2016;(June). Accessed July 5, 2019. https://essay.utwente.nl/69969/

170. Matsubara M, Augereau O, Sanches CL, Kise K. Emotional arousal estimation while reading comics based on physiological signal analysis. In: *Proceedings of the 1st International Workshop on CoMics ANalysis, Processing and Understanding - MANPU '16*. ACM; 2016:1-4. doi:10.1145/3011549.3011556

171. Greene S, Thapliyal H, Caban-Holt A. A Survey of Affective Computing for Stress Detection: Evaluating technologies in stress detection for better health. *IEEE Consum Electron Mag*. 2016;5(4):44-56. doi:10.1109/MCE.2016.2590178

172. Prabhu VG, Linder C, Stanley LM, Morgan R. An affective computing in virtual reality environments for managing surgical pain and anxiety. In: *Proceedings - 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality, AIVR 2019*. ; 2019. doi:10.1109/AIVR46125.2019.00049

173. Prabhu VG, Stanley L, Morgan R. A biofeedback enhanced adaptive virtual reality environment for managing surgical pain and anxiety. *Int J Semant Comput*. 2020;14(3):375-393. doi:10.1142/S1793351X20400152

174. Hoonakker P, Carayon P, Gurses AP, et al. Measuring workload of ICU nurses with a questionnaire survey: the NASA Task Load Index (TLX). *IIE Trans Healthc Syst Eng*. 2011;1(2):131-143. doi:10.1080/19488300.2011.609524

175. Hart SG. Nasa-Task Load Index (NASA-TLX); 20 Years Later. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol 50. SAGE PublicationsSage CA: Los Angeles, CA; 2006:904-908. doi:10.1177/154193120605000909

176. Yurko YY, Scerbo MW, Prabhu AS, Acker CE, Stefanidis D. Higher Mental Workload is Associated With Poorer Laparoscopic Performance as Measured by the NASA-TLX Tool. *Simul Healthc J Soc Simul Healthc*. 2010;5(5):267-271. doi:10.1097/SIH.0b013e3181e3f329

177. Ruiz-Rabelo JF, Navarro-Rodriguez E, Di-Stasi LL, et al. Validation of the NASA-TLX Score in Ongoing Assessment of Mental Workload During a Laparoscopic Learning Curve

in Bariatric Surgery. *Obes Surg*. 2015;25(12):2451-2456. doi:10.1007/s11695-015-1922-1

178. Malik M, Bigger JT, Camm AJ, et al. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *Eur Heart J*. 1996;17(3):354-381. Accessed August 17, 2019. https://www.escardio.org/static_file/Escardio/Guidelines/Scientific-Statements/guidelines-Heart-Rate-Variability-FT-1996.pdf

179. Peltola MA. Role of editing of R-R intervals in the analysis of heart rate variability. *Front Physiol*. 2012;3(148):1-10. doi:10.3389/fphys.2012.00148

180. Shaffer F, Ginsberg JP. An Overview of Heart Rate Variability Metrics and Norms. *Front public Heal*. 2017;5:1-32. doi:10.3389/fpubh.2017.00258

181. Castaldo R, Montesinos L, Melillo P, James C, Pecchia L. Ultra-short term HRV features as surrogates of short term HRV: A case study on mental stress detection in real life. *BMC Med Inform Decis Mak*. 2019;19(1):1-13. doi:10.1186/s12911-019-0742-y

182. Fowles DC, Christie MJ, Edelberg R, GRINGS WW, Lykken DT, Venables PH. Publication Recommendations for Electrodermal Measurements. *Psychophysiology*. 1981;18(3):232-239. doi:10.1111/j.1469-8986.1981.tb03024.x

183. Benedek M, Kaernbach C. A continuous measure of phasic electrodermal activity. *J Neurosci Methods*. 2010;190(1):80-91. doi:10.1016/j.jneumeth.2010.04.028

184. Storm H. Skin conductance and the stress response from heel stick in preterm infants. *Arch Dis Child - Fetal Neonatal Ed*. 2000;83(2):143-147. doi:10.1136/FN.83.2.F143

185. Storm H, Myre K, Rostrup M, Stokland O, Lien MD, Raeder JC. Skin conductance correlates with perioperative stress. *Acta Anaesthesiol Scand*. 2002;46(7):887-895.

doi:10.1034/j.1399-6576.2002.460721.x

186. Harrison D, Boyce S, Loughnan P, Dargaville P, Storm H, Johnston L. Skin conductance as a measure of pain and stress in hospitalised infants. *Early Hum Dev*. 2006;82(9):603-608. doi:10.1016/J.EARLHUMDEV.2005.12.008

187. Reinhardt T, Schmahl C, Wüst S, Bohus M. Salivary cortisol, heart rate, electrodermal activity and subjective stress responses to the Mannheim Multicomponent Stress Test (MMST). *Psychiatry Res*. 2012;198(1):106-111. doi:10.1016/j.psychres.2011.12.009

188. Setz C, Arnrich B, Schumm J, Marca R La, Tr G, Ehlert U. Using a Wearable EDA Device. *Technology*. 2010;14(2):410-417. doi:10.1109/TITB.2009.2036164

189. Svetlak M, Bob P, Cernik M, Kukleta M. Electrodermal complexity during the Stroop colour word test. *Auton Neurosci*. 2010;152(1-2):101-107. doi:10.1016/j.autneu.2009.10.003

190. Sano A, Picard RW, Stickgold R. Quantitative analysis of wrist electrodermal activity during sleep. *Int J Psychophysiol*. 2014;94(3):382-389. doi:10.1016/j.ijpsycho.2014.09.011

191. Braithwaite JJ, Derrick D, Watson G, Jones R, Rowe M. *A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments*.; 2015.

192. Acharya UR, Joseph KP, Kannathal N, Lim CM, Suri JS. Heart rate variability: A review. *Med Biol Eng Comput*. 2006;44(12):1031-1051. doi:10.1007/s11517-006-0119-0

193. McCorry LK. Physiology of the autonomic nervous system. *Am J Pharm Educ*. 2007;71(4):270-276. doi:10.1111/j.1399-6576.1964.tb00252.x

194. Thayer JF, Åhs F, Fredrikson M, Sollers JJ, Wager TD. A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health. *Neurosci Biobehav Rev*. 2012;36(2):747-756. doi:10.1016/j.neubiorev.2011.11.009

195. Brown DK, Barton JL, Gladwell VF. Viewing nature scenes positively affects recovery of autonomic function following acute-mental stress. *Environ Sci Technol*. 2013;47(11):5562-5569. doi:10.1021/es305019p

196. Hall M, Vasko R, Buysse D, et al. Acute Stress Affects Heart Rate Variability during Sleep. *Psychosom Med*. 2004;66(1):56-62. doi:10.1097/01.PSY.0000106884.58744.09

197. Salahuddin L, Jeong MG, Kim D, Lim SK, Won K, Woo JM. Dependence of heart rate variability on stress factors of stress response inventory. In: *Proceedings of 9th International Conference on E-Health Networking, Application and Services: Ubiquitous Health in Aging Societies*. ; 2007:236-239. doi:10.1109/HEALTH.2007.381638

198. Sloan RP, Shapiro PA, Bagiella E, et al. Effect of mental stress throughout the day on cardiac autonomic control. *Biol Psychol*. 1994;37(2):89-99. doi:10.1016/0301-0511(94)90024-8

199. Tarvainen MP, Niskanen J-P, Lipponen JA, Ranta-Aho PO, Karjalainen PA. Kubios HRV--heart rate variability analysis software. *Comput Methods Programs Biomed*. 2014;113(1):210-220. doi:10.1016/j.cmpb.2013.07.024

200. Jensen GM, Shepard KF, Hack LM. The Novice Versus the Experienced Clinician: Insights into the Work of the Physical Therapist. *Phys Ther*. 1990;70(5):314-323. doi:10.1093/ptj/70.5.314

201. Weenk M, Alken APB, Engelen LJLPG, Bredie SJH, van de Belt TH, van Goor H. Stress measurement in surgeons and residents using a smart patch. *Am J Surg*. 2018;216(2):361-368. doi:10.1016/j.amjsurg.2017.05.015

202. Michels N, Sioen I, Clays E, et al. Children's heart rate variability as stress indicator: Association with reported stress and cortisol. *Biol Psychol*. 2013;94(2):433-440. doi:10.1016/j.biopsycho.2013.08.005

203. Li Z, Snieder H, Su S, et al. A longitudinal study in youth of heart rate variability at rest and in response to stress. *Int J Psychophysiol*. 2009;73(3):212-217. doi:10.1016/j.ijpsycho.2009.03.002

204. Birhanu M, Gebrekidan B, Tesefa G, Tareke M. Workload Determines Workplace Stress among Health Professionals Working in Felege-Hiwot Referral Hospital, Bahir Dar, Northwest Ethiopia. *J Environ Public Health*. 2018;2018. doi:10.1155/2018/6286010

205. Gördeles Beşer N. *Nursing Studies and Practice International Perceived Stress Levels of Physicians and Other Health Personnel Working in the 112 Emergency Service and Associated Factors OPEN ACCESS*. Vol 1.; 2018.

206. Lesage F-X, Berjot S, Deschamps F. Clinical stress assessment using a visual analogue scale. *Occup Med (Chic Ill)*. 2012;62(8):600-605. doi:10.1093/occmed/kqs140

207. Porges SW. Cardiac vagal tone: A physiological index of stress. *Neurosci Biobehav Rev*. 1995;19(2):225-233. doi:10.1016/0149-7634(94)00066-A

208. Rose RM. Endocrine Responses to Stressful Psychological Events. *Psychiatr Clin North Am*. 1980;3(2):251-276. doi:10.1016/S0193-953X(18)30965-1

209. Girishan Prabhu V, Taaffe K, Pirrallo R, Shvorin D. Stress and burnout among attending and resident physicians in the ED: a comparative study. *IISE Trans Healthc Syst Eng*. 2020;11(1):1-10. doi:10.1080/24725579.2020.1814456

210. Choi E, Bahadori MT, Schuetz A, Stewart WF, Sun J. Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. In: *Proceedings of the 1st Machine Learning for Healthcare Conference*. ; 2016:301-318. Accessed April 25, 2019. http://proceedings.mlr.press/v56/Choi16.html

211. Deng L, Yu D. Deep Learning: Methods and Applications. *Found Trends Signal Process*. 2013;7(3):197-387. doi:10.1113/expphysiol.1998.sp004170

212. Lipton ZC, Kale DC, Elkan C, Wetzel R. Learning to Diagnose with LSTM Recurrent Neural Networks. *arXiv Prepr arXiv151103677*. Published online 2018. Accessed April 25, 2019. http://arxiv.org/abs/1511.03677

213. Razavian N, Marcus J, Sontag D. Multi-task Prediction of Disease Onsets from Longitudinal Laboratory Tests. In: *Proceedings of Machine Learning Research*. ; 2016:73-100. Accessed April 25, 2019. http://proceedings.mlr.press/v56/Razavian16.html

214. Kwon JM, Lee Y, Lee Y, Lee S, Park J. An algorithm based on deep learning for predicting in-hospital cardiac arrest. *J Am Heart Assoc*. 2018;7(13):1-11. doi:10.1161/JAHA.118.008678

215. Ballinger B, Hsieh J, Singh A, et al. DeepHeart: Semi-Supervised Sequence Learning for Cardiovascular Risk Prediction. In: *Proceedings of The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*. ; 2018. http://arxiv.org/abs/1802.02511

216. Prabhu VG, Taaffe K, Pirrallo R. A multi-layered LSTM for predicting physician stress during an ED shift. In: *Proceedings of the 2020 IISE Annual Conference*. ; 2020:1223-1228. doi:10.21203/rs.3.rs-318589/v1

217. Kalatzis A, Stanley L, Prabhu VG. Affective State Classification in Virtual Reality Environments Using Electrocardiogram and Respiration Signals. In: *Proceedings - 2021 4th IEEE International Conference on Artificial Intelligence and Virtual Reality, AIVR 2021*. Institute of Electrical and Electronics Engineers Inc.; 2021:160-167. doi:10.1109/AIVR52153.2021.00037

218. Kalatzis A, Teotia A, Prabhu VG, Stanley L. A Database for Cognitive Workload Classification Using Electrocardiogram and Respiration Signal. In: *Lecture Notes in Networks and Systems*. Vol 259. Springer Science and Business Media Deutschland GmbH; 2021:509-516. doi:10.1007/978-3-030-80285-1_58